FACIAL ACTION DETERMINANTS

OF

PAIN JUDGMENT

by

DOUGLAS SPENCER LEE

B.A., The University of British Columbia, 1974

M.Ed., The University of British Columbia, 1978

M.A., The University of British Columbia, 1983

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

PSYCHOLOGY

We accept this dissertation as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

October 1985

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Psychology

The University of British Columbia
1956 Main Mall
Vancouver, Canada
V6T 1Y3

Date Oct 16 1985

DE-6 (3/81)

ABSTRACT

Nonverbal indices of pain are some of the least researched sources of data for assessing pain. The extensive literature on the communicative functions of nonverbal facial expressions suggests that there is potentially much information to be gained in studying facial expressions associated with pain. Results from two studies support the position that facial expressions related to pain may indeed be a source of information for pain assessment.

A review of the literature found several studies indicating that judges could make discriminations amongst levels of discomfort from viewing a person's facial expressions. Other studies found that the occurrence of a small set of facial movements could be used to discriminate amongst several levels of self-reported discomfort. However, there was no research directly addressing the question of whether judges ratings would vary in response to different patterns of the identified facial movements. Issues regarding the facial cues used by naive judges in making ratings of another person's discomfort were investigated.

Four hypotheses were developed. From prior research using the Facial Action Coding System (FACS) (Ekman & Friesen, 1978) a small set of facial muscle movements, termed Action Units (AUs), were found to be the best facial movements for discriminating amongst different levels of pain. The first hypothesis was that increasing the number of AUs per expression would lead to increased ratings of

discomfort. The second hypothesis was that video segments with the AUs portrayed simultaneously would be rated higher than segments with the same AUs portrayed in a sequential configuration. Four encoders portrayed all configurations. The configurations were randomly editted onto video tape and presented to the judges. The judges used the scale of affective discomfort developed by Gracely, McGrath, and Dubner (1978). Twenty-five male and 25 female university students volunteered as judges.

The results supported both hypotheses. Increasing the number of AUs per expression led to a sharp rise in judges' ratings. Video segments of overlapping AU configurations were rated higher than segments with non-overlapping configurations. Female judges always rated higher than male judges.

The second study was methodologically similar to the first study. The major hypothesis was that expressions with only upper face AUs would be rated as more often indicating attempts to hide an expression than lower face expressions. This study contained a subset of expressions that were identical to ones used in the first study. This allowed for testing of the fourth hypothesis which stated that the ratings of this subset of expressions would differ between the studies due to the differences in the judgment conditions.

Both hypotheses were again supported. Upper face expressions were more often judged as portraying attempts by the encoders to hide their expressions. Analysis of the fourth hypothesis revealed that the expressions were rated higher in study 2 than study 1. A sex of judge X judgment condition interaction indicated that females rated higher in study 1 but males rated higher in study 2.

The results from these studies indicated that the nonverbal communication of facial expressions of pain was defined by a number of parameters which led judges to alter their ratings depending on the parameters of the facial expressions being viewed. While studies of the micro-behavioral aspects of facial expressions are new, the present studies suggest that such research is integral to understanding the complex communication functions of nonverbal facial expressions.

TABLE OF CONTENTS

page

ABSTRACT.. ii

LIST OF TABLES.................................. ix

LIST OF FIGURES................................. x

ACKNOWLEDGEMENTS................................ xi

INTRODUCTION.................................... 1

General Types of Pain Assessment.................. 2

Rationale for Studying Facial Expressions of Pain. 4

Rationales for Conducting Judgment Studies of Facial Expressions........................... 6

Defined Lines of Research......................... 7

Pain Expression vs. Emotional Expression.......... 8

The Concept of Facial Behavior.................... 10

Summary of Rationale for the Present Research..... 11

LITERATURE REVIEW............................... 13

Model for Conducting Judgment Studies of Pain..... 13

Classification of Empirical Studies of Facial Expressions of Pain............................. 16

Summary of A - B Findings......................... 24

Limitations of A - B Research..................... 25

Summary of A - C Findings......................... 26

Limitations of A - C Findings..................... 29

Summary of B - C Findings......................... 31

Limitations of B - C Findings..................... 32

General Summary of Reviewed Research.............. 32

Evaluation of Rosenthal's Model.................. 34

Strengths.................................... 34

Model relationships not yet addressed.......... 36

Weaknesses................................... 36

Conclusions from evaluating Rosenthal's model... 37

HYPOTHESES... 39

Hypothesis 1 - Impact of the Complexity of AU Configurations............................. 40

Hypothesis 2 - Overlapping versus Non-Overlapping AUs....................................... 42

Hypothesis 3 - Hidden vs Non-Hidden AU Configurations of Pain....................... 43

Hypothesis 4 - Effects of Judgment Context on Judges' Ratings............................. 45

GENERAL METHODS.................................... 47

Construction of Video Segments.................. 47

Encoders..................................... 48

Encoder Training.............................. 48

Rules for Selecting Video Segments.............. 49

Rules for equating posed with spontaneous expressions.............................. 49

Rules for eliminating movement artifacts........ 50

Rule for standardizing segment context.......... 50

How rules were applied....................... 51

Equipment.................................... 51

Reliability of Encoders' AU Configurations....... 51

STUDY 1... 54

Method....................................... 54

Subjects.. 54
Judges.. 54
Encoders.. 55
Procedures for Developing Stimulus Tape.......... 55
Selection of AU configurations to be tested..... 55
Construction of stimulus tape................ 57
Procedures for Judges.......................... 60
Measures.. 61
Design.. 62
Results.. 62
Reliability of Judges.......................... 62
Post-Experimental Questionnaire Data............ 64
Analysis of Hypothesis 1...................... 66
Analysis of Hypothesis 2...................... 69
Summary.. 77
STUDY 2.. 80
Method.. 80
Subjects.. 80
Judges.. 80
Encoders.. 80
Procedures for Developing Stimulus Tape.......... 80
Selection of AU configurations to be tested..... 80
Construction of stimulus tape................ 81
Procedures for Judges.......................... 81
Measures.. 83
Design.. 83
Results.. 84

