UNIVERSITY OF ALBERTA

GENERALIZATION GRADIENTS IN THE EROTIC PREFERENCE PROFILES OF NORMAL AND PEDOPHILIC MEN.

BY

ROY R. FRENZEL ©

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF EDUCATION.

DEPARTMENT OF EDUCATIONAL PSYCHOLOGY

EDMONTON, ALBERTA

FALL 1990

ISBN 0-315-64982-8

UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: Roy René Frenzel

TITLE OF THESIS: Generalization Gradients in the Erotic Preference Profiles of Normal and Pedophilic Males.

DEGREE: Master of Education

YEAR THIS DEGREE GRANTED: 1990

Roy R. Frenzel

#103 - 140 Harrow Circle
Edmonton, Alberta
T5A 4H7

Date: October 2, 1990

UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

THE UNDERSIGNED CERTIFY THAT THEY HAVE READ, AND RECOMMEND TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH FOR ACCEPTANCE, A THESIS ENTITLED: Generalization Gradients in the Erotic Preference Profiles of Normal and Pedophilic Men SUBMITTED BY: Roy René Frenzel IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF: Master of Education

Michael Szabo, Ph.D.

Thomas Maguire, Ph.D.

George Fitzsimmons, Ph.D.

DATE: September 17, 1990

In loving memory of my grandparents, Johann and Irmgard Scherr, and Linus.

Abstract

This study was undertaken to test the hypothesis that generalization gradients would appear across age categories in the data from erotic partner preference tests. It was further hypothesized that individuals deliberately attempting to manipulate the test results would not be successful in producing profiles showing generalization. Three groups of men (normal controls, faking deviants, and nonfaking deviants) were given erotic preference tests using volumetric plethysmography. Among controls and nonfaking deviants a significant proportion evidenced generalization gradients. The number of profiles in the faking group showing this feature was not significant. It is therefore concluded that the hypotheses were supported by the data. This paper begins with an extensive discussion of erotic preference testing, and the various methods of scoring and representing the data.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

Introduction

Recent years have witnessed a growth in a specialized branch of psychology which deals with sexual behaviour. This area, known as behavioral sexology, is delimited by a series of systematic inquiries designed to investigate various aspects of human sexual behavior. Among the more important applications of the knowledge gained through this process is the accurate diagnosis of sexual preferences; particularly those of sexual deviants who pose a potential risk to those around them. The principal means of performing this task involves a specialized procedure known as phallometric, or erotic preference testing (EPT). Although these assessments are carried out in a variety of different ways, all versions share certain features in common. For example, all require that the examinee fit some type of transducing device to his penis, which is in turn connected to a recording instrument. The aim is to measure changes in penile dimensions as a carefully chosen assortment of potentially erotic visual and/or audio stimuli is presented to the examinee.

The purpose of the present study was twofold: first, to determine whether or not generalization gradients (of the type described by Hanson, 1959) are present among the erotic response profiles of normal and pedophilic men, and secondly; if they are, can this fact be exploited to assess the validity of an examinee's assessment data. As such, this investigation is largely validation work on an existing measurement procedure which will hopefully describe a strategy for detecting deliberate faking on the part of examinees.

In clinical use, EPT is of tremendous value in determining the nature of a patient's sexual anomaly. Diagnostic impressions can be formed by comparing an individual's penile responses across different stimulus categories (e.g. males to females, or consenting to nonconsenting sexual contact). Providing that adequately normed stimulus materials are utilized, it is then possible to draw conclusions regarding that person's erotic interests with a high degree of confidence. In addition, the technique is ideally suited for monitoring the effect of treatment interventions. Prior to the advent of EPT, any monitoring of this nature relied on self-reports either in the form of sexual fantasy logs or specially developed paper-and-pencil inventories.

Instruments of this nature are strongly influenced by factors such as legal status, incarceration, the absence of sexual stimuli and, most of all, the honesty of the patient. Although penile responses are not immune from the influence of these variables (as is discussed at length below), they are generally regarded as far more accurate indices (Murphy & Barbaree, 1987). Nevertheless, it is possible for certain subjects to voluntarily distort their penile responses to certain categories of erotic stimuli. Though some deceptions of this nature are readily apparent to an experienced examiner, others are more elusive and frequently escape notice. This impediment to diagnostic precision owes largely to the fact that test data are examined one trial at time. It would be preferable to assess the validity of the responses to a given category with reference to responses in other categories. Unfortunately, we simply do not know enough about the way in which responses to different stimuli relate to each other. Should the first hypothesis be supported, however, (i.e. predictable generalization gradients are proven present among response profiles), it would then be possible to statistically identify unnaturally large departures from expected response ranges. If deliberate attempts at faking erotic preference testing contribute to such departures, then a new method of assessing the validity of a set of test data could be based on this finding. This would represent a substantial contribution to behavioral sexology.

The nature of stimulus materials and their mode of presentation varies somewhat depending on the nature of the deviation being tested for. For the purposes of partner preference evaluation (i.e., the age and sex of the preferred partner) it is usual to present slides or movie segments of nude individuals covering a range of ages, and both genders (e.g. Freund, 1965, 1967). When an adequate number of stimuli from each age/gender category have been presented, the results (usually in the form of category means) are compared.[1]

An analogous approach is taken with tests of erotic activity preference. These assessments generally involve the presentation of audiotaped narratives describing a range of sexual behaviors. Some activity descriptions, such as consenting intercourse, are included to provide a benchmark to which other responses can be compared. This is necessary since individuals differ greatly in their overall level of responsiveness to the test materials. In addition, factors such as recency of ejaculation may affect the overall responsiveness of individual subjects from one point in time to the next. Some subjects produce very small penile blood volume (PBV)[2] changes, while others will show large responses to almost all categories. To illustrate the importance of this, consider an examinee who evidences a "large" reaction to a description of rape. That response may or may not be indicative of a pathological interest in rape depending on how it compares in magnitude to responses in other categories. If it were twice the size of his reaction to a description of consenting intercourse it would likely be of diagnostic importance. If, by contrast, it was only a small fraction of that size, it would be less relevant diagnostically. All of this depends, of course, on the use of stimuli which are capable of discriminating between rapists and nonrapists. Of equal importance, the examiner must be able to distinguish the responses of a deviant from those of normal control subjects. After all, even violent rapists will respond to descriptions of normal sexual contact (Abel, Blanchard, Barlow & Guild, 1977), and it would be incorrect to assume that normal men do not respond to sexual scenes with violent components. They might simply be focusing on the purely sexual aspects to the exclusion of any violent content[3]. In any case, if certain response patterns prove characteristic of a specific offender subtype, they must be identified empirically. Despite this, published research occasionally emerges which ignores the difference between face and discriminative validity. This may be partly due to the fact that it is generally assumed that an individual devotes a portion of his sexual attention to an activity (or partner type) which is proportionate to the degree of satisfaction achieved through involvement with that activity (or partner) (Murphy & Barbaree, 1987).

<u>The Efficacy of EPT as a Research/Diagnostic Instrument</u>

EPT has a well established record of success in discriminating between groups on the basis of sexual orientation. The first systematic studies were conducted by Kurt Freund, a pioneer in the area of behavioral sexology (e.g. Freund, 1963, 1967) who published data demonstrating that he could distinguish between pedophiles (those who erotically prefer pre-pubertal children), hebephiles (those preferring pubescents), and normals who were most strongly attracted to adults.

Furthermore, it was possible to determine whether the examinee was heterosexual or homosexual in gender preference. More recent studies (e.g. McConaghy, 1967, 1974; Frenzel & Lang, 1990) employing techniques very similar to Freund's have replicated and built upon his initial findings.

In addition to the early investigations by Freund, other research based on somewhat different equipment (discussed below under the heading: Circumferential penile transducers) has also been successful in finding significant differences between deviant groups using penile responses. Marshall, Barbaree and Christopher (1986) found significant differences between groups of intra- and extrafamilial child molesters, and extrafamilial abusers and controls. Differences between intrafamilial abusers and controls, however, were not significant. Similar results had previously been reported by Quinsey, Chaplin and Carrigan (1979). As a result of using EPT, the long-held suspicion that incest offenders were true pedophiles who just happened to molest family members was displaced. Results of this nature prompt researchers to sharpen their focus on other factors in attempting to understand deviant sexual behavior, and therefore serve to streamline the investigative process.

In addition to questions of age and gender preference, EPT has also proven useful in distinguishing between individuals differing in activity preference. This is true of rapists and nonrapists (Abel et al., 1977) and violent versus nonviolent child molesters (Avery-Clark & Laws, 1984). The phallometric method has also been gainfully applied to the study of transsexualism (Barr, 1973; Barr, Raphael & Hennessey, 1974), the effects of alcohol (Barbaree, Marshall, Yates & Lightfoot, 1982) and induced anger (Yates, Barbaree & Marshall, 1984) on the responses of normal men to rape scenes. The latter two studies exemplify the usefulness of EPT in the experimental investigation of sexual deviance at a very basic level.

In addition to the above-listed research applications, the penile plethysmograph[4] can be incorporated into biofeedback equipment (e.g. Laws & Pawlowski, 1973) to aid in treatment, or can simply be used as a method of obtaining pre-and post treatment data. One things is certain, however, and that is that the measurement of erectile responses to controlled erotic stimuli will grow in importance as behavioral sexology continues to develop, and the number of researchers and clinicians in that area increases.

The Measurement of Sexual Arousal in Human Males

Instruments used to transduce (convert) changes in penile dimensions to electrical output fall into three distinct categories; circumferential, volumetric, and bioimpedant. Each of these will now be described in detail.

Circumferential penile transducers.

First described by Bancroft, Jones and Pullman (1966) this variety of strain gauge generally consists of a flexible narrow diameter tube filled with an electroconductant fluid such as mercury. The tubing is closed at either end and fastened so as to form a loop. Wires are positioned at the ends to allow connection to an appropriate amplifier. The resulting device can be placed around the shaft of the examinee's penis, and acts as a variable resistor whose impedance varies in direct proportion to its circumference (i.e. as the examinee's penis increases in circumference, so does the gauge's electrical resistance). A variation on this instrument was constructed by Barlow, Becker, Leitenberg and Agras (1970) who produced a unit consisting of two rigid semicircular arms joined on one end by a small hinge-like device. The hinge contained a wire-wound strain gauge which changed in resistance as the arms were moved apart by increased penile circumference. Like the mercury-in-rubber strain gauge, this device, often referred to as a Barlow gauge, is only sensitive to circumferential changes.

The function relating degree of gauge deformation and electrical resistance is almost perfectly linear (Richards, Kalucy, Wood & Marshall, 1990), and calibration of either variety of strain gauge can be easily performed by placing the device around a number of known circumferences prior to the testing session (Earls & Jackson, 1981). The output recorded during calibration can be used as a reference against which actual test data can be scored.

Volumetric penile transducers.

Devices in this second category are known as "volumetric" because they register true penile volume changes originating both in circumference and length variations. The apparatus consists of an inflatable cuff which the examinee places around the base of his penis, and a glass cylinder which encompasses both his penis and the cuff. The latter is inflated to produce an air-tight seal between the penis and cylinder, which is closed except for a tube which carries displaced air to a suitable pressure sensor (Freund, Sedlacek & Knob, 1965). McConaghy (1974) described a volumetric transducer which differed from Freund et al.'s only in that the airtight seal was created by a latex diaphragm instead of an inflated cuff. No direct comparisons have been reported between Freund's and McConaghy's arrangements, however, the latter may have the advantage of allowing more of the penile shaft to protrude into the cylinder and may prove somewhat less cumbersome for the subject to attach.

Figure 1 is a plot showing actual changes in penile volume over the course of a single trial. The vertical ticks toward either end of the abscissa indicate where stimulus presentation began and ended respectively. The smaller ripples within each trial are artefacts caused by the subject's respiratory movements. The scale along the vertical axis is in cubic centimeters (cc or cm^3) of air displacement. Note the relative stability of the tracing in the pre-stimulus interval, and the gradual increase in volume once the material is presented. The rate at which blood volume increases varies considerably from one subject to the next, and within any given trial.

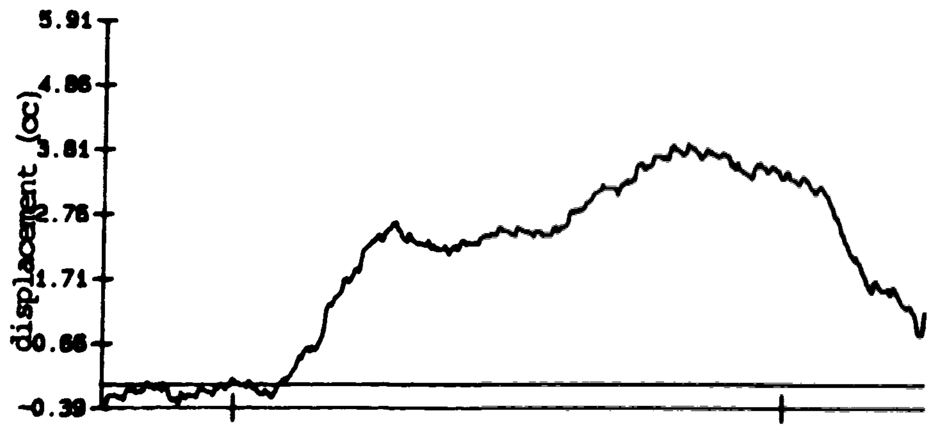


Figure 1. Sample volumetric tracing showing data gathered during the presentation of erotic material.

Hence the slope of the curve increases and decreases. When the stimulus is switched off (at the point indicated by the second tick mark), there is a gradual loss of PBV. Many subjects will continue to show some increase, however, even after the presentation has ended. The subject of the test from which this data was taken had engaged in sexual contact with 9 boys ranging in age from 9 to 14 over a period of approximately 8 years.

Bioimpedance analyzers.

This final category of penile transducer exploits the fact that skin, not unlike a column of mercury, increases in electrical resistance as it is stretched. Accordingly, the resistance changes measured during the presentation of erotic stimuli should provide a good index of the examinee's level of arousal (Bradford, 1986). A strength of this method is that it allows the simultaneous, yet independent, recording of circumferential and longitudinal changes. It's principal drawback is that the large changes in galvanic skin response (GSR) which occur in all humans independently of sexual arousal may confound the relatively small changes due to erection. This might not be a problem were it not for the fact that GSR has proven useless for the purposes of measuring changes in sexual arousal state (Quinsey, Steinman, Bergersen & Holmes, 1975; Bancroft & Mathews, 1971). Bancroft and Mathews measured changes in GSR shown by sex offenders and non-sex offenders while sexually neutral or explicit materials were presented. Their data did not reveal any significant group differences. Moreover, combining the data from the two groups indicated that responses to neutral stimuli could not be distinguished from those to erotically explicit stimuli. Due to the insufficiency of this measure, there is no basis for concluding that the bioimpedant changes resulting from a combination of GSR and arousal would be of use to sex researchers. Hence, recording equipment used in bioimpedance analysis must have the capability of separating out and rejecting GSR artefacts. To date, bioimpedance analysis has met with only limited acceptance and, to the writer's knowledge, no reports have emerged indicating that it has been successfully used in discriminating between groups or individuals on the basis of sexual preference.

As the two principal means of diagnosing sexual preference, the volumetric and circumferential methods of plethysmography have been compared to a limited degree in the literature (e.g. Wheeler & Rubin, 1987; Rosen & Keefe, 1978), with no conclusions regarding the overall superiority of either approach being apparent. Proponents of circumferential methods cite the relatively low costs of operation, short training periods, and ready availability of commercial polygraphs and other recording instruments as it's key advantages. By contrast, the volumetric apparatus is costly, must all be custom built, is quite difficult to use (training periods of several months are typical), and comparatively intrusive from the examinee's standpoint. On the other hand volumetric measures are inherently more sensitive owing to the fact that both widening and elongation contribute to the phenomenon of interest. In addition, since circumferential techniques are less sensitive, it is often necessary to present more provocative (frequently salacious) stimuli to produce adequate arousal. Some examinees will refuse to view and/or hear stimuli which they perceive as obscene, particularly if they have begun treatment and are being actively encouraged to avoid sexually explicit materials.

Although the conjecture has yet to be formally tested, the use of highly provocative stimuli may also have the effect of reducing the ability of the test to discriminate between deviants and nondeviant control subjects. This is for two reasons, the first being that more powerful stimuli tend to contain a greater number of features which are extraneous to the stimulus dimension (e.g. age) in question. Hence, the response of interest is confounded with responses to uninteresting factors. For example, a slide showing sexual contact between a man and a child may elicit larger responses than one showing a child alone, but also makes it unclear whether the examinee was aroused by the child, the man, or the voyeuristic aspect of observing a sexual encounter between two other people. Since stimuli of this nature are typically seized materials passed on to researchers by police, it is impossible to control for the presence of undesirable features. Strictly from the standpoint of experimental/test design then, materials of this nature are not appropriate.

The second reason that very salient stimuli are unsuitable is that even if confounding sexual stimuli are not present, there is an increased probability of arousing normal controls to the point of eradicating between-group differences. Since there is a ceiling effect imposed by subjects' finite erectile capability, stimuli of medium intensity should provide maximal variance between subjects showing different amounts of erotic valence toward any stimulus category. The notion of maximizing discrimination by administering items of medium salience (or difficulty) is well established in psychometric theory (e.g., Crocker & Algina, 1986).

There is a further advantage associated with volumetric methods which is a consequence of male erectile physiology. As Freund, Langevin, and Barlow (1974), and McConaghy (1974) have noted, the human penis is <u>nonisotropic</u>, meaning that at low to moderate levels of arousal, there is poor correspondence between circumferential increases and length increases. In fact, penile length may increase by as much as fifty percent before increases in circumference are noted (Freund et al., 1974). In some cases, penile tumescence may begin with length increases accompanied by a <u>decrease</u> in circumference. Initial widening, then, could indicate either tumescence or detumescence. Since the greatest level of inter-stimulus discrimination may well occur at low levels of arousal, this compromises the interpretability of circumferential data.

In view of some of the shortcomings inherent to circumferential measures, it is the writer's opinion that the advantages of the volumetric method outweigh the practical difficulties associated with its implementation. Accordingly, that technique was chosen for the present study.

The Problem of Faking

As mentioned earlier, there is a problem associated with all forms of EPT which threatens it's validity as a research and diagnostic instrument. Simply stated, many subjects can and will exert some degree of voluntary control over their penile responses (e.g., Hensen & Rubin, 1971; Kaine, Crim & Mersereau, 1988).

Keeping in mind that the typical examinee has been referred through legal channels and is charged with some form of sexual misdemeanour, it is reasonable to conclude that there may be strong motivation to produce normal-appearing test results. This tendency, however, is certainly not restricted to EPT assessment data. As Murphy and Barbaree (1987) have pointed out, the use of conventional psychometric instruments such as self-report questionnaires and personality inventories yields highly unreliable data which are essentially useless for the purposes of studying sexual deviance. Also, such tests have consistently failed to provide any means of differentiating sex offenders from other patient/inmate groups. With these issues in mind, it is clear that in order for EPT to be optimally useful to sexologists, the mechanisms of deception, to which it is vulnerable, must be understood. Some degree of progress has already been made in this regard, and some the more frequently employed faking methods are discussed below.

Pumping.

One commonly encountered faking strategy involves the voluntary contraction of the perineal and/or abdominal musculature (Freund, 1967; Freund, Watson & Rienzo, 1988). These muscles are capable of creating movement in the penile shaft which will often result in increased penile blood volume. This practice is often referred to as "pumping". From the examinee's perspective this would be a useful behavior to engage in whenever seemingly appropriate stimuli are being presented. For example, a man accused of molesting his stepdaughter may pump at a time when an image of an adult female is being presented, reasoning that the appearance of arousal might be in his best interests. Out of all the faking strategies available, this one is probably the easiest to detect (using volumetric equipment) since the movement involved results in large, fast, rather unmistakable artefacts in the output tracing (See Figure 2). The forced angulation of the penis causes a distortion in the shape of the inflated cuff which results in the pressure

Figure 2. Examples of pumping artefacts in a volumetric tracing.

changes represented in Figure 2. In this trial, the subject was shown a film clip of an adult women. In an attempt to produce tumescence, he pumped shortly after the stimulus onset, as indicated by the rapid decrease which appears immediately after the first tick mark on the horizontal axis. As indicated by the small rise following the first contraction, this effort produced a slight increase in PBV. Since the subject was not attracted to adult females, however, his PBV soon decreased. Becoming aware of this drop, he contracted the subject pumped three more times, which once again resulted in a rise. Comparing this tracing to the one in

Figure 1 highlights the rather conspicuous nature of pumping artefacts. In addition, the perineal muscles tend to fatigue quickly, and a predictable decrease in the amplitude of subsequent contractions begins almost immediately, thus adding another easily identifiable feature to the tracing.

When confronted with output curves similar to those shown, examinees will often admit to faking. Others will claim they were nervous, or that they were not consciously aware of the pumping. In any case, test tracings which contain the pumping pattern must be considered invalid, and no attempt to diagnose the examinee can made using them. The invalidation of faked test results, of course, it necessary whenever any form of faking is evidant; not just pumping. Clinically, the fact that an attempt at deception was made may prove informative in its own right as it reveals much concerning the honesty and motivation of the examinee for therapy. Information of this nature can then be used in deciding whether or not that man should be taken into treatment.

Pumping is less apparent when strain gauges are used since the muscular contractions which lead to the characteristic displacement of the penis have no immediate effect on it's circumference. To combat this problem some researchers have resorted to recording perineal contractions by placing an air bladder attached to a pressure sensor against the subject's chair at roughly mid-lumbar height. This measure is less direct and sensitive however, and perineal contractions are often difficult to separate from those resulting from normal movement. Also, unless the output from this apparatus is recorded on the same instrument used to monitor tumescence, the data from each must somehow be collated to determine where, in the stimulus presentation, any pumping took place.

Avoidance of stimuli.

A less sophisticated and less frequently encountered method of faking involves looking away from (Laws & Rubin, 1969) or not listening to the critical stimuli when they are presented. In the case of visual material, one need only focus a video camera on the subject's face to determine whether or not he is paying attention. In the case of auditory material presented through headphones the subject has little choice but to hear what is played. Ensuring that the subject actually attends to the material is less straightforward. In addressing this problem, Quinsey and Chaplin (1987) have utilized a semantic tracking task which requires the subject to depress one of two switches during stimulus presentation. When sexual material was played through a set of headphones, one of the switches was to be pressed, and when non-sexual material was played the other was pressed. Two groups of normal men were used, with only one group tested under these conditions. Subjects in both groups were instructed to suppress their penile responses to the erotically explicit stimulus materials. The results were encouraging as the control subjects succeeded in distorting their response profiles, and the semantic tracking subjects did not.

<u>Mechanical manipulation</u>.

Undoubtedly the crudest among faking techniques is any which involves a direct manual attempt by the subject to simulate changes in his arousal state. Laws and Holman (1968) for example, observed one patient place the shaft of a ball-point pen between a mercury-in-rubber strain gauge and his penis. He then used the pen to distort the gauge attempting to fabricate a normal response. Similarly, the present author has observed a small number of examinees pinch the tubing that conducts air from the volumetric cylinder to the recording device in an attempt to prevent any responses from being registered. Tactics such as these invariably result in the recording of artefacts which could not possibly follow from natural changes in PBV, hence they are not difficult for a competent examiner to identify.

Cognitive mediation.

Data manipulations involving techniques in this group are by far the most difficult to detect and most problematic for investigators employing arousal measurement methods. Very simply, it is apparent that many men are able to induce or prevent changes in PBV without resorting to any of the methods described thus far (e.g. Geer & Fuhr, 1976; Freund et al., 1989, Quinsey & Chapman, 1987). A few attempts to devise methods of either foiling or predicting faking efforts based on cognitive mediation have been made.