Reliability of Judges.............................. 84

Post-Experimental Questionnaire Data.............. 85

Analysis of Hypothesis 3.......................... 85

Analysis of Hypothesis 4.......................... 88

Summary.. 93

GENERAL DISCUSSION................................. 94

Level of Conplexity............................... 94

Sex of Judge...................................... 95

Type of Configuration............................. 98

Hidden vs Not Hidden Facial Configurations........ 99

Placing the Present Research within the

Judgment Model................................... 101

Facial Activity as a Nonverbal Measure of Pain.... 103

Limitations of the Present Studies................ 104

Conclusions....................................... 106

REFERENCES... 107

Appendix A.. 121

Appendix B.. 131

Appendix C.. 135

Appendix D.. 138

Appendix E.. 141

Appendix F.. 143

Appendix G.. 145

Appendix H.. 147

Appendix I.. 149

Appendix J.. 151

Appendix K.. 153

Appendix L.. 155

Appendix M.. 157

Appendix N.. 159

Appendix O.. 161

Appendix P.. 165

Appendix Q.. 171

Appendix R.. 174

Appendix S.. 176

Appendix T.. 178

List of Tables

Table		page
1	Summary of Empirical Research on Judgments of Facial Expressions of Pain Categorized using Rosenthal's Model	17
2	Possible AU Configurations	56
3	AU Configurations for Hypothesis 1	58
4	AU Configurations for Hypothesis 2	59
5	Study 1 - Analysis of Post-Experimental Questionnaire Data	65
6	Mean Gracely Ratings for Each Level of AU Configuration by Sex of Judge	67
7	Mean Gracely Ratings for Type of Facial Configuration across Levels of Complexity for each Sex of Judge	71
8	Upper and Lower Face AU Configurations	82
9	Study 2 - Analysis of Post-Experimental Questionnaire Data	86
10	Mean Proportion of Times Configurations of the Upper and Lower Face were Judged as Attempts to Hide	87
11	AU Configurations Common to Studies 1 and 2	90
12	Mean Gracely Ratings for each Level of Sex of Judge and Judgment Condition	91

List of Figures

Figure		page
1	Rosenthal's Model for Conducting Judgment Studies	15
2	Interaction of Sex of Judge and Levels of Complexity	68
3	Distribution of Overlapping AU Configurations within Levels of Complexity	70
4	Interaction of Type of Configuration and Level of Complexity	73
5	Interaction of Sex of Judge and Type of Configuration	74
6	Interaction of Sex of Judge and Level of Complexity	75
7	Interaction of Judgment Condition and Sex of Judge	92

ACKNOWLEDGEMENTS

I am indebted to my supervisor, Ken Craig, for five years of support and unrelenting pushing to do a better job. Your encouragement to do better has contributed greatly to the development and refinement of my research skills.

Without the nimble faces of Andy Gotoviec, Chris Patrick, Delphin Schwalm, and Cara Zaskow, and Brenda Gerhardt to tell the FACS about their faces this research would never have happened. Thank you all.

A note to my mother and father who must have wondered at times what I would be when I grew up. Wonder no more, I think I'll be a doctor.

My eternal love to my family Pat, Carol, Dianne, and Monica. Now that school is out what shall we do?

INTRODUCTION

Pain is not only an almost universal experience but data indicates its impact on western society is staggering. Estimates of various pain conditions indicate there are millions of sufferers (Bonica, 1983; Loeser, 1980; Paulley & Haskell, 1975; Main & Waddell, 1982). Bresler (1979) found that up to 80% of all patients who consulted physicians did so for pain related problems. Treatment of pain sufferers costs billions of dollars annually (Brena, 1978; Loeser, 1980; Main & Waddell, 1982). In addition to the societal costs recent work suggests that it is common for families of pain sufferers to be adversely affected (Block, 1981; Block & Boyer, 1984; Mohamed, Weisz & Waring, 1978).

In recognizing the enormity of the problem it must also be recognized that we lack understanding of many facets of the pain experience. How well individual sufferers, their families, researchers, and society can learn to understand and control pain will depend in large part on how accurately pain can be assessed. As stated by Melzack (1983) "The measurement of pain in man is essential for the study of pain mechanisms and for the evaluation of methods to control pain" (p.1). Recognizing that the task of assessing the nature and severity of pain is ultimately tied to interpersonal judgment, the present studies addressed a number of important questions focusing on how observers use nonverbal facial expressions when judging another person's pain.

General Types of Pain Assessment

In keeping with a multifaceted concept of pain (Melzack & Wall, 1965, 1982) four general types of assessment are available for measuring pain. These methods are self-report (including written and verbal reports), observation of tissue damage, physiological measurement, and observation of nonverbal behavior. Chapman et al. (1985) overviews the self-report, physiological, and observation of nonverbal behavior methods of measuring pain. Tissue damage and insult can be assessed by a wide range of medical diagnostic and assessment techniques. Physiological measures of pain utilize autonomic response (e.g. Block, 1981), electromyographic measurement (Vaughan & Lanzetta, 1980) and electroencephalographic measurement (Blanchard, 1981; Chudler & Dong, 1983). The McGill Pain Questionnaire (MPQ) (Melzack, 1975) and scales for rating the affective and sensory qualities of pain developed by Gracely, McGrath, and Dubner (1978, 1979) are among the best self-report measures and both have proven to be useful measures of pain (Gracely, 1983; Melzack, 1983). However, until recently assessment procedures of nonverbal pain behavior have either been isolated to research measures unique to single studies (Ganchrow, Steiner, Kleiner & Edelstein, 1978; Leventhal & Sharp, 1965), or have been so recently developed that the measure has not yet gained wide acceptance (see Keefe and Block, 1982 for a global coding system and Craig & Patrick, 1985 for specific facial expressions of pain).

While assessment of pain via self-report, physiological measures, and tissue damage are better developed than nonverbal measures of pain no method is without problems. Distortion and response bias of self-report measures are well known (Craig & Prkachin, 1980; Kazdin, 1980). Also to be recognized is the fact that certain segments of the population (eg. children, adults who are illiterate or lack facility with the test language, or retarded individuals) do not have the requisite skills to properly complete a self-report assessment (Craig, 1980). Assessment of pain by amount of tissue damage has been shown to be a poor indicator of pain (Beecher, 1959; Loeser, 1980; Stewart, 1977; Melzack & Wall, 1982). In addition, physiological measures are often reactive and intrusive (Ekman, 1982; Kazdin, 1980), have ambiguous meanings, display large individual differences, and generally have low intercorrelations among measures (Cacioppo, Marshall-Goodell & Gormenzano, 1983). These problems often make interpretation of physiological data difficult. Development of nonverbal measures of pain will not only widen the scope of assessment but possibly will also provide measures which will counter some of the problems inherent in the other types of assessment.

As indicated by Harper, Wiens, and Matarazzo (1978) studies of nonverbal behavior suggest that people use a wide range of nonverbal cues for a variety of interpersonal judgments. From anecdotal accounts (Turk, Meichenbaum & Genest, 1981; Fordyce, 1976) and clinical research (Chambers & Price, 1967; Jacox, 1980; Johnson, 1977; Teske, Daut &

Cleeland, 1983) it is clear that the use of nonverbal cues for assessing and describing pain is common and seen as an important source of information (Craig & Prkachin, 1983). It is as yet unclear, however, what types of nonverbal behavior are used in judging pain. For instance, even though it has been demonstrated that observers of facial expressions of pain can accurately judge the occurrence and intensity of pain (Gubar, 1966; Lanzetta & Kleck, 1970; Patrick, Craig, & Prkachin, in press) it is as yet unknown what features of facial expressions are used by the observers in making their judgments. The present studies have attempted to clarify this question.