Studies such as Quinsey and Chapman's (1987), which was outlined previously, have examined the data produced by subjects who were directly instructed or influenced by the researchers to alter their penile responses. To gain clues as to how suppression might be achieved Geer and Fuhr (1976), utilizing a dichotic listening task, found that penile responses to erotic material played through one channel of a headset could be significantly reduced if nonerotic materials were simultaneously presented in the other channel. The results of that study suggest that it is possible to suppress responses by concentrating on nonerotic topics or engaging in complex cognitive tasks. Investigating the role of demand characteristics, Quinsey, Chaplin and Varney (1981) presented auditory stimuli describing rape to a group of normal controls. In the briefing which took place prior to the experiment some subjects were told that a certain level of arousal to rapistic themes was usual in normal men. The remaining subjects received no such information. An analysis of the data revealed that the group receiving the briefing evidenced significantly higher levels of arousal to descriptions of rape than their control counterparts. The writers postulated that some measure of disinhibition had taken place. In a more directive approach, Abel, Blanchard and Barlow (1980) instructed both deviant and nondeviant subjects to alternately enhance and suppress their responses to stimuli covering a range of activities. The results clearly indicated that subjects were able to follow these instructions closely enough to produce a significant effect; that is, they were able to successfully alter their penile responses.

The semantic tracking task (Quinsey & Chapman, 1987) described earlier represents an effort to interfere with the ability of men to suppress their responses to audiotaped descriptions of violent rape. The authors suggested that their treatment may have forced the subjects to attend more carefully to the stimulus material. An analogous procedure was implemented by Freund et al. (1989) who superimposed a small flickering light onto movie segments depicting nude individuals. The location of the light on the body was randomly varied so that the subject would have to scan the entire image in order to correctly report its location. Surprisingly, this procedure did not succeed in preventing suppression. Earlier efforts by Freund (1965) involved the injection of testosterone several hours prior to testing. This hormone, present in all men, is largely responsible for sexual drive, and increasing it's level will often lead to an increase in sexual arousability. The rationale behind administering it to examinees was that faking (specifically suppression) might be more difficult under conditions of heightened sexual responsiveness. Unfortunately, this treatment did not significantly diminish subjects' capacity to voluntarily alter their responses. Hence, apart from the semantic tracking task (Quinsey & Chapman, 1987) no procedure has as yet been shown to effectively prevent response suppression. Moreover, this single success has only been reported with rapists, and to the writer's knowledge, no effort has been made to replicate the findings.

In view of the difficulties associated with preventing cognitively mediated faking, it may prove more fruitful to concentrate on finding ways to detect it. At a rudimentary level Freund, Chan and Coulthard (1977) recommended the use of standardized questionnaires designed to assess the examinee's willingness to acknowledge deviant erotic interests. These investigators found that subjects who denied their deviant erotic inclinations were less likely to be assigned a phallometric diagnosis corresponding to their index offense than their admitting counterparts. Of course it is possible that these individuals were truthful in their claims, and had offended for reasons other than primary erotic attraction to their victim's age and gender group.

Although success in predicting or discovering faking would not allow one to diagnosis erotic preference, it would at least minimize misdiagnosis, and researchers could benefit by knowing which subjects to exclude from their data pool.

Since suppression and fantasizing have not been shown to produce any characteristic markings on the trial tracing, methods of detection based on the analysis of entire response profiles (i.e. comparative summaries based on data gathered during the presentation of <u>all</u> stimulus categories) deserve investigation. To these ends, Freund et al. (1989) have tested the hypothesis that two distinct age preference profiles exist which are suggestive of faking. The first of these is present when the mean response to erotically neutral stimuli occupies the highest or second highest position within the response magnitude hierarchy. The second occurs when the second highest mean response to the various age/gender categories is to a category of the gender opposite that eliciting the greatest overall mean response. For example, if the examinee's greatest response was to adult females, and the second highest mean response was to prepubescent males, this criterion would be met. Freund et al. added the requirement that the greatest response must be to an adult. These investigators found support for the use of both of these profiles as indicators of faking. It must, however, be noted that the subjects in this study were all paid heterosexual nonoffenders who were instructed only to fake a pedophilic age preference, during the first session and a homosexual orientation in the second. Hence, it is possible that either of these configurations could appear in the response profiles of true pedophiles or hebephiles even if no faking had taken place.

Cognitive mediation might result in either a decrease or an increase in PBV. In the first case, the subject might simply initiate some favourite fantasy he knows will increase his level of arousal. Since this process is not observable the investigator is likely to conclude that any change arousal he measures is a response to the stimuli being presented. Alternatively, and probably more frequently, some subjects will attempt to suppress increases in PBV by performing mental arithmetic, cognitively rehearsing nonsense syllables, or concentrating on something incompatible with sexual arousal. On the output tracing this will usually appear as a gradual drop in PBV or a failure to evidence tumescence in response to the critical stimulus. Since EPT diagnosis is based on differences in one's responses to stimuli representing various activity or partner categories, being suspicious of a response only because it seems incongruent with a client's history is tantamount to doubting the procedure's validity. More often the examiner concludes, as the subject had intended, that the corresponding stimulus was simply not arousing. Indeed a decrease in PBV coinciding with the onset of the critical stimulus might be construed as evidence that the subject found the stimulus erotically aversive.

Pertinent Learning Principles

Since the concept of generalization is central to the present paper, it is worth reviewing here along with some related concepts. Frequent reference to both operant and classical learning models will be made for the sake of explanatory ease but this is not intended to endorse or refute the position that sexual preferences are learned phenomena. Generalization, and generalization gradients are observed throughout psychology and physiology, and may be discussed without adopting a learning perspective.

Stimulus generalization.

Stimulus Generalization refers to the process whereby stimuli other than the original Conditioned Stimulus (CS) comes to elicit a Conditioned Response (CR), or a response similar to it. If an animal learns to salivate upon presentation of a 200 cycle per second (Hz) tone as a result of its prior pairing with an Unconditioned Stimulus (UCS) such as food, the tone is said to be a CS. In addition, the salivation which follows the CS is termed a Conditioned Response (CR). If, through subsequent learning, a frequency other than 200 Hz comes to elicit the CR, stimulus generalization, in the Pavlovian sense, has occurred.

There is also an operant definition. In this case, assume that an animal has learned to bar press for food reinforcement when a 200 Hz tone is sounded. The relationship can be denoted as follows:

$$S_D:R \dashrightarrow Rf$$

where :

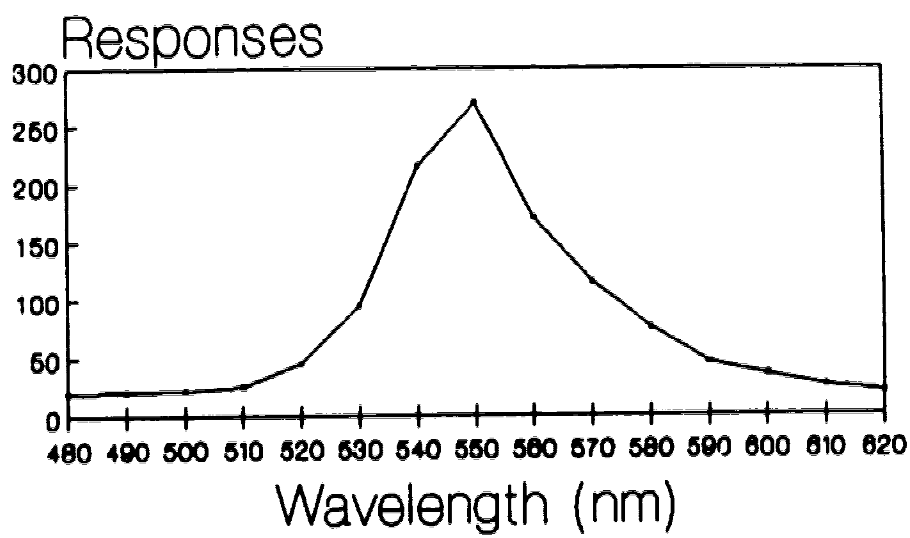
- S_D is known as a discriminative stimulus (200 Hz tone)
- R is some operant response (bar pressing)
- Rf is a reinforcement, or consequence of the behavior (food)

In the operant usage of the term, stimulus generalization refers to the process whereby discriminative stimuli (S_D's) other than the original come to signal the R, Rf relationship and thereby control responding.

Generalization gradients.

An interesting phenomenon associated with both Pavlovian and Skinnerian paradigms exists: The appearance of generalization gradients. These describe the ability for stimuli qualitatively similar to, but quantitatively somewhat different from the original S_D or CS to elicit responses similar to R or CR. As the stimuli becomes increasingly distinct from that employed during the initial acquisition trials, the magnitude or rate of responding decreases correspondingly. If a graph is drawn which relates rate (or magnitude) of responding to eliciting stimulus (S_D or CS), a characteristic unimodal distribution[5] emerges with a peak corresponding to the value of the S_D (or CS) used during conditioning. The other points along the abscissa would represent stimuli which differ increasingly from the original S_D along some quantifiable dimension such as size, shape, wavelength, etc.

Hanson (1959) performed a typical operant learning task in which pigeons were initially reinforced for pecking a key switch illuminated by a 550 nanometer (nm) light source. No other lights were used as discriminative stimuli. During testing phase, lights of either shorter or longer wavelengths were presented. The number of



<u>Figure 3</u>. Excitatory generalization gradient. After Hanson, 1959.

responses made are graphed as function of light wavelength in Figure 3. The slope of the curve is characteristic of a generalization gradient with maximum responding occurring to the original S_D^+. As the light wavelength is increased or decreased, the rate of operant responding decreases. If responses to light frequencies other than the original are reinforced, future presentations of that stimulus would be marked by increased responding. Conversely, unreinforced frequencies will come to elicit fewer responses. This process whereby an organism comes to narrow the range of stimuli that it will respond to is referred to as discrimination learning. Alternatively, reinforcing responses which occur in the presence of light of other wavelengths will widen the gradient. In essence, the organism learns the preposition: respond in the presence of a light. Frequency becomes an irrelevant variable.

For the purposes of the present discussion, there are two points to kept in mind with respect to discrimination and generalization, both of which are likely to have direct relevance for understanding penile responses during EPT. Firstly, a certain amount of stimulus generalization is inherent to all learning. Therefore, even without instituting any further training, stimuli other than the intended S_D or CS will affect some aspect of responding. Hence the appearance of generalization gradients. This is equally true of UCRs which occur naturally, as such, it was previously stated that one needn't adopt a learning perspective in order to study these processes. Secondly, as stimuli come closer and closer to the S_D or CS in similarity, the degree of responding they result in will increase.

If an organism is exposed to an inhibitory contingency in which it learns that a certain stimulus signals the non-availability of reinforcement, or that bar pressing will no longer be followed by food reinforcement, the symbol used to represent that stimulus is often suffixed with a '-'. Therefore, S_D becomes S_D^-, and CS becomes CS^-. The opposite is also true, meaning that stimuli signalling availability are suffixed with a '+' sign to denote an excitatory relationship. When either CS or S_D are written alone, it is generally understood that an excitatory ('+') as opposed to inhibitory ('-') contingency is in effect.

As has already been stated, it is a hypothesis of the present study that sexual responses follow a generalization gradient. This implies that erectile responses will be at a maximum in the presence of certain critical stimuli, and will gradually decline as the S_D is varied somewhat.

Though it is beyond the scope of the present thesis, it is certainly interesting to consider the role that discrimination learning may play in the acquisition and refinement of sexual responses. For a thorough discussion of this, the reader is referred to recent paper by Laws and Marshall (1990).

Research Hypotheses

Examination of hundreds of erotic preference profiles has led the writer to suspect that the progression of response magnitudes from one age category to the next is characterized by a predictable transition such as that observed in generalization gradients. In addition, the responses contributing to the mean of any single category seem to evidence little variability. With these two observations in mind, a series of regression analyses on the data gathered from individual erotic preference tests were performed to determine whether or not such gradients were present among a significant number of EPT profiles. The rationale for using regression results as the criterion of generalization is explained more fully under the Analyses heading in the Method section.