Rationale for Studying Facial Expressions of Pain

Attempts to describe nonverbal pain behavior, particularly facial expressions, can be traced back to Darwin (1872). However, due to the only recent development of adequate measuring techniques (Ekman & Freisen, 1978; Izard, 1979) and the complexity of facial expressions (Ekman, 1982), work on facial expressions of pain is in its infancy.

The decision to focus the present studies on facial expressions of pain was based on three considerations. First, and foremost, the question of what facial expressions are used by judges when rating a person's level of pain has not received any direct experimental study. The use of a recently developed micro-analytic measurement method for scoring facial expressions (Ekman & Friesen, 1978) allowed the present studies to empirically examine the question of what facial expressions a judge uses when rating qualities of pain of another person. The utility of the Facial Action Coding System (FACS) (Ekman & Friesen, 1978) as a source of dependent variables which discriminate among different pain intensities has already been demonstrated (Craig & Patrick, 1985; Patrick et al., in press).

Second, the face has been identifed as a potentially rich and important source of information for assessing pain (Craig & Prkachin, 1983; LeResche & Dworkin, 1984; Polunin, 1977). Although research has been conducted which indicated differential effectiveness of the face, body movements, and speech qualities as indicators of emotion (Cline, Atzel & Holmes, 1972; Dimatteo, Friedman & Taranta, 1979; Ekman, Friesen & Scherer, 1976), the more recent work by Ekman, Friesen, O'Sullivan, and Scherer (198Ø), Ekman and Oster (1982), and Scherer and Ekman (1982) argues that much of the debate regarding differential effectiveness of nonverbal modes is due to limitations of research methodology and the situational specificity of most nonverbal behavior. It is suggested that more information on each mode of nonverbal behavior, using more satisfactory research methods, is needed before studying the possible differential effectiveness of nonverbal modes and their possible interactions in specific situations. The present research has attempted to supply more adequate information on facial expressions of pain.

The final reason for concentrating on facial expressions of pain is concerned with providing a theoretical framework for interpreting the past and present research within an integrative structure. Prior research on facial expressions of pain has been fragmented with no overriding thrust to the research. The use of a general model for conducting judgment studies (Rosenthal, 1982) was used to allow for an integration of prior research with the present studies to provide a framework for guiding further research.

<u>Rationales for Conducting Judgment Studies of Facial Expressions</u>

The three rationales that have been offered by others for studying observer judgments of facial expressions are that the expressions serve a communicative function, that they reflect important psychological phenomena, and they may have clinical utility as a measure of treatment impact. While these rationales share certain commonalities they need to be differentiated as they each have unique features which may significantly affect the research based on the particular rationale.

The assumption that facial expressions serve a communicative function is the most common rationale (Craig, 1978, Ekman, 1965; Keefe, Wilkins & Cook, 1984; Lanzetta & Kleck, 1970; Munn, 1940). This assumption is based on impressions of everday experiences in which people appear to infer the subjective status of other people from their nonverbal behavior, most commonly their facial expressions. The second rationale suggests that measurements of facial expressions may be used as dependent variables in order to augment data from other measures (Craig & Patrick, 1985;

Ekman & Friesen, 1969; LeResche, 1982). This rationale is predicated on the assumption that facial expressions are less prone to voluntary manipulation by the subject than either other nonverbal behavior (Ekman, 1982) or self-report (Craig & Prkachin, 1983). The third rationale is related to the second but focuses on a different need. Several studies have suggested that information gained from measuring facial expressions may prove to have valuable clinical utility, particularly as a measure of intervention impact and as a source of feedback for the therapist (Block, 1981; Ekman & Friesen, 1968; Ganchrow et al., 1978). The present research has focused on the communicative function of facial expressions.

Defined Lines of Research

Allied to the different rationales used for studying judgments of facial expressions are the types of research done in this area. The history of judgment studies of facial expressions can be traced along four lines of investigation. The first line has dealt with the question of whether observers could make accurate judgments of emotions or pain based on observing facial expressions (Boucher, 1969; Ekman, 1972; Izard, 1969; Langfeld, 1918; Lanzetta, Cartwright-Smith & Kleck, 1976; Lanzetta & Kleck, 1970; Patrick et al., in press). The second line traces research that has attempted to devise measurement tools for coding facial expressions (Birdwhistell, 1952; Ekman & Friesen, 1978; Izard, 1979). The third line, which is related to the second, focused on questions concerned with what areas of the face observers use when making judgments (Dunlap, 1927; Ekman, Friesen & Tomkins, 1971; Hanawalt, 1944). The final line of research, which has only recently been developed, has focused on physiological measures of facial musculature to investigate the role of unobservable facial muscle activity in experiencing various emotions (Ekman, Levenson & Friesen, 1983; Schwartz, Brown & Aher, 1980; Schwarz, Fair, Salt, Mandel & Klerman, 1976; Vaughan & Lanzetta, 1980).

When combined these four lines of research represent a broadband research front attacking the many questions asked regarding the utility and role of nonverbal facial behavior as these behaviors relate to the expression and judgment of pain and emotion. The present studies were based mainly on the first and third lines of research.

Pain Expression vs. Emotional Expression

Craig and Patrick (1985), Patrick et al. (in press), LeResche (1982), and Hyde (1985) have shown that FACS, which was developed to study facial expressions of emotion, can be used to study facial expressions of pain. Ganchrow et al. (1978), in studying judgments of pain responses, used a simple coding system reminiscent of the Facial Affect Scoring Technique developed by Ekman et al. (1971). However, the adaptability of methods for studying emotional expression to studying pain expression must not hide the theoretical differences between the concept of affect or emotion (Izard, 1971; Tomkins, 1982) and pain (IASP, 1979; Leventhal & Everhart, 1979; Melzack & Wall, 1965; Weisenberg, 1977).

Theories of pain and affect are different. Pain experiences are conceptualized as comprising affective, cognitive, and sensory components which may or may not be closely interrelated under various circumstances (Leventhal & Everhart, 1979; Melzack & Wall, 1982). Within theories of pain, affect is seen as an important but not the only central component (Craig, 1985). The concept of affect however is the obvious central component in theories of emotion, the two terms often being used interchangeably. Tomkins (1982) stated "I view affect as the primary innate biological motivating mechanism, more urgent than drive deprivation and pleasure, and more urgent even than physical pain" (p. 354).

A key question when discussing the use of a single measurement tool across two different fields is how the results based on the tool are interpreted. In theories of affect, the focus is on hypotheses of the cultural determinants, genetic contributions, and communicative functions of facial expressions (Ekman, 1978; Izard, 1971; Redican, 1982; Tomkins, 1982). For theories of pain the data from the same measurement method may be related to the sources of pain, the sensory or evaluative qualities perceived by the person, the aforementioned hypotheses of emotion as they relate to pain, as well as complex interactions of these hypothesized factors (Kleck, et al. 1976).

It is clear then that the similarity of methods used and possible overlapping hypotheses between research on pain expression and emotional expression only represent a circumscribed area of joint theoretical interest between research in these two areas. It is also evident that these two areas are separable and deserve focused attention as unique topics of investigation. This shared methodology but theoretical difference between pain and emotion will be evident throughout the remainder of this paper.

The Concept of Facial Behavior

Whether the focus of research has been on facial expressions of pain or affect, and regardless of the theoretical predilection of the researcher, within the realm of measurement techniques the term "facial expression" has caused considerable confusion. Ekman, Friesen, and Ellsworth (1982a) and O'Sullivan (1982) point out several problems with the terminology used in the study of the face. Ekman et al. (1982a) proposed the term "facial behavior" to indicate observable movement on the face. The terms emotion or expression are reserved for the inferences one makes from observing facial behavior. This distinction, which clearly separates what is observed from that which is inferred, opens the door for researchers in all areas to use common measures of facial behavior without adopting other researchers' inferences about what is being measured. The concept of facial behavior reduces the debate over what is being measured to the problem of defining what a behavior is without also including in this debate the question of how to interpret the meaning of the behavior. The terms facial behavior, movement, and action will be used to indicate facial events that have been objectively coded. The term pain expression will refer to inferences or general descriptions of configurations of facial behaviors that are presumed to represent the subjective state of the individual who demonstrates these specified facial behaviors.

By stripping inference from observation of facial behavior it becomes possible to compare observations from diverse areas in order to investigate the similarities and differences in facial movement which are observed, regardless of the inferences made about these movements. It would be possible to compare the facial behavior observed when people were subjected to fearful or stressful stimuli with the facial behavior of people exposed to painful stimuli. This comparison could suggest the common facial behaviors in various circumstances and identify those which are unique to each stimulus. Within the area of pain research these unique facial behaviors could also be investigated to see if they were related to either specific qualities of the pain stimulus and its sensory impact, to physiological characteristics of the person's reaction or to the person's subjective interpretation of the event. The present studies have focused on the relationship between specific facial behaviors related to pain and judges' ratings of pain based on viewing these facial behaviors.

Summary of Rationale for the Present Research

The thrust of the present research can be quickly summarized. There is a need for more and better pain assessment methods. Of the four major areas where pain assessment can be made only the assessment of nonverbal behavior lacks any adequate measurement methods. The use of nonverbal behavior in assessing pain has clinical and experimental importance. Although several areas of nonverbal behavior could be investigated there are practical, theoretical, and clinical reasons for studying facial behavior. The relationships among certain facial behaviors that have been found concomitant to the report of pain and judges' ratings of the subjective states portrayed by these facial behaviors was investigated in the present research.

LITERATURE REVIEW

Due to the diverse research on facial expressions of pain there is a need for a conceptual framework for examining the strengths and weaknesses of judgment studies of pain. The literature to be reviewed will be organized following Rosenthal's (1982) model for conducting judgment studies.

This review contains four major sections. The first introduces Rosenthal's model and outlines how it is used to structure the review. The second provides a system for classifying empirical research on judgments of pain expression. This system will follow the model's guidelines. The third provides a summary of the findings in each area and a discussion of methodological issues involved in research on facial expressions of pain. The fourth section assesses the utility of the model based on the classification of the empirical research and how well the model helped in identifying new areas for investigation.

Model for Conducting Judgment Studies of Pain

Except for research on facial expressions that utilize physiological recording methods, all studies of facial behavior use observers either as coders (people trained to use a specific recording device or method) or as judges (people asked to make personal decisions based on observations of facial behavior). These two observer roles may overlap but usually fulfill different research needs. For instance, Craig and Patrick (1985) and LeResche (1982) used highly trained coders, Lanzetta and Kleck (1970) used naive judges as observers, and Ganchrow et al. (1978) provided their observers with mimimal training and a structured scoring form.

The aforementioned studies by Craig and Patrick, Lanzetta and Kleck, and Ganchrow et al. all investigated pain expressions, but each study used a different observer role. From the diversity of strategies the question arises as to how best to integrate the findings. The model proposed by Rosenthal (1982) provides an integrative framework for assessing and conducting judgment studies. This model provides a structure for a systematic review and for contrasting the different observer roles and the data collected by each method.

Figure 1 is a graphic display of the model of judgment studies proposed by Rosenthal. The following outline briefly describes the model's components and component interactions.

A - one or more attributes of the person exposed to a known stimulus (eg. pain or fear).

B - the observable behavior of the person (eg. facial or body movement, skin tone change).

C - the judgment of the observer of the A attributes based on perception of B (eg. rates level of pain).

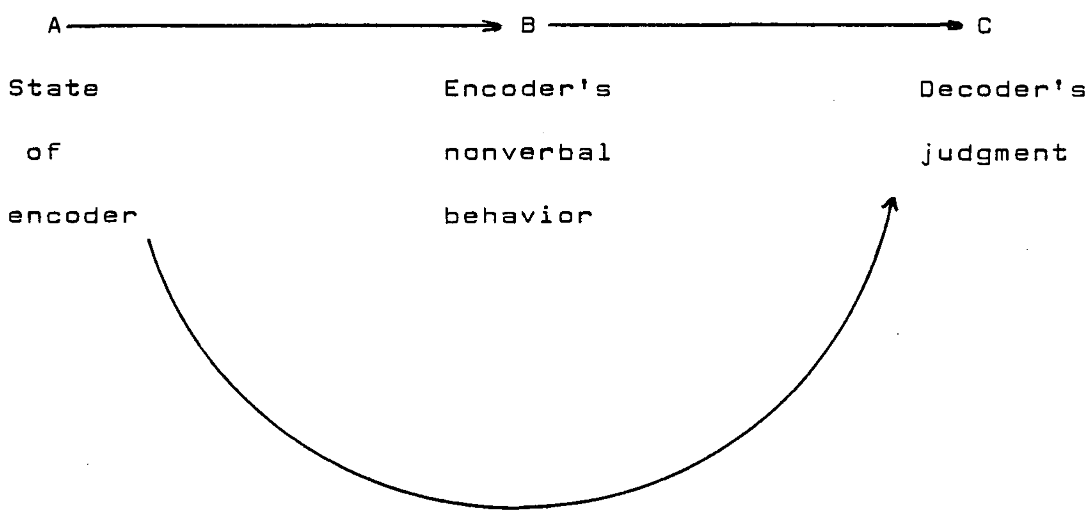
A B - the study of nonverbal behavior as a dependent variable (eg. are there a limited number of facial actions specific to high levels of noxious stimulation?).

B C - the primary interest is on using the encoder's nonverbal behavior as an independent variable (eg. do

Figure 1

Rosenthal's Model for Conducting

Judgment Studies



observers make different judgments dependent upon the type of nonverbal behavior they see?).

A C - the primary focus is on the observer's accuracy in identifying the attributes of the encoder (eg. how accurately can the observers discriminate between pain and fear states of the subjects?).

A B C - the simultaneous interest is all interactions in the model (eg. finding what facial actions relate to specific states of the encoder and testing which facial actions observers use to infer the state of the encoder).

Using this model it is possible to classify empirical research by the particular relationship of the components being studied.