Since neither response suppression nor the recruitment of fantasy are readily detectable by either circumferential or volumetric means when simply inspecting the response curves, it is necessary to examine alternative methods of assessing the validity of

EPT data. It is the writer's belief that the erotic response profile must be examined as a whole in order to find clues to the presence of cognitive manipulation. The second hypothesis to be tested here, therefore, is that although a subject may be able to fake a trial tracing which appears convincing, it will be far more difficult to feign generalization in his profile. Accordingly, faked tests will be identifiable by an inordinately high proportion of error variance in the regression analysis of data from the preferred gender categories. The above two hypotheses can be stated more formally:

Hypothesis 1: generalization gradients will be evident among the data gathered for normal and pedophilic men as evidenced by significant *F*-tests of a regression model fit to the data of the preferred gender.

Hypothesis 2: subjects who fake, or are instructed to fake will be identifiable by a disproportion of error variance, as measured by an insignificant *F*-test.

The Scoring of Phallometric Data

Regardless of which type of transducer is employed, the investigator must eventually decide on a method of representing the test data. Three distinct approaches to this problem have been suggested, each of which is discussed below. Briefly, they are: raw displacement data, maximum displacement expressed as a percentage of full erection, and maximum displacement converted to ipsative *z*-scores. Each of these is discussed in detail below. It is necessary to preface any discussion of scoring practices with a comment on *what* is being scored. There are basically two options; one is to score the area under the curve defined by the data, and the other is simply to use the maximum displacement from baseline noted during a trial. Though different researchers have chosen variantly from these options, it turns out that the two are closely related, with correlations ranging from 0.86 to 0.95 (e.g. Quinsey & Harris, 1976) typically reported. Hence, the decision should probably be based on questions of computational convenience, and the type of recording equipment used. Although maximum displacement is probably the simpler of the two strategies, computerized data collection methods have greatly simplified the task of scoring multiple data points such as the type necessary to produce area approximations. In addition, area measures are less readily swayed by artefactual fluctuations, and will therefore be less likely to require editing prior to test scoring.

Raw data.

The simplest way to represent penile responses is in terms of the units displayed on a recording instrument. In the case of a polygraph, it is usual to record millimeters of pen deflection from the position occupied under baseline conditions. If the apparatus has been precalibrated with known circumferences, these values can be easily converted to linear quantities. With volumetric equipment it is standard practice to inject known volumes into the airway once it is attached. This is analogous to precalibrating a strain gauge.

The chief advantage of scoring data in this form is convenience. The principal limitation is that it complicates the comparison of different individuals who naturally differ in overall responsiveness. That is, one examinee may show large responses throughout the test while another responds to a much lesser degree.

Percentage of full erection.

To offset the problems associated with scoring raw data, many researchers have resorted to expressing each response as a percentage of the examinee's maximum erectile capability, or percent of full erection (PFE).

This practice yields data in a form which retains information about the individual's level of responsiveness, yet equates all examinees in terms of relative erectile capability. It is also useful for tests in which full erection is likely to be encountered frequently and ceiling effects are therefore of concern.

A problem associated with this approach is that a full erection must be achieved at some point during the testing session. Perhaps not surprisingly, the clinical surroundings may prohibit many individuals from reaching this extent of arousal, and it is not always possible to utilize this measure (e.g., Earls, Quinsey & Castonguay, 1987). A further difficulty is that many men who are able to achieve full erection when directed to do so will nonetheless evidence only very small responses during the actual stimulus presentation. This is particularly true when direct physical stimulation is initially permitted, or the examinee is shown very provocative materials in the interests of producing a full erection. Investigators using this approach generally require that some minimal level of responding, typically ten to twenty PFE, be achieved before the data can be considered meaningful (e.g. Murphy & Barbaree, 1987). This practice is intended to ensure that the observed displacements are not merely the result of random fluctuation. Unfortunately, penile reactions are most difficult to voluntarily influence at low levels (Malcolm, Davidson & Marshall, 1985). Therefore, using stimuli which yield higher responses will make the test more vulnerable to faking.

There is a further scoring procedure which is based on PFE, but uses mean response to a normal stimulus category in the place of the value corresponding to full erection. For example the response to a rape scene, could be divided by that to consenting sex. The quotient is often referred to as an "index" (Abel et al., 1977), and simply compares responses resulting from exposure to deviant stimuli with those to non-deviant material. This index, of course, eliminates any information concerning subjects' overall responsiveness. The advantages of this measure are twofold, firstly; full erection need never be attained, and secondly, subjects are equated on the basis of overall responsiveness.

Ipsative z-scores.

The third, and probably most sophisticated, method of scoring EPT data involves the expression of each response as a z-score (e.g. Freund, 1967). Like the index measure described above this strategy does not require the subject to reach a full erection, and all subjects are equated in terms of responsiveness.

As Earls et al. (1987) have pointed out, *z*-scores better reflect stimulus effects because they are insensitive to individual differences in overall responsiveness. Unfortunately, this means that error variance will also play a more prominent role. Barbaree (1987) has pointed out that this creates a problem for researchers using group data because subjects who respond randomly contribute as much to the data as those who respond exclusively to one stimulus category. Indeed Quinsey and Chapman (1984) found that coefficient α values ranged from 0.50 to 0.84 among responses to stimuli of the same categories in their test when using raw scores, but fell to the 0.35 to 0.65 range when the same data were expressed in terms of *z*-scores. Of course this was largely a result, as they correctly noted, of removing that component of variance due to subject responsiveness.

A more meaningful comparison of *z*-scores and raw scores was offered by Earls et al. (1987) who conducted a study which was methodologically similar to Quinsey and

Chaplin's. These investigators, however, chose to compare the value of the *F* statistic resulting from the use of data expressed in different forms, and found that among raw, PFE and *z*-scores, the latter produced the largest value. This finding is interesting here for two reasons: first, the use of error terms in the assessment of test validity is central to the present study and; second, it addresses Barbaree's (1987) concern about the relative contributions of error and stimulus effects to total variance. Had Barbaree's concern been justified, using *z*-scores would not have produced a higher *F* ratio. It is importanc to keep in mind that the use of strain gauges with subjects who evidence low overall levels of arousal is complicated by the effect Freund et al. (1974) identified. Once again, the penis may elongate significantly before increasing in circumference, or circumference may actually *decrease* while this is occurring. Consequently, a researcher employing circumferential measures cannot necessarily be certain if preliminary circumferential changes reflect an increase or a decrease in arousal. By contrast, volumetric strategies *are* able to utilize data gathered in this range since length, as well as circumference contribute to penile volume. It is possible that this phenomenon exacerbates the problem of high error variance in circumferential recordings made at low to moderate levels of responding.

A point should also be made here regarding the use of appropriate stimulus materials for the purpose of EPT. Since multiple presentations of stimuli from each category are required to provide the test with adequate reliability, some of the error variance must be attributed to the differences among the materials within categories. Ideally, all stimulus materials should be carefully controlled so that the influence of factors extraneous to the stimulus dimensions under test (e.g. age, gender, degree of force) are eliminated. In addition, the sequence in which stimulus categories are presented should not be predictable by the examinee and presentation order should be counterbalanced, where possible, to counteract serial order effects. As Kolarsky and Madlafousek (1977) found, large responses tend to increase the magnitude of responses in following trials. Similarly, small responses tend to be followed by small responses.

To anyone versed in experimental/test design the need for appropriate stimuli may seem self-evident, however, practical problems such as the limited availability of photographs containing unclothed children remain a factor. As was previously mentioned, sexually explicit materials dealing with deviant themes are generally available to researchers through police departments, but are usually inappropriate for precisely those reasons discussed above. Better alternatives to traditional sources of photographs may evolve from advancements in graphical computer simulations. Also, the potential for using image enhancing software to modify existing materials has yet to be exploited.

Method

Subjects

Incarcerated sex offenders in the province of Alberta are eligible to apply, at their option, for admission to the Sex Offender Program at Alberta Hospital Edmonton. Each one is given erotic preference testing as a standard part of his intake evaluation. All admissions to the program must be at least 18 years of age, though there is no upper age limit or restriction placed on the nature of the offence. The present study included only those individuals who were referred for questions related to partner age preference. That is, they had a legal history of sexual misconduct with one or more minors, and did not deny this involvement at the time of their application to the program. Personal data such as age and education were available for all subjects.

Three types of examinees formed the subject pool for this study. The first was a group of 155 deviants (D) 155 who had applied to the program through the correctional system. Members of this group displayed profiles which contained no evidence of faking but were consistent with their self-reported erotic preferences and offence history. There were also 62 normal controls (NC) (described below), and 22 deviant examinees who had been identified as fakers (FD) by the presence of pumping artefacts, and/or their efforts to avert their gaze from the stimulus materials. They were otherwise similar to members of the D group.

NC subjects had no police record of sexual misconduct and acknowledged no sexual interest in children. They were recruited from Canada Manpower Centers in the city of Edmonton or a newspaper advertisement which was identical in content to those posted in the Manpower offices. NC subjects were tested according to the same protocol as the other subjects, but were paid a total of $50.00 for their participation in this study, and completing a number of self administered questionnaires which were not part of the present study.

A computer program was specially written (included in Appendix 1) to randomly select 10 subjects from each group. While the 3 pools were not of equal size, all subjects within each were selected randomly by the program. The groups were restricted in size to 10 members each primarily to avoid having to select an inordinately high proportion of subjects from either the NC or FD groups. Group sizes of 20, for example, would have led to an inclusion of all but two subjects in the FD group.

Apparatus

The testing laboratory consisted of a nine by fourteen foot sound insulated room with a projection screen on one end, and a comfortable 'Lazy-Boy' chair positioned opposite the screen. A Canon infrared camera was positioned on the wall between the chair and screen in plain view of the examinee. Attached to the camera there was a small infrared light source which made it possible to observe the examinee during testing, even in the absence of any visible light. A cabinet and sink were located immediately next to the chair, but were out of view to an examinee looking toward the screen.

Adjacent to the testing room there was a control area from which the examiner conducted each test. This area housed an IBM PC XT microcomputer, TECMAR Model 1004 12 bit LabMaster (an analog-to-digital and digital-to-analog converter which allows data to be gathered by the computer), a Bell and Howell Filmosound 16mm movie projector, two Kodak Carousel diapositive slide projectors, and an air pressure-to-voltage transducer which can register volume changes as small as 0.02 cm^3

Film and slide projectors were interfaced to the microcomputer so that stimulus presentation could be achieved automatically and easily coordinated with data acquisition. Software was available to analyze, score, plot, and archive all data following the test, and to automate the actual testing and data acquisition processes.

Stimuli

Stimuli were in the form of a 16mm color film (Freund, 1967) showing nude men, women, and children of both sexes walking toward the examinee (i.e. becoming psychologically less distant). Each segment showed two models (in the same age/gender category), one after the other. Stimulus models were shown individually, and no segment depicted any sexual behavior or gesturing. All models were rated according to the Tanner system which, once again, is a method of objectively rating physical maturation according to the presence of secondary sexual characteristics and overall body shape. The scale ranges from 1, (indicating development slightly higher than that of infancy), to 5 which indicates adulthood. The distribution of Tanner scores used in the present study is shown in Table 1.

In all cases, the models were filmed against a non-descript background consisting of a carpeted floor and draped wall. All individuals depicted fell into one of four age categories, 6 to 8 years (corresponding to a mean Tanner score of 1), 10 to 12 years, a mean Tanner score of 2.5, puberty, with a mean Tanner score of 3.5, or adulthood, a Tanner score of 5. These age and Tanner groups were used for both genders, yielding a total of 8 critical stimulus categories. In addition, a ninth category was included which depicted only erotically neutral stimuli such as landscapes and footage of a lake shore. Each stimulus category was represented once in

Table 1

Mean Tanner Ratings of Models Appearing in the Test Stimuli

Category	Model Number					
	1	2	3	4	5	6
A	1.2	1.0	1.0	1.0	1.5	1.0
B	2.0	2.0	2.0	2.0	2.0	2.5
C	3.5	3.0	3.5	3.0	3.0	3.5
D	5.0	5.5	5.0	5.0	4.5	5.0
E	1.0	1.0	1.0	1.0	1.0	1.0
F	1.5	2.0	2.0	2.5	2.0	2.0
G	4.0	3.0	3.5	3.5	3.5	3.5
H	5.0	5.0	4.5	4.5	4.5	5.0

each of 3 stimulus blocks, (consisting of one segment from each of the nine categories). Within blocks, the categories were pseudo-randomly arranged so that order effects would be controlled for. In all, each subject viewed 27 test trials prefaced by three warmups for a total of 30 trials. This stimulus material has been in use for several years, and has been distributed among a number of research and diagnostic centers throughout North America and Europe. A large body of data has been amassed attesting to its ability to discriminate between homosexual and heterosexual pedophiles, as well as those with an adult orientation (E.g. Freund, 1967; Freund et al., 1977).

Procedure

Upon referral for erotic preference testing, each examinee was provided with a thorough explanation of the procedure, an opportunity to view the laboratory and equipment, and was invited to ask any questions he might have. Prior to testing, it was stressed that the evaluation could only take place with the express (and written) consent of the examinee. If agreeable to testing, he was ushered into the testing room and asked to be seated. At that point he was asked to lower his trousers and underwear as far as his ankles and attach the volumetric devices under the supervision of the examiner. The subject was then provided with a linen sheet to cover himself with, and a pair of headphones through which instrumental music was played during testing. Music was played in order to mask the noise created by the film projector. Each participant was encouraged to remain still for the duration of the procedure, and cautioned to allow his responses to occur without deliberate cognitive or physical interference. In all cases, testing was completed in a single session.

Departures from the baseline penile volume were closely monitored by the examiner throughout the procedure using the apparatus described above. Each trial was begun only when the examinee's resting penile volume had been within ±0.5 cm^3 of baseline for a minimum of 5 seconds. Data were gathered during a 25 second stimulus presentation interval, as well as 5 seconds prior to, and 5 seconds following the stimulus epoch. Between trials, data was stored on floppy disk, and the film was automatically advanced to the start of the next segment by the computer. No image was projected while the film advanced. This sequence of events was repeated until the conclusion of the test, at which time the penile cuff was deflated and the subject removed the equipment. All devices coming in contact with the examinee were placed immediately in an activated glutaraldehyde sterilizing solution (Cidex), and the examinee was required to scrub his hands with a bactericidal soap prior to leaving the laboratory. After all analyses had been completed, subjects were advised as to the outcome of their testing. This took place as time permitted, but generally within a few days of testing.

<u>Analyses</u>

The data gathered during each test were in the form of 27 distinct trials; each one representing penile volume change versus time. In keeping with accepted scoring standards, the area between the baseline (0.00 cm^3) and the response curve was approximated and used to determine if the response was positive or negative in direction. If positive, the largest positive displacement represented that trial. Conversely, if the trial was found to be negative, the largest negative value was taken. Maximum displacements were used because subjects often begin to evidence a loss of PBV before the end of the trial. Also, recording maximum displacement wherever it occurs in the presentation ensures that the trial duration is not effectively shortened as would be the case if data were consistently gathered at some fixed point between the beginning and end. The three blocks of nine trials were then separately converted to *z*-scores in order to remove between block error resulting from fatigue effects. The data were then subjected to a regression analysis which fit a curve relating response to the mean Tanner score represented by each stimulus category. For these purposes, a program was written which would read in the data, convert them to *z*-scores and perform the regression analyses. This accuracy of this program was verified by comparing it's output with that of existing statistical software. The decision to create this program, rather than rely on existing software, was based primarily on the archived data format, which was incompatible with available statistical packages.

Results were output to a file for later inspection, as well as output to the screen so as to provide a less cluttered summary. The program and a sample of the output are included in Appendices 2 and 3 respectively. A separate equation was derived for male and female stimuli, and neither included the neutral category, though these values were not removed until after the data were converted to *z*-scores. To produce a nonfaking group, data from the NC and D groups were combined, and significant *F* tests of the regression models were the criterion used to determine whether or not a generalization gradient was present in each case. A binomial test was then conducted to determine whether or not a significant number of subjects had evidenced this pattern. Finally, the second hypothesis was tested by subjecting the data from the FD group to similar analyses.

Where the regression line fitted to these data is

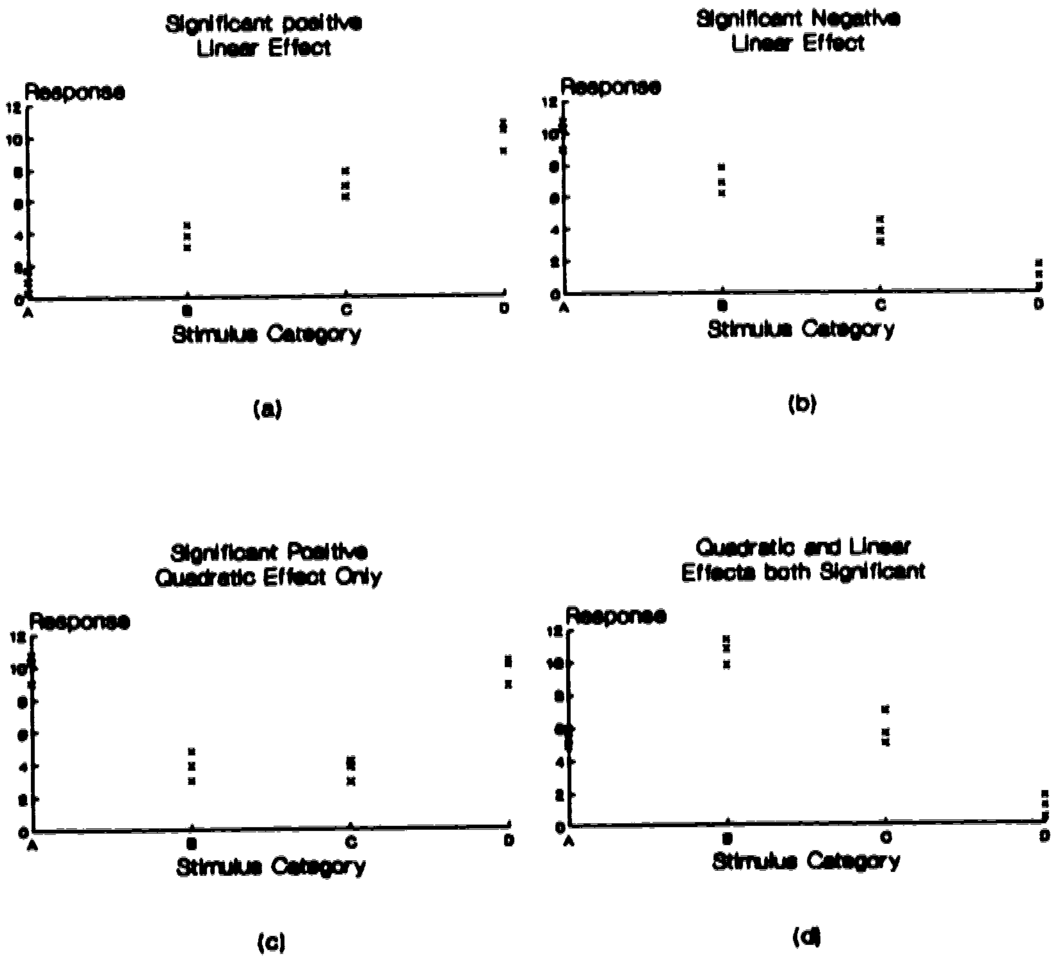


Figure 4. Generalization gradients in four hypothetical erotic preference profiles. The categories are: A - Tanner 1, B - Tanner 2.5, C - Tanner 3.5, D - Tanner 5.

noncurvilinear, the slope direction, as indicated by a positive or negative value of the linear effect coefficient, β_1, indicates whether the subject's largest responses occurred to younger or older individuals (see Figures 4a and 4b). Moreover, the absolute magnitude of β_1 indicates the strength of the linear relationship between age and penile response. Where β_1 is small, the regression curve will either be fairly flat, indicating little age discrimination, or it will be curvilinear. Where the value is large, the degree of discrimination is high.

The quadratic effect coefficient, β_{11} provides an index of the degree of linearity among the category means. If the greatest difference in category means does not occur between reponses to the most and least developed females, β_{11} will take on increased importance. Where it is large and negative in sign, the curve is convex, indicating a peak response corresponding to a Tanner score of 2.5, or 3.5 (see Figure 4d). Conversely, a sufficiently large positive value for β_{11} will indicate that the curve is essentially "U" - shaped (Figure 4c).

Regression coefficients beyond the second order were not calculated in the analysis because with only four stimulus categories per gender, it would be too easy to fit the data with higher order models and the test of the generalization hypothesis would be weakened. Ideally, a test containing several, for example eight or ten, distinct age groups would be preferable. The amount of time needed to administer a test of that length, however, would be prohibitive. In addition, that number of morphologically distinct groups may not exist a priori. According to the Tanner system, there are essentially five.

Lastly, since sexual responding is likely an excitatory rather than inhibitory process, "U" - shaped profiles, characteristic of inhibatory generalization gradients, (see Figure 4d) were not expected to emerge. If they do, it is expected that the F value from that test will be very low.

CHAPTER II

Results

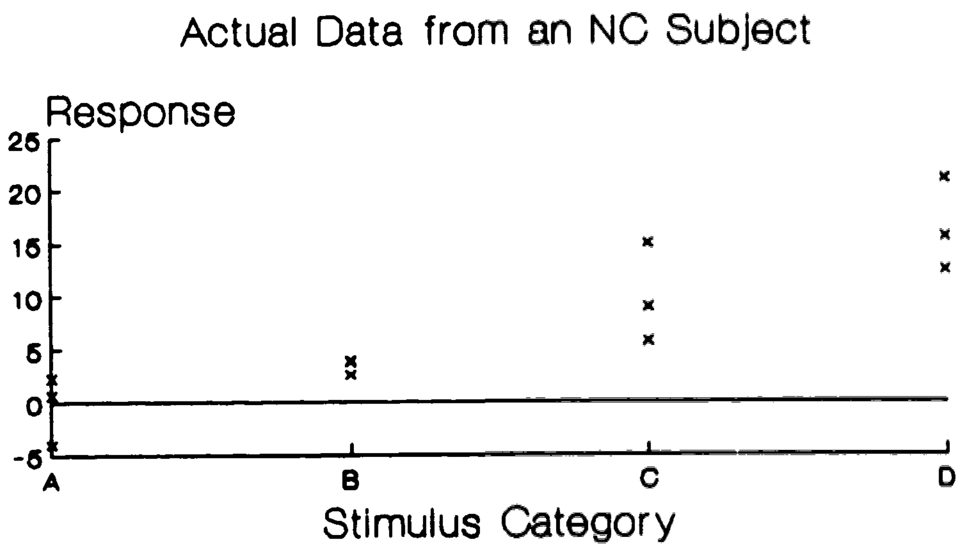
Since EPT referrals are made without regard to variables such as subjects' age and education, one-way analyses of variance were carried out to determine whether or not significant differences existed between the groups in mean age or education. As can be seen in Table 2 below, the results were significant for age, but not for education.

Table 2

Demographic Features of Faking and Non-Faking Subjects

Group	Control	Non-Faking Deviants	Faking Deviants
Measure			
n	10	10	10
Age *			
Mean	28.0	38.3	36.0
S.D.	4.2	8.1	6.8
Education **			
Mean	12.8	11.0	10.8
S.D.	2.3	3.8	2.3

* $F(2,27) = 6.779$, $p = 0.0041$
** N.S.