Classification of Empirical Studies of Facial Expressions of Pain

Table 1 summarizes the recent work on facial expressions of pain. Studies conducted prior to 1965 were not included because they contained particularly problematic methodological flaws (Ekman, 1982; Ekman, Friesen & Ellsworth, 1982b), and often did not separate pain from the larger focus on facial expressions of emotion. In a review of research on facial expressions of pain conducted by LeResche and Dworkin (1984) only one study investigating facial expressions of pain (Chapman & Jones, 1944) was found prior to 1965. Also not included in Table 1 are research summaries of work on the physiological factors involved during the observations of a subject receiving pain eliciting stimuli.

Physiological research is outside the realm of the present model of judgment studies. If a study focused simultaneously on both judgments of facial expressions and physiological measures a summary of the judgment section is included in Table 1.

Summary of A - B Findings

It is evident from the studies summarized in Table 1 that research on different component relationships in Rosenthal's model produce unique types of results distinguishable from research on other component relationships. The nine experiments that examined the A-B relationship all listed specific facial features, movements, or areas of activity which appear to be related to facial expressions of pain. This type of research relies heavily on adequate measurement of facial behavior. Comparing the FACS coded data (Craig & Patrick, 1985; Hyde, 1985; LeResche, 1982; Patrick et al., in press) to the more global coding of Ganchrow et al. (1978) and Leventhal and Sharp (1965) and the limited range of Keefe and Block (1982) highlights the gains in comprehensiveness and specificity when using sophisticated coding system versus global or limited coding for this type of research. With the exception of the Keefe and Block (1982, exp. 1) finding that grimacing was not correlated with the patient's pain rating all other experimental studies found that facial behaviors were significantly related to measures of the patient's or subject's pain. Ganchrow et al. (1978) also found that the state of the subject affects the facial behaviors demonstrated during painful stimuluation.

Table 1

Summary of Empirical Research on Judgments of Facial Expressions of Pain Categorized using Rosenthal's Model

Author Model Segment A - B	A State Subject	B[a] Nonverbal Behavior of Subject	C Rating of Judge	Ss in A	Coders in B	Judges in C	Relia-bility	Findings
Craig & Patrick (1985)	Cold pressor. Gracely scale of Affective Discomfort used as self-report. Ss exposed to modeling influence.	FACS		72 F	1 F 1 M		71% overall 85% for signif. AUs[b].	AUs 6,7,10,12,25,26,27,43& 45 discriminated baseline from hand emersion. AUs described as eyes narrow upper lip pull, pulling lips oblique, mouth open, eyes closed or blinking. Sig. tolerant & intolerant modeling effect on self-report.
Ganchrow Steiner & Kleiner (1978)	Cold pressor. Self-report pain 0-10 scale. Hamilton Depress. scale & self-rate depression 0-100.	Coders rated global level of pain & affect.		24 F 8 M 16 depres patients 16 normal volunteer	4 F 2 M		Not given.	Pain tolerance lower for depressed group. At neutral depressed group had higher freq. corrugated brow, squinting, closed eyes & turned down mouth. During pain depressed group had higher freq. of squinted eyes, turned down mouth & compressed lips. Coders rated both groups in more pain during pain than base.

[a] As the sole focus of this table is on facial expressions it would be redundant to list facial expressions under "B". Instead, either the method of measuring facial expressions or specific characteristics of "B", besides facial expressions, will be given.

[b] AU = Action Unit. This is a term used by Ekman and Friesen (1978) to describe movements of specific facial muscles.

Author	A	B	C	Ss in A	Coder in B	Judge in C	Reliability	Findings
Hyde (1985)	Ss were chronic back pain pats. who received a range of motion test. Pats. instructed to hide & pose pain express for one segment. Self-report MPQ Gracely scales.	FACS. Global coders rated overall expressiveness.		60 F 60 M	FACS 2 F 1 M Global 1 F 1 M		FACS 78% overall. Global ratings inter-class correl. = 0.92.	Males rated globally as less expressive. No difference on self-report scales. No sex differences for AU frequency. Overall MANOVA found AUs 2,4, 6,7,10,12,25,43 & 45 sig. different amongst baseline, pain, masked & posed segments. Post hoc analysis found a) AUs 4 & 43 sig. more in pain than base. b) AU 45 sig less during masked than base. c) AUs 4,6, 12 & 43 sig. more & AU 45 sig. less in posed than base. d) AU 25 sig more & AUs 4,6 & 12 sig less in pain than posed.
Keefe & Block (1982) Exper. 1	Ss rated pain 0-10 prior to assess of static & dynamic states. All Ss had chronic low back pain.	5 codes for static & dynamic states & concomitant behavior codes for guarding, bracing, rub, grimacing & sighing		13 F 14 M	2		Effective % 80-96	Correlation of guarding & grimace with other behaviors was highest. Correlation of pain rating with all behaviors except grimace was sig. Mean frequency of combined behaviors sig. higher during movement than static assessment.
Keefe & Block (1982) Exper. 4 See also in B-C below.	Ss were either pain patients, normal or depress. controls. All were assessed as in experiment 1.	Same as in Exper. 1		10 Ss in each group.	Blind to Ss status		Not given	All behaviors sig. higher in pain patients than controls, except bracing which was not sig. different between pain patients and normals.

Author	A	B	C	Ss in A	Coder in B	Judge in C	Relia- bility	Findings
Keefe Wilkins & Cook (1984)	Ss were low back pain patients receiving neuro- surgical evaluation. Ss rated pain on 0-10 scale.	Same as in Keefe & Block (1982).		32 F 48 M	2		91.9%.	Total pain behaviors sig. related to pain ratings. Sig. more total pain behaviors dur- ing movement than when still. Mult. regress. found number of operations predicted total pain behaviors, guarding & bracing. Exam findings predicted total pain & guarding. Duration of pain predicted guarding. No patient characteristics predicted rubbing, grimacing & sighing.
LeResche (1982)	Still photos of unanesthetized people experiencing pain from acute severe trauma.	FACS		7 Ss in 16 photos	1 F		None	Indices of pain were AUs 4,6,7, 11,20,25,26,27 & 43. AUs described as brow lower, horiz- ontal lip stretch, closed eyes mouth open, cheeks raised, eye lids tight & nasolabial furrow.
Leventhal & Sharp (1965)	Hours in labor & cervical dilation.	Birdwhistell coding system (1952).		71 F	2 F		75.5% mean agree over sig. moves.	Brow, forehead & eye movements sig. increased with dilation & hours in labor.
Patrick Craig & Prkachin (in press) See also A-C below.	Ss received shocks while exposed to models. Used Gracely scale of Affective Discomfort.	FACS		30 F	1 F 1 M		74% overall 84% for sig. AUs.	Sig. AUs discriminating highest shock from baseline were AUs 4, 6,10 & 45. Frequency of AUs sig. higher in tolerant than intolerant or control condition Au frequency sig. greater in pain threshold or high shocks than baseline.