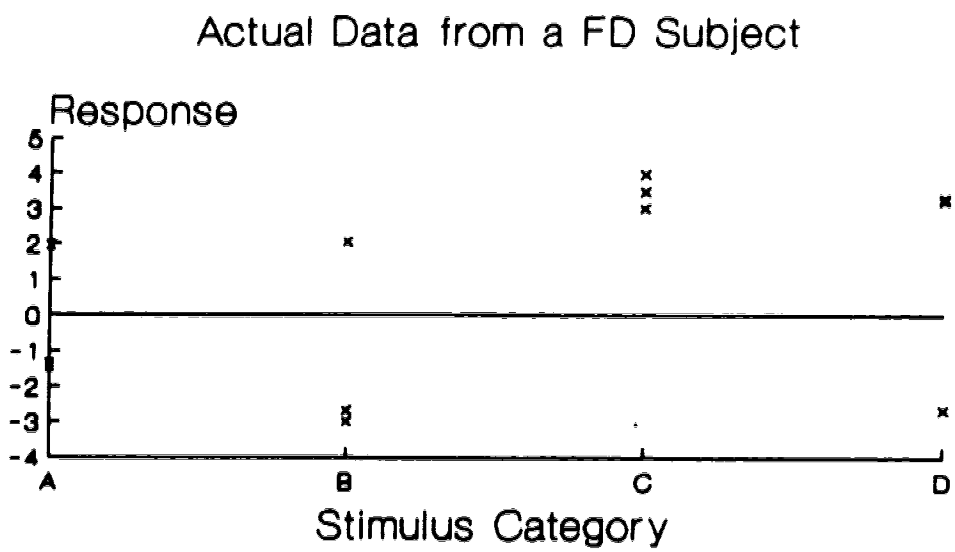


F (Full Model) = 27.338 (P < 0.001)
Lin. Eft Coef. = 0.393 (P < 0.001)
Quad. Eft Coef/X1 = 0.058 (P = 0.478)

Figure 5. Scatterplot of data from a nonfaked test.

Figures 5 and 6 are scatterplots showing actual data supplied by a Normal Control and a Faking Deviant respectively. In the first case, there is a strong, consistent increase in response magnitude corresponding to depictions of more developed females. In addition, the variability among the scores in any given category is fairly small. As noted on the graph, a highly significant *F* statistic followed a regression analysis of these data. In addition, the variance was adequately

explained by the linear effect coefficient alone, as the quadratic effect coefficient (given the presence of the linear component) did not account for a further significant decrease in error variance.



F (Full Model) = 0.523 (P = 0.528)
Lin. Eft Coef. = 0.524 (P = 0.286)
Quad.Eft Coef/X1 = 0.058 (P = 0.776)

Figure 6. Scatterplot of data taken from a faked test.

The Faked data presented in Figure 6 show a much higher degree of variability then those of the previous example. Although there seems to be a slight tendency to evidence larger responses to more developed females, the large degree of scatter introduced to the scores by pumping rendered that trend insignificant. Note that this data was contributed by the same subject whose pumping artefacts were highlighted in Figure 2.

Of the 20 subjects in the two nonfaking groups (NC and D), 15 (75%) showed generalization gradients, indicated by a significant F test of the full regression model. A two-tailed binomial test indicated that this proportion was statistically significant (N = 20, p = .041). By contrast, only 2 of the 10 faking subjects (20%) showed generalization to stimuli corresponding to their preferred gender. This result was not significant.

A Chi-square test, corrected for continuity, indicated that these proportions were also significantly different from each other (Chi-Square(1, N = 30) = 6.126, p <.05). The results of these three analyses are summarized in Table 3 below.

A similar comparison was carried out among the data representing responses to the less preferred gender. Data from the FD group was not included in view of the finding that faking interferes with the emergence of generalization in the profile. Among the 20 nonfaking subjects, the profiles of only three contained generalization as defined in this study. This proportion is significantly <u>less</u> than what would be predicted if this variable was randomly distributed ($N = 20$, $p < .01$). Hence, a pattern of generalization was not a characteristic of subjects' responses to members of the gender opposite that which they preferred.

Table 3

<u>Number of Faked and Non-Faked Profiles with Significant Generalization Gradients</u>

Generalization	Profile		
	Not Faked	Faked	Total
Absent	5	8	13
Present	15	2	17
Total	20	10	30

Chi-Square = 6.126, df = 1, p = 0.0133

The above findings suggest that generalization gradients are a property of erotic preference profiles both in normal and sexually anomalous men, and that subjects who attempt to manipulate the outcome of testing will introduce enough error to the data to obscure this feature of the profile. In other words, a regression analysis of the distorted data will yield an insignificant *F* test. Hence, both hypotheses under investigation were supported by the data.

The prediction that "U"-shaped curves would not emerge was tested by combining the data from all three groups and noting the total number of profiles which showed significant positive quadratic effect coefficients among the data from the preferred gender. This prediction was supported by the observation that only 3 of the thirty profiles contained such configurations.

Discussion

As positive as these results are, it would be wise to exercise caution before incorporating them into any assessment of EPT validity at the individual test level. Subsequent investigations should be conducted to determine whether the generalization criteria is equally effective against different faking strategies. The reason for this may not be apparent, but rests with the fact that some of the FD subjects in the present study utilized pumping in their efforts to fake. Since this results in relatively large artefacts, it is possible that a larger component of error was entered into the regression analysis than would be by other faking strategies. On the other hand these rapid volume changes are characteristically negative in direction (see Figure 2), and therefore may not alter the trial data if scoring is based on maximum displacement rather than area under the curve. If this is the case, pumping probably will not add any more error than other methods of faking.

Another question arises as to how applicable these results are to erotic preference tests concerned with activity rather than partner preference. It is important to recognize the fact that regression analyses require an independent variable which can be readily measured on an interval or ratio scale. In the present case this presented little difficulty since physical maturation proceeds through predictable stages. Also, a convenient and widely accepted scoring structure, the Tanner system, was already available. Without these advantages, applying regression techniques to other tests might be less straightforward. For example, designing a test of activity preference would require that all stimuli be rated in salience along one or more continua, none of which would be as easily defined as was physical maturity. A progression such as: voyeurism, toucherism, rape might represent points along a continuum of aggressiveness, yet it would be very difficult to control for the effects of other factors such as the presence or absence of physical contact with the partner/victim and the degree of partner familiarity, to name only two.

Perhaps the most immediate question concerns the usefulness of the generalization criterion in tests utilizing non-volumetric transducers. As was previously pointed out, penile enlargement proceeds in a non-isotropic manner. Consequently, it will be necessary to determine whether generalization gradients appear in data based on circumferential measures. The fact that circumferential increases at low levels of arousal may reflect either an increase or decrease in penile blood volume virtually assures a more substantial component of error variance in the profiles of subjects attaining only low levels of arousal. This would tend to lower the value of the *F* statistic in a regression analysis. Where the subject is consistently producing responses which are large and positive, however, this should prove less problematic.

Since one goal in conducting the present study was to find a method of reliably detecting faking, some discussion of this application is fitting. Although the results of the main analyses were statistically significant, they must be weighed against the actual success rates of the generalization criteria in distinguishing between faked and non-faked profiles. If this were the only means employed to assess the validity of the test, one quarter of the non-faking subjects (5 out of 20) would have been incorrectly suspected of attempting to manipulate the test results. On the other hand, profiles from 3 of the 10 (30%) known FD subjects would have been considered valid. As an indicator of faking, therefore, one must apply this test with great caution, and should use it only as an adjunct to other methods of assessing validity.

A comment should be made regarding the dependency of the preceding results on the number of subjects in this study. It is important to remember that the non-faking group was comprised of men in both the D and NC control groups which led to a group size of 20. Of the 5 insignificant results, 2 were in the NC group and 3 were in the D group. This left a total of 8 and 7 significant profiles respectively. Neither of these values would have yielded a significant binomial test result had they been considered separately, though comparing either these values against that found in the FD group would produce a significant outcome. The implication is that while a group of faking subjects will certainly produce significantly fewer valid profiles using this criterion, an equally sized groups of honest subjects may also fail to show a significant number. With a sample size of 10, only observed values of 0,1,9 or 10 generalized profiles would have led to a statistically significant binomial test. On the other hand, had a proportionate number of generalized profiles emerged in a group of 20 (i.e. a total of 4) faking subjects, a significant binomial test ($N = 20$, $p < .05$) would have resulted. In this example the test would have indicated a significant absence of generalized responses.

It would have been desirable to utilize a greater number of faking subjects. Since EPT referrals are made for clinical reasons, however, it was not possible to control their availability directly. Future investigations could avoid this problem by including a group of non-referred subjects who were instructed by the examiner to attempt faking. Moreover, specific faking strategies could be systematically employed to determine whether the generalization criterion is more useful against some than others.

Apart from the utility the present results have for assessing test validity, it is important to directly consider the implications of the central finding: that the degree of erotic interest one has in a given age group is not independent of the attraction he feels toward members of that gender who are different in age. One interpretion of this is that true pedophilia is not a discreet disorder. Rather, it is a syndrome which can be expressed in terms of the coefficients defining a regression line fit to a set of data. Where $ß_1$ is negative and significant, there is an overall tendency toward smaller responses as older models are depicted. Where this value is very small, and the quadratic effect coefficient is insignificant, the examinee is failing to discriminate between age groups.

Regarding the value of $ß_{11,}$ which represents the importance of the curvilinear component, its meaning must be interpreted in the context of $ß_1$. In any case, when it is negative, the profile curves upward (as in Figure 4d). In the context of a small $ß_1$ coefficient, this indicates a peak response in the intermediate age groups. Where β_1 is large however, it simply indicates poor linearity among the category scores. As mentioned above, large (significant), positive values appear not to emerge, and may provide an additional clue to faking if present.

The findings suggest that although it is certainly possible for one's responses to be greatest to children, this does not preclude that individual from functioning in the context of an age appropriate relationship. The corollary is even more important: the fact that an individual engages in, perhaps evens prefers, sexual contact with an adult does not constitute evidence of his inability to respond to children.

What factors influence the shape of a generalization gradient is a question well beyond the scope of the present paper yet one must wonder at the large number of sexual assaults against minors that take place while the perpetrator is under the influence of drugs or alcohol. It would be interesting to administer controlled quantities of alcohol to a group of subjects who had been previously tested. In this manner, it would be possible to determine whether the gradient underwent a shift in peak, a flattening, or no change. It is also possible that the curve would remain essentially the same shape but the gradient would become nonsignificant. In this case, one could postulate that the alcohol interfered with the degree of stimulus control involved in sexual responding.

Lastly, it deserves comment that a highly significant number of nonfaking subjects failed to evidence generalization among their responses to individuals of the nonpreferred gender. That such a shift took place in the relative contribution of error variance seems somewhat surprising. At least during the early stages of development, there is more similarity between members of the two genders then there is difference. Hence one might expect comparable measures of generalization even though the coefficients could have drastically different values. This was clearly not the case and residual analyses based on scatterplot inspection did not suggest any difference in the degree of error as a function of stimulus age. It would seem therefore, that sexual responses to individuals opposite in gender to what the examinee prefers are not under stimulus control. There is one other possibility, however, and that is that negative penile responses (which are frequently evidenced upon exposure marginally aversive erotic material) are not as reliable as their positive counterparts. Also, since the range over which negative responses can occur is likely smaller, it may be that a lower degree of measurement precision is available. Either of these factors could be a source of error.

Recommendations

It seems appropriate to include some suggestions which might prove useful in future investigations of the type reported herein. Most of what follows concerns the implementation of procedures intended to provide controls for variables along which subjects will differ.

First, since recency of ejaculation affects the overall responsiveness of subjects being tested, it would be desirable to know when this last occurred. The simplest way to get this information would be to ask the subject before testing. Given that some subjects may find this sort of questioning very embarrassing, however, the answers given to this question would have to be considered with caution.

Where it is possible, it would be worthwhile observing the examinee for several hours (at least) prior to testing. That failing, he could spend that time responding to questionnaires, or doing whatever other activities that might be related to his assessment. The purpose of observing the examinee may not be apparent. Very simply, subjects will sometimes masturbate, or engage in sex, prior to EPT; presumably with the intention of reducing their responsiveness to the test stimuli. Similarly, potential examinees should be discouraged from consuming alcohol or excessive quantities of coffee, as these may adversely affect one's penile responses.

Secondly, the issue of access to pornography must be considered an important one. Examinee's should be asked whether or not they have used erotic material in the recent past since prior use on the part of some individuals would introduce a confounding variable.