Author Segment A - C	A	B	C	Ss in A	Coder in B	Judge in C	Relia-bility	Findings
Gubar (1966)	Anticipation of shock or reward.		Judges yoked to Ss & could avoid shock by making correct judgments of S's condition. Js were naive or experienced with shock.	23		47	N/A	Experienced Js better at knowing when to respond. Naive Js failed to discriminate between shock & reward trials but experienced Js made accurate discriminations.
Kleck, Vaughan, Cartwright-Smith, Vaughan & Lanzetta (1976) Exper. 1	Received no/med/high shocks alone or being observed. Rating of shock 0-100. Ss were selected as high or low self-monitors.		Judges rated shock 0-100 scale.	20 M		3 F 1 M	Kendalls coeff. of concordance p .001	Hypothesized shock level X experimental condition did not reach sig. For self-report main effect for shock level but hypothesized shock level X condition was not sig. Less discomfort judged in observed vs alone condition.
Kleck et al. (1976) cont. exper. 2	Replication of exper 1. No selection for high low monitoring.		Replication of experiment 1.	40 M		3 F 3 M	Same as above	No sex effect. Less discomfort judged in observed vs alone condition. Self-report rating lower when observed than alone.
Lanzetta Cartwright-Smith & Kleck (1976) exper. 2	Anticipation of no/low/high shock under instructions to be demonstrative or hide facial expression or baseline		Js indicated if Ss were anticipating shock & its level. Js observed self & others. All Js were also Ss.	9 M		9 M	None given	Js made fewer errors judging expressions under demonstrative than hide or baseline conditions

Author	A	B	C	Ss in A	Coder in B	Judge in C	Relia-bility	Findings
Lanzetta et al. (1976) cont. exper. 3	Anticipation & reception of low/med high shocks. Posed no pain & intense pain. Rated pain 0-100.		Js rated intensity on 0-100 scale.	10 F 10 M		2 F 6 M	.90 mean correl. for test retest reliab.	No sex differences in expressions. Posing had sig. effect on Js ratings of shock. Sig. posing X intensity. Ss reported lower pain in pose "no pain" vs pose "intense pain".
Lanzetta & Kleck (1970)	Anticipation of mod. level shock or no shock.		Js viewed tape of self & others. Js received shock if incorrect.	12 M		12 M	None given	Js could discriminate shock vs no shock facial expressions. Ss who made few judgment errors were themselves difficult for others to judge. Those making many judgment errors were relatively easy for others to judge.
LaRusso (1978)	Ss receiving shocks or simulating receiving shocks.		Js viewed tape & scored segments as real or simulated. Half of Js received single shock.	2 M 2 F		24 paranoid 24 normal contols	N/A	Group X tape condition. Pats. more accurate with genuine expressions than contols. Patients viewing simulated expressions less accurate than controls. Contols viewing genuine express. were less accurate than when viewing simulated expressions.
Patrick et al. (in press) See also A-B above	Videos of Ss receiving shocks while exposed to tolerant intolerant or contol models.		Js given analgesic, hypersensitive, or control instructions. Rated Ss on 7 pt. scale of discomfort.	30 F		15 F	None given	Js rated Ss receiving higher shocks to be in more pain. Instructions to Js has sig. effect at highest shock level. Js rated tolerant Ss to be in more pain than other two groups Mult. regress. showed AUs 4,6, 10 & 45 sig related to Js ratings.

Author	A	B	C	Ss in A	Coder in B	Judge in C	Relia-bility	Findings
Prkachin & Craig (1985)	Videos of Ss receiving shocks while exposed to tolerant intolerant & control model		Js rated shocks as low/med/high & rated Ss' overall distress & difficulty in rating S.	27 F		15 F	SDT analysis Lx value non-sig.	Js discriminated shock levels. Discrimination between high & med shocks easier than med & low. Modeling affected Js rating of intensity. Js global ratings of distress differed amongst modeling groups. Js ratings of difficulty sig. different across modeling groups.
Prkachin Currie & Craig (1983)	Videos of Ss receiving shocks while exposed to tolerant intolerant & control models. Rated pain on 7 pt. scale		Js given analgesic hyperanalgesic, or control instructions Rating pain on 7 pt. scale.	30 F		15 F 15 M	N/A	Js discriminated shock intensity. Js' ratings differed across model conditions. Sig. shock X model which showed most sig. differences at high shock. Js rated face as more important than body cues. No difference due to sex of judge. Hypersensitive instruct. lead to higher judgment of pain
<u>B - C</u>								
Block (1981)		Videos of spouses or confederate displaying pain or neutral expression on command. Spouse was pain patient.	Non-patient spouses & neutral observers judged expressions on 11 pt. scale.		Not given	16 spuses 2 neutral observers	Not given	Spousal judges & neutral observers rated pain expressions as more pain than neutral for both their spouses & the confederate.

Author	A	B	C	Ss in A	Coder in B	Judge in C	Relia- bility	Findings
Boucher (1969)		Still photos of faces previously judged to show fear/suffering; fear; distress, as well as psychiatric controls	Judges rated photos as fear; sadness/ sorrow; actual pain		Photos from previous research	47	Not given	Chi square test indicated distribution of judgments differed sig. from chance for 36 of 39 photos.
Keefe & Block (1982) See also A-B above		Coding was done for 22 video segments using system from experiments in A-B above.	Js used 11 pt scale, visual analogue & Gracely scale of sensory distress.		Not given	9	Not given	Total level of coded pain behaviors sig. correlated with all 3 types of Js ratings. Guarding & grimacing correlated independently with all 3 pain rating scales used by Js.
Von Baeyer, Johnson & McMillan (1984) Exper. 1		Videos of low or high nonverbal pain expressions by patient. Verbal content was standard. Male doctor & female patient were acting.	Js rated video on 21 item visual analogue scale. Js were senior undergrad. nurses.		Not given. Pre- test was done.	44 F	Not given	Js in high condition gave sig. higher pain concern scores. Js did not differ in rating need for aid or support for patient. In high condition neg. correlation between experience and pain concern.
Von Baeyer et al. (1984) Exper. 2		Same as above except low/med/high expressions.	Js rated videos on 16 items using 6 pt. scale. Js were divided into low/ med/high nurturance.		Same as above no pre- test for med.	88 F	N/A	Face & body movements & voice tone were rated as sig. diff- erent between high & low with med. in between. Js ratings increased monotonically with increases in pain display. Sig. interaction between nurturance and pain express- ion on solicitude scale.

Brow furrowing, eyes tight and/or closed, mouth open, and lips stretched (upper lip pulled up or lips pulled horizontally or obliquely) are the general findings of these studies. The consistency of these findings is remarkable for such a small number of studies, given the fact that five very different clinical and experimental pain conditions were studied (cervical dilation/hours in labor, chronic low back pain, cold pressor, electric shock, and photos of unanaesthetized people experiencing acute severe traumas). Also of note is that four different methods of coding nonverbal facial behavior were used. Although in its infancy this research appears to provide a good beginning towards the identification of possible systematic facial expressions of pain.

Limitations of A - B Research

LeResche (1982), in an uncontrolled case study, provided the weakest of the studies. However, it is interesting to note the similarity of her findings to the general findings in the area. While the variety of facial coding methods could be seen as a strength of this area the tremendous variablility in comprehensiveness and specificity between methods could be artificially increasing the overall consistency of the findings. For instance, the findings by Leventhal and Sharp (1965) are open for interpretation as FACS coding of the brow, forehead, and eyes uses 10 AUs (AU or Action Unit is a term used by Ekman and Friesen, 1978, to describe movements of specific facial muscles) for coding movement in these areas, only five of which are identified as relating to pain in research using FACS. It must remain uncertain if Leventhal and Sharp reported the same facial behaviors as FACS based research even though there is agreement that the upper face expresses important pain cues. Keefe and Block (1982) used a predefined set of facial pain behaviors. Using a predefined set of behaviors precluded the possibility of finding other facial behaviors which may have been powerful indicators of pain. Also, the description of "grimace" given by Keefe and Block is not in accord with the FACS descriptors found in other A-B research.

Summary of A - C Findings

Before discussing the results of A-C research it is important to recognize the role of component B, or the subject's nonverbal behavior, in the A-C relationship. The critical difference between A-C research and the other areas of research is that the nonverbal behavior of the subject is not measured even though it is the sole source of direct information from the subject that is available to the judge. The sole focus of A-C research is on whether judges, given access solely to the nonverbal behavior of the subject, can accurately rate the state of the subject.

The findings of studies focusing on the A-C relationship indicated, in broad terms, that judges could make discriminations between facial expressions of subjects anticipating a shock or no shock (Gubar, 1966; Lanzetta & Kleck, 1970) and amongst facial expressions related to different levels of shock intensity (Kleck et al., 1976, exp.

1; Lanzetta et al., 1976, exp. 3; Patrick et al. in press; Prkachin & Craig, 1983; Prkachin et al., 1983). Several subject and judge manipulations were found to affect judges' ratings. Judges rated subjects who were observed to be in less discomfort than subjects who were unobserved (Kleck et al., 1976, exp. 1 & 2). Modeling influences on subjects altered judges' ratings (Patrick et al., in press; Prkachin & Craig, 1985; Prkachin et al., 1983). Instructions given to subjects to alter their facial expressions affected judges' ratings (Lanzetta et al., 1976, exp. 2 & 3; LaRusso, 1978). Information given to judges regarding the state of the subject led to differential judge responses (Patrick et al., in press; Prkachin et al., 1983). The present state of the judge (normal or paranoid) affected their ability to discriminate between spontaneous and simulated pain expressions (LaRusso, 1978). Finally, giving judges repeated experience with the pain stimulus increased their accuracy (Gubar, 1966) while a single exposure did not (LaRusso, 1978).

The findings regarding the A-C relationship are instructive in three areas. First, it is obvious that judgment outcomes are sensitive to an array of social, instructional, and knowledge factors. Social, instructional, and knowledge factors when applied to the subject appear to affect judges' ratings. Judges' knowledge of the subjects' state and whether the judges were paranoid or not also affected judges' ratings. Second, the studies to date have focused on the relative magnitude of the effects of the various manipulations and not on the absolute accuracy. That is, research has shown discriminable results across levels of most manipulations but almost no information regarding how close the judges came to rating the subjects' state as defined by the subjects' self-report or to the absolute magnitude of the stimulus. Kleck et al. (1976) in both their experiments on the effects of subjects being observed or unobserved gave self-report and judges' ratings of pain intensity that suggested that, regardless of condition, self-report far exceeded judges' ratings for mid and high intensities but judges' ratings far exceeded self-report at zero intensity. Lanzetta et al. (1976, exp. 3) also indicated large discrepancies between judges' ratings and self-report during posed "intense" and posed "no pain" conditions. Judges again were consistently lower than self-report except for posed "intense" at low shock levels where the ratings were similar. The major clinical question which arises from these results is what differences in pain management efficacy are there between management based on judgments of relative difference versus judgments of absolute accuracy? More research is needed in order to adequately address this question.

The final area where the results from A-C research are instructive is in the area of individual differences. Lanzetta and Kleck (1970) and Kleck et al. (1976) pointed out that there were substantial differences in judges' reliability across subjects. Some subjects were reliably judged, while for others judges showed little reliability. This variability is probably due to both individual differences in subjects' facial behaviors as well as judges' ability to make certain discriminations. The amount of individual difference in subjects' nonverbal behavior is best portrayed in the research by Patrick et al. (in press). Patrick et al. found that 55% of the variance in judges' ratings could be accounted for by changes in four specific components of the subjects' facial expressions (AUs 4, 6, 10, and 45). While this is an impressive result for just four facial behaviors it also indicates that the source of 45% of judges' variance is left unaccounted.

Limitations of A - C Research

The intent of A-C research within the model is to study how well judges map the state of the subject and what factors other than nonverbal behavior may affect this mapping. Two problems in the reviewed literature have limited the fulfillment of this intent. First, only Kleck et al. (1976, exp. 1 & 2) and Lanzetta et al. (exp. 3) used identical self-report and judges' rating scales which would allow for direct comparison between these scales. As noted in the findings above there were major discrepancies between the subjects' and judges' ratings. Unless further attention is paid to the issue of accuracy this discrepancy will remain unclarified. Second, the measurement properties of most of the measures of the subjects' state are not well known. The Gracely scale of affective discomfort has research supporting its measurement properties (Gracely et al., 1978, 1979). While in common use, the serial rating scales (eg. 1-100) often lack adequate support for either their reliability or validity (Gracely, 1982; Tursky, Jamner & Friedman, 1982). Thus, although such scales have proven useful for allowing subjects' and judges' ratings to discriminate events it is unknown what the parameters are for such discriminations. It is particularly important to specify what is being assessed when studying pain because of its multidimensional nature (Chapman et al. 1985). This need for specificity is indicated by Gracely et al. (1978) producing separate affective and sensory scales and the MPQ being divided into sensory, affective, and evaluative subscales (Melzack, 1975; Prieto & Geisiger, 1982). The lack of specificity in the research to date hinders a clear interpretation of what aspects of pain the judges are rating.

Two factors limit the generalizability of the A-C research. Electric shock is the only pain stimulus used. This is clearly an insufficient base for generalizing results from this area (Harris & Rollman, 1983; Turk et al., 1983; Wolff, 1971). Also, very little has been done towards studying the attributes of the judges and how they may contribute to the judges' ratings. Only the sex of judge (Kleck et al., 1976, exp. 2; Prkachin et al., 1983) and judges' experience with the pain stimuli (Gubar, 1966; LaRusso, 1978) have been studied, with there being no effect for sex of judge. Although a number of judge attributes may prove to be important the research by Von Baeyer et al.

(1984) regarding the B-C relationship would suggest that plausible attributes to study would be judges' work experience in rating pain and level of nurturance.