Next, it would be appropriate in future studies to vary the sequence of stimulus presentation in order to control for the effects of stimulus arrangement. This may be complicated when using movie film, but digitally-based storage media such as optical disks could deal with the problem with relative ease owing to their random access nature.

In addition to these factors, steps need to be taken to control for the fact that individuals differ on the basis of sexual experience, frequency of alcohol consumption, prior criminal records, and access to potential victims. While these factors are as yet poorly understood, controlling for their influence, if any, is important.

Footnotes

[1]Investigations which purport to measure age preference are actually measuring body shape preference since the latter characteristic is observable, and the former is not. Accordingly, some writers such as Fuller, Barnard, Robbins and Spears (1988) have advocated the use of Tanner-scored stimulus materials in such tests. Tanner scores are numbers used to rate an individuals' overall physical maturity using standard physical markers such as breast development, amount of pubic hair, etc. (Tanner, 1978). Each model is assigned a single number ranging from 1 (indicating the least degree of physical maturity) to 5 (corresponding to an adult level of development).

[2]Changes in the v .ume of blood retained in the <u>corpus cavernosa</u> of the penis are responsible for the observable changes in outer dimensions (Rosen, Shapiro and Schwartz, 1975). Hence, all EPT measures ultimately reflect changes in penile blood volume.

[3]Abel et al. demonstrated that although rapists tend not to discriminate between rapistic stimuli and those describing consensual intercourse, normal controls evidenced significantly higher responses to the latter

category. This finding has led other researchers (e.g. Barbaree, Marshall & Lanthier, 1979) to suggest that a failure to inhibit aggressive impulses in the presence of violent cues is an underlying factor in the commission of rape.

[4]A plethysmograph is a device used to measure changes in blood volume.

[5]Generalization gradients are unimodal only when dealing with excitatory stimuli. In the case of inhibitory stimuli, the gradient inverts to produce a "U"-shaped curve.

References

Abel, G.G., Barlow, D.H., Blanchard, E.B., & Guild, D. (1977). The components of rapists' sexual arousal. Archives of General Psychiatry, 34, 895-903.

Abel, G.G., Blanchard, E.B., & Barlow, D.H. (1980). Measurement of sexual arousal in several paraphilias: The effects of stimulus modality, instructional set and stimulus content on the objective. Behavior Research and Therapy, 19, 25-33.

Avery-Clarke, C.A., & Laws, D.R. (1984). Differential reaction response patterns of sexual child abusers to stimuli describing activities with children. Behavior Therapy, 15, 71-83.

Bancroft, J.H., Jones, H.G., & Pullman, B.R. (1966). A simple transducer for measuring penile erection, with comments on its use in the treatment of sexual disorders. Behaviour Research and Therapy, 9, 119-130.

Bancroft, J., & Mathews, A. (1971). Autonomic correlates of penile erection. Journal of Psychosomatic Research, 15, 154-167.

Barbaree, H.E. (1987). Z-score transformation of erectile responses: Increased error in studies of stimulus control of arousal. Unpublished manuscript.

Barbaree, H.E., Marshall, W.L., & Lanthier, R.D. (1979). Deviant sexual arousal in rapists. Behavior Research and Therapy, 17, 215-222.

Barbaree, H.E., Marshall, W.L., Yates, E., & Lightfoot, L.O. (1982). Alcohol intoxication and deviant sexual arousal in male social drinkers. Behavior Research and Therapy, 21, 365-373.

Barlow, D.H., Leitenberg, H., & Agras, S. (1969). The experimental control of sexual deviation through manipulation of the noxious scene in covert sensitization. Journal of Abnormal Psychology, 74, 596-601.

Barr, R.F. (1973). Responses to erotic stimuli of transsexual and homosexual males. British Journal of Psychiatry, 123, 579-585.

Barr, R.F., Raphael, B., & Hennessey, N. (1974). Apparent heterosexuality in two male patients requesting change-of-sex operation. Archives of Sexual Behavior, 3, 325-330.

Bradford, J.M.W. (1986). The use of a bioimpedance analyzer in the measurement of sexual arousal in male sexual deviants. Canadian Journal of Psychiatry. 31, 44-47.

Crocker, L. M., & Algina, J. (1986). Introduction to Classical and Modern Test Theory. New York: CBS College Publishing.

Earls, C.M., & Jackson, D.R. (1981). The effects of temperature on the mercury-in-rubber strain gauge. Behavioral Assessment, 3, 145-149.

Earls, C.M., Quinsey, V.L., & Castonguay L.G. (1987). A comparison of three methods of scoring penile circumference data. Archives of Sexual Behavior, 16, 493-500.

Frenzel, R.R., & Lang, R.A. (1989). Identifying sexual preferences in intrafamilial and extrafamilial child sexual abusers. Annals of Sex Research, 2, 255-275.

Freund, K., Chan, S., & Coulthard, R. (1977). Phallometric diagnosis with 'nonadmitters'. Behaviour Research and Therapy. 17, 451-457.

Freund, K. (1963). A laboratory method for diagnosing predominance of homo- or hetero- erotic interest in the male. Behavior Research and Therapy, 1, 85-93.

Freund, K. (1967). Diagnosing homo- or heterosexuality and erotic age-preference by means of a psychophysiological test. Behavior Research and Therapy, 5, 209-228.

Freund, K., Langevin, R., Barlow, D. (1974). Comparison of two penile measures of erotic arousal. Behaviour Research and Therapy, 12, 355-359.

Freund, K., Sedlacek, F., & Knob, K. (1965). A simple transducer for mechanical plethysmography of the male genital. Journal of the Experimental Analysis of Behavior, 8, 169-170.

Freund, K., & Blanchard, R. (1981). Assessment of sexual dysfunction and deviation. In M. Hersen & A.S. Bellack (Eds.), Behavioral Assessment (pp. 427-455). New York: Pergammon Press.

Freund, K., Watson, R. & Rienzo, D. (1988). Signs of faking in the phallometric test. Behavior Research and Therapy, 26, 105-112.

Geer, J.H., & Fuhr, R. (1976). Cognitive factors in sexual arousal: The role of distraction. Journal of Clinical and Consulting Psychology, 44, 238-243.

Hanson, H.M. (1959). Effects of discrimination training on stimulus generalization. Journal of Experimental Psychology, 58, 321-334.

Kaine, A., Crim, M., & Mersereau, G. (1988). Faking sexual preference. Canadian Journal of Psychiatry, 33, 379-385.

Kolárský, A., & Madlafousek, J. (1977). Variability of stimulus effect in the course of phallometric testing. Archives of Sexual Behavior, 6, 135-141.

Laws, D.R., & Marshall, W.L. (1990). A conditioning theory of the etiology and maintenance of deviant sexual preference and behavior. In W.L. Marshall & D.R. Laws (Eds.), Handbook of Sexual Assessment (pp. 209-229). New York: Plenum.

Laws, D.R., & Pawlowski, A.V. (1973). A multi-purpose biofeedback device for penile plethysmography. Journal of Behavior Therapy and Experimental Psychiatry, 4, 339-341.

Laws, D.R., & Rubin, H.B. (1969). Instructional control of an autonomic sexual response. Journal of Applied Behavior Analysis, 2, 93-99.

Malcolm, P.B., Davidson, P.R., & Marshall, W.L. (1985). Control of penile tumescence: The effects of arousal level and stimulus content. Behavior Research and Therapy, 23, 273-280.

McConaghy, N. (1967). Penile volume changes to moving pictures of male and female nudes in heterosexual and homosexual males. Behavior Research and Therapy, 5, 43-48.

McConaghy, N. (1974). Measurements of change in penile dimensions. Archives of Sexual Behavior, 4, 381-388.

Murphy, W.D., & Barbaree, H.E. (1987). Assessments of sexual offenders by measures of erectile response: Psychometric properties and decision making. Unpublished manuscript.

Quinsey, V.L., Chaplin, T.C., & Varney, G. (1981). A comparison of rapists' and non-sex offenders' sexual preferences for mutually consenting sex, rape, and physical abuse of women. Behavioral Assessment, 3, 127-135.

Quinsey, V.L., Chaplin, T.C., & Carrigan, W.F. (1979). Sexual preferences among incestuous and nonincestuous child molesters. Behavior Therapy, 10, 562-565.

Quinsey, V.L. & Harris, G. (1976). Comparison of two methods of scoring the penile circumference response: magnitude and area. Behavior Therapy, 7, 702-704.

Quinsey, V.L., Steinman, C.M., Bergersen, S.G., & Holmes, T.F. (1975). Penile circumference, skin conductance, and ranking responses of child molesters and "normals" to sexual and nonsexual visual stimuli. Behavior Therapy, 6, 213-219.

Richards, J.C., Kalucy, R.S., Wood, M.M., & Marshall, V.R. (1990). Linearity of the electromechanical penile plethysmograph's output at large expansions. The Journal of Sex Research, 27, 283-287.

Rosen, R.C., & Keefe, F.J. (1978). The measurement of human penile tumescence. Psychophysiology, 15, 366-376.

Rosen, R.C., Shapiro, D., & Schwartz, G.E. (1975). Voluntary control of penile tumescence. Psychosomatic Medicine, 37, 479-483.

Tanner, J.M. (1978). Foetus into Man: Physical growth from Conception to Maturity, Harvard University Press, Cambridge, MA.

Wheeler, D., & Rubin, H.B. (1987). A comparison of volumetric and circumferential measures of penile erection. Archives of Sexual Behavior, 16, 289-299.

Yates, E., Barbaree, H.E., & Marshall, W.L. (1984). Anger and deviant sexual arousal. Behavior Therapy, 15, 287-294.

Zuckerman, M. (1971). Physiological measures of sexual arousal in the human. Psychological Bulletin. 297-329.

APPENDIX I

This appendix contains the random subject selection program. Although it was written in Turbo Pascal, it should run under most other Pascal language compilers with minimal modification.

```
Program Choose_Subjects; { Listing starts here }

{$N+}    { Enables the 80X87 math coprocessor. The
directive must be switched to $N- if no coprocessor is
installed }

Uses
   DOS, Crt;  { These are standard Turbo Pascal unit
                libraries }

Type
   Subject_Status          =   (Taken, Not_Taken);
Var
   n, Choice, Draw, Pool   :   Byte;
   Subject_Pool            :   Array[1..255] of
                               Subject_Status;
                               { Records whether or not
                                 a subject has already
                                 been chosen }

Procedure Shake_the_Dice;
{ Initializes the random seed }
Begin
   Write('Press  ENTER <--  to begin:');
   ReadLn;
   Randomize;
End;

Procedure Choose_a_Subject; { Returns a randomly selected
                              subject and checks if that
                              subjects has already been
                              picked. If so, another is
                              chosen }
Begin
   Choice := Random(Pool);
   If (Subject_Pool[Choice] = Taken) then
      Choose_a_Subject;
End;
```

```
BEGIN { Main body of Program }
   ClrScr;
   FillChar(Subject_Pool, SizeOf(Subject_Pool),
            Not_Taken);
   Write('Size of pool: ');
   ReadLn(Pool);
   Write('Number to draw: ');
   ReadLn(Draw);
   WriteLn;
   Shake_The_Dice;
   For n:= 1 to Draw do Begin
      Choose_a_Subject;
      Subject_Pool[Choice] := Taken;
   End;
   WriteLn('The following subjects were chosen:');
   For n:= 1 to Pool do
      If (Subject_Pool[n]) = Taken then
         WriteLn(n);
END. { Of program Choose_Subjects }
```

APPENDIX II

The following is the main data analysis program.

```
Program Thesis_Data_Analysis;

{$N+}   { Enables the 80X87 math coprocessor, switch
          to $N- if none is installed }

Uses
   DOS, Crt, Common, LeastSqr, FTest;

   { These are standard libraries, like IMSL which
     contain various routines }


Const
   XValues               : Array [1..12] of Real =
                 (1,1,1,2.5,2.5,2.5,3.5,3.5,3.5,5,5,5);


  { These are the Tanner scores which comprise the
    independent variable }

Type
   MatrixRecordPointer = ^MatrixRecord;
   MatrixRecord        =  Record
                            Elements : Integer;
                            Matrix   : TNColumnVector;
                          End;

  { This data type contains a vector as well as
    another variable which records its length. A
    MatrixRecordPointer can later be passed to a user
    defined procedure or function with compatible
    input parameters }


Var
   InFile, SummaryFile : Text;
   MaleData, FemaleData, Block1,
   Block2, Block3, Z1, Z2, Z3,
   YFit, Residuals :  MatrixRecordPointer;
   Coefficients    : Array[1..3] of TNRowVector;
   GN, SN                             : Integer;
   XData                              : TNColumnVector;
```

```
    Hour, Min, Sec, Sec100          : Word;
    Hourf, Minf, Secf, Sec100f      : Word;

 Procedure Initialize_Arrays;
 { Dynamically allocates memory space
   for the variable listed below }

 Var
    m : Integer;
 Begin
    New(MaleData);
    New(FemaleData);
    New(Block1);
    New(Block2);
    New(Block3);
    New(Z1);
    New(Z2);
    New(Z3);
    New(YFit);
    New(Residuals);
    For m := 1 to 12 do
       XData[m] := XValues[m];
 End;

 Procedure Clear_Array_Space;

       { Deallocates RAM when data structure
         is no longer needed }
 Begin
    Dispose(MaleData);
    Dispose(FemaleData);
    Dispose(Block1);
    Dispose(Block2);
    Dispose(Block3);
    Dispose(Z1);
    Dispose(Z2);
    Dispose(Z3);
    Dispose(YFit);
    Dispose(Residuals);
 End;

 Procedure Open_Data_File;
 { Opens the file outfile1.txt which contains the
   raw data }
 Begin
```

```
   Assign(InFile,'outfile1.txt');
   Reset(InFile);
End;

Procedure Open_Summary_File; { Opens the file
                                summary.out
                                on the current
                                directory }

Begin
   Assign(SummaryFile, 'summary.out');
   Rewrite(SummaryFile);
End;

Procedure Get_Observation; { Reads data from f.le
                             outfile1.txt }
Var
   FileLine        : String[23];
   i, j, n, z,
   code            : Integer;
Begin
   If (not EOF(InFile))  then Begin { Ensure that the
                                      end of file
                                      marker hasn't
                                      been encountered
                                    }
      ReadLn(InFile, FileLine);
      If (Copy(FileLine,2,1) = ' ') then
         Val(Copy(FileLine,1,1), GN, Code)
      else
         Val(Copy(FileLine,1,2), GN, Code);
      If (Copy(FileLine,6,1) = ' ') then
         Val(Copy(FileLine,5,1), SN, Code)
      else
         Val(Copy(FileLine,5,2), SN, Code);
      n := 1;
      For n := 1 to 9 do Begin
         ReadLn(InFile, FileLine);
         Val(Copy(FileLine, 1,7), Block1^.Matrix[n],
              Code);
         Val(Copy(FileLine, 8,8), Block2^.Matrix[n],
              Code);
         Val(Copy(FileLine,16,8), Block3^.Matrix[n],
              Code);
      End;
      Block1^.Elements := 9; Block2^.Elements := 9;
```

```
      Block3^.Elements := 9;
      Z1 := Block1; Z2 := Block2; Z3 := Block3;
   End;
End;

Procedure   Standardize(InArray,   OutArray   :
MatrixRecordPointer);

{ This procedure converts raw data into standarized data
}


Var
   n                                    : Integer;
   divisor, mean, SS, StandardDeviation : Real;

Begin { Of procdure standardize }
   mean := 0;
   divisor := 0;
   For n := 1 to InArray^.Elements do Begin
      divisor := divisor + 1;
      mean := mean + InArray^.Matrix[n];
   End;
   mean := mean / divisor;

   SS := 0;
   For n := 1 to InArray^.Elements do
      SS := SS + Sqr(InArray^.Matrix[n] - mean);

   StandardDeviation := Sqrt(SS / (divisor - 1));

   For n:= 1 to InArray^.Elements do
      OutArray^.Matrix[n] := (InArray^.Matrix[n] - mean)/
                              StandardDeviation;
   OutArray^.Elements := InArray^.Elements;
End;

Procedure Arrange_Data_by_Stimulus_Gender;

{ Separates data gathered during presentation of male and
female stimuli }

Var
   n, increment : Integer;
Begin
   n := 1;
```

```
   increment := 1;
   While (n <= 10) do Begin
      FemaleData^.Matrix[n] :=
      Z1^.Matrix[increment];
      FemaleData^.Matrix[n + 1] :=
      Z2^.Matrix[increment];
      FemaleData^.Matrix[n + 2] :=
      Z3^.Matrix[increment];
      MaleData^.Matrix[n] :=
      Z1^.Matrix[increment  + 5];
      MaleData^.Matrix[n + 1]   :=
      Z2^.Matrix[increment + 5];
      MaleData^.Matrix[n + 2] :=
      Z3^.Matrix[increment + 5];
      increment := increment + 1;
      n := n + 3;
   End;
   FemaleData^.Elements := 12;
   MaleData^.Elements   := 12;
End;


Procedure Report_Values(DataArray :
                        MatrixRecordPointer);

{ Prints values held in record pointed to by passed
variable }

Var
   n : Integer;
Begin
   For n := 1 to DataArray^.Elements do
      WriteLn(XData[n]:7, DataArray^.Matrix[n]:10:2);
End;


Procedure   Perform_Regression(InArray                    :
MatrixRecordPointer;
                               GenderString : String);

{ It is this routine which actually does the regression
analysis }

Const
   Fit                                  = Poly;
```

```
Var
   StandardError, Count, YMean, SSTO   : Float;
   SSR, SSE, MSR, MSE, F, P            : Array[1..3] of
                                         Float;
   Error, NumTerms                     : Byte;
   m                                   : Integer;

Begin
   YMean := 0;
   Count := 0;
   For m := 1 to InArray^.Elements do Begin
      Count := Count + 1;
      YMean := YMean + InArray^.Matrix[m];
   End;
   YMean := YMean / Count;

   For NumTerms := 2 to 3 do Begin
      LeastSquares(InArray^.Elements, XData,
                   InArray^.Matrix,
                   NumTerms, Coefficients[NumTerms],
                   YFit^.Matrix,
                   Residuals^.Matrix, StandardError,
                   SSR[NumTerms],
                   Error, Fit);
      MSE[NumTerms] := Sqr(StandardError);
      SSE[NumTerms] := MSE[NumTerms] *
         (InArray^.Elements -  NumTerms);
      SSR[NumTerms] := 0;
      For m:=1 to InArray^.Elements do
         SSR[NumTerms] := SSR[NumTerms] +
                          Sqr(YFit^.Matrix[m] -
                          YMean);
      MSR[NumTerms] := SSR[NumTerms] / (NumTerms -1);
      F[NumTerms] := MSR[NumTerms]/MSE[NumTerms];
      P[NumTerms] := SigF(F[NumTerms],(NumTerms - 1),
                     (InArray^.Elements - NumTerms));
   End;

   NumTerms := 3; { Number of Beta coeeficients to
                    determine,
                    including intercept }

   SSTO := SSE[3] + SSR[3];

   SSR[1] := SSTO - SSR[2] - SSE[3];
```

```
   F[2] := MSR[2] / MSE[3];
   P[2] := SigF(F[2], 1, 9);

   MSR[1] := SSR[1];
   MSE[1] := MSE[3];
   F[1]   := MSR[1] / MSE[1];
   P[1]   := SigF(F[1], 1, (InArray^.Elements -
                 NumTerms));

   NumTerms := 3;

   WriteLn('********************************');
   WriteLn('Coefficients:                       Gn:
            ',GN:3,'  Sn:
            ',SN:3,
            '   Gender: ',GenderString);
   WriteLn('********************************');
   For m:=1 to NumTerms do
      WriteLn('Beta: ',pred(m),Coefficients[3,m]:17:4);
   WriteLn;

WriteLn('*************************************************
******************************          ');
   WriteLn('Source of Variance     Sum of Squares    df
  Mean Square     F      P              value');

WriteLn('*************************************************
******************************          ');
   WriteLn('Regression               ', SSR[NumTerms]:12:4,
(NumTerms - 1):8 , MSR[NumTerms]:14:4,
            F[NumTerms]:10:4, P[NumTerms]:10:4);
   WriteLn('Error                     ', SSE[3]:12:4,'
',(InArray^.Elements - NumTerms):2 ,MSE[3]:14:4);

WriteLn('*************************************************
******************************          ');
   WriteLn('Test of quadratic effect:');

WriteLn('*************************************************
******************************          ');
   WriteLn('X1                       ', SSR[2]:15:4, '1':8,
SSR[2]:14:4, F[2]:10:4,P[2]:10:4);
   WriteLn('X2 / X1                  ', SSR[1]:15:4, '1':8,
SSR[1]:14:4, F[1]:10:4,P[1]:10:4);
   WriteLn('Error                     ', SSE[3]:12:4,'
',(InArray^.Elements - NumTerms):2 ,MSE[3]:14:4);
```

```
WriteLn('**************************************************
****************************          ');
   WriteLn;
   WriteLn('X1 = Tanner score');
   WriteLn('X2 = Tanner score squared');
   WriteLn('Y  = volumetric displacement in ml.');
   Write(SummaryFile, GN:3, SN:4,GenderString:8);
   For m := 1 to NumTerms do
      Write(SummaryFile, Coefficients[3,m]:8:3);
   Write(SummaryFile, P[2]:8:3, P[1]:8:3);
   WriteLn(SummaryFile, F[3]:8:3, P[3]:8:3);
End;


BEGIN { Main body of program }
   GetTime(Hour, Min, Sec, Sec100);
   Write('Execution initiated: ',Hour:2, Min:2, Sec:2,
          Sec100:2);
   Initialize_Arrays;
   Open_Data_File;
   Open_Summary_File;


   While (not EOF(InFile)) do Begin
      Get_Observation;
      If (GN in [1..21]) then Begin
         Standardize(Block1, Z1);
         Standardize(Block2, Z2);
         Standardize(Block3, Z3);
         Arrange_Data_by_Stimulus_Gender;
         Perform_Regression(FemaleData, 'FEMALES');
         ClrScr;
         Perform_Regression(MaleData, 'MALES');
         WriteLn(SummaryFile);
         ReadLn;
      End;
   End;
   Close(InFile);
   Close(SummaryFile);
   Clear_Array_Space;
   GetTime(Hourf, Minf, Secf, Sec100f);
   Write('Execution complete: ',Hourf:2, Minf:2, Secf:2,
Sec100f:2);
END.
```

APPENDIX III

The following is sample of the output generated by the program in Appendix II. A page such as this was produced for each subject. Summary data are presented in the Results section.

```
Gn:   1  Sn:  11   Gender: FEMALES
**************************
Coefficients:
**************************
Beta: 0          -1.5564
Beta: 1           0.9917
Beta: 2          -0.0893
************************************************************
Source  Sum of Squares  df  Mean Square     F      P value
************************************************************
Regression   5.6340     2     2.8170     17.7391   0.0008
Error        1.4292     9     0.1588
************************************************************
Test of quadratic effect:
************************************************************
X1           5.2974     1     5.2974     33.3584   0.0003
X2 / X1      0.3366     1     0.3366      2.1197   0.1794
Error        1.4292     9     0.1588
************************************************************
Gn:   1  Sn:  11   Gender: MALES
**************************
Coefficients:
**************************
Beta: 0          -2.1226
Beta: 1           1.8470
Beta: 2          -0.3430
************************************************************
Source  Sum of Squares  df  Mean Square     F      P value
************************************************************
Regression   6.1031     2     3.0515      5.2772   0.0304
Error                  5.2042              9       0.5782
************************************************************
Test of quadratic effect:
************************************************************
X1           1.1385     1     1.1385      1.9688   0.1941
X2 / X1      4.9646     1     4.9646      8.5856   0.0168
Error        5.2042     9     0.5782
************************************************************
X1 = Tanner score
```

X2 = Tanner score squared
Y = volumetric displacement in cc.