<u>Summary of B - C Findings</u>

The model's major criterion for studying the B-C relationship is that there is experimental control over the nonverbal behavior shown to the judge. This permits specification of the nonverbal (ie. facial) behaviors to be used as stimuli for the judges and is an independent issue from the manipulation of the subjects' state which is hypothesized to alter their nonverbal behavior. If this criterion was strictly followed only Boucher (1969) and Keefe and Block (1982) would have been included in Table 1. Block (1981) and Von Baeyer et al. (1984) were included because they made some attempt to control or define the nonverbal behaviors shown to the judges although they did not specify the explicit parameters of these behaviors.

The findings suggest that direct manipulation of the nonverbal facial behavior of the subject can lead to alteration of the judges' ratings independent of explicit manipulation of the subjects' state. The studies indicated that judges can differentiate between levels of pain and neutral facial behavior (Block, 1981; Von Baeyer et al., 1984), between fear, sadness/sorrow, and actual pain facial behavior (Boucher, 1969), and that pain related behaviors, especially grimacing and guarding behaviors, correlate with three types of judges' rating scales (Keefe & Block, 1982, exp. 4).

Limitations of B - C Research

As stated above only Boucher (1969), and Keefe and Block (1982) came close to fulfilling the requirement for B-C research of specifying the nonverbal behavior viewed by the judges. By comparison to the specificity achieved in coding nonverbal behavior in A-B research it is evident that the research by Boucher, and Keefe and Block only barely meet the levels of specificity and comprehensiveness obtainable by current coding systems such as FACS. The results of all B-C studies must be viewed as somewhat inconclusive due to the lack of specifying particular facial behaviors which affected the judges' ratings. Thus it is not known if judges focused on the upper or lower face movements, on lip or eye movements, or some complex of facial behaviors.

It is also disturbing to note that no data were given regarding the reliability of the coding done to specify the nonverbal behaviors used. Without such data it is unknown how well the facial behaviors fit the description of pain expressions used to describe the nonverbal behavior seen by the judges. From the perspective of fulfilling the conditions of Rosenthal's model this is a serious problem. Thus, even though the B-C research used a variety of judges, which is important, the lack of specificity of the nonverbal behaviors viewed by these judges means these studies have not adequately fulfilled the intent of the model's guidelines.

General Summary of Reviewed Research

A-B research has moved generally from molar to more molecular coding of facial behavior with a concomitant increase in coding comprehensiveness. Specific facial movements have been found to discriminate amongst different pain states and pain versus non-pain states. A-C research has shown that judges can discriminate among various pain states of the subjects, although, as expected from the model, what cues the judges use for their discriminations is not known. It has also been shown that judges' ratings are affected by a variety of social, instructional, and knowledge factors. Electric shock is the only pain stimulus used to date in this research. B-C research indicated that judges' ratings do vary as a function of the facial behaviors shown to them. However, due to a lack of specificity these facial behaviors are not well defined. Also, as expected from the model, this research does not tie specific facial behaviors to the subjects' state. Overall the model's basic component relationships have been investigated, albeit the quality and focus of the research was variable.

The research by Patrick et al. (in press) and Keefe and Block (1982) which was listed in Table 1 as investigating two types of relationships, A-B and A-C, and A-B and B-C respectively, will be the focus of the final summary remarks. At first glance it may appear that these two studies could have been used to investigate more complex component relationships. However, this is not the case. What the studies revealed was that simply studying the simple component relationships separately was not enough to guarantee that direct comments could be made about more complex relationships. In the Patrick et al. study it was not possible to take the separate A-B and A-C research and make a clear statement about an A-B-C complex relationship due to using a signal detection design in the A-C part of the study, a judges' rating scale that was different than the subject's self-report, and the testing of the judge's accuracy being collapsed across sub-pain to pain tolerance stimuli segments.

In the Keefe and Block series of studies differences in subject samples, limited assessment of the subjects' state, and focusing on the correlation between coded behavior and multiple judgment rating scales interfered with attempts to make a statement about the larger relationships than those directly addressed in the studies.

The difficulties of dealing with the complexity of the entire model is recognized. In developing the present research, careful attention was given to the design of the experiments so that they could be clearly integrated into the overall framework of this model.

Evaluation of Rosenthal's Model

At this point it is possible to provide a tentative evalution of Rosenthal's model for conducting judgment studies. This section will profile the strengths and weaknessess of the model.

Strengths

From use of the model to date three areas of strength have been identified. First, the fact that prior to this research few of the reviewed studies had been cited together indicates that the model is powerful enough to allow previously divergent research to be organized into a systematic structure. A benefit of this structure is that several specialty areas can be identified (eg. the A-B relationship). The just completed review of the literature is an example of the benefits of the identification of these specialty areas of research. Using the model allows for the integration of research on special topics so that larger and more coherent statements can be made on specific issues than if the research was left unstructured.

The second strength is that the model sets guidelines for what should and should not be included in or expected from research in a particular area. As pointed out in the general summary of the research it was expected that if the question of interest is focusing on the cues used by judges then B-C research is the place to look. However, as revealed above, the B-C research to date does not meet this expectation and in not meeting this expectation one of the weaknesses of this area of research was identified. This example demonstrates that the expectations of the model not only help in deciding where to search for research on a particular question but also helps in evaluating the research pertinent to a specific question.

Finally, this model has proven useful for gaining a fuller appreciation of the intricacies of studying judgments of nonverbal facial behavior. The problems in using the

Patrick et al. (in press) and Keefe and Block (1982) studies, both of which focused on two types of component relationships, for discussing the more complex relationships suggested by the model highlights some of the intricacies of researching this area.

Model relationships not yet addressed

As is obvious by now the major area not yet researched is the A-B-C relationship, with, as mentioned earlier, the B-C relationship awaiting more solid research. Research on the whole model is critical if links between the model's segments are to be forged so that it is possible to evaluate the interrelationships between the A-B, B-C, and A-C components. This would allow for a proper evaluation of the model's utility for guiding research in the area of studying judgments of facial behavior.

Weaknesses

Besides the fact the model as a whole has not been tested, which is not a weakness of the model per se but rather a caution that the model's utility is unknown, there are two weaknesses in the model. The first weakness is that the model has no explicit connections to other areas of related research. Thus the model stands in isolation to work done on physiological factors related to either encoding of facial behavior (Lanzetta et al., 1976) or making judgments (Bandura & Rosenthal, 1965; Block, 1981; Vaughan & Lanzetta, 1970), or to vocal factors as other indicators of the subjects' state (Cline et al., 1972; Ekman et al., 1980). Cognitive research dealing with such issues as self-observation of pain behavior (Bandler, Madaras & Bem, 1968) and judgment heuristics (Kahneman, Slovic & Tversky, 1982) are other areas to which the model may need to be connected. The trend suggested by these examples is that each general component could be broken down into several subcomponents and research would focus on the relationships amongst the subcomponents.

The second problem with the model is its unidirectional character. The model is not formulated to represent reciprocal interactions which, from such perspectives as social learning theory (Bandura, 1977), are critical in explaining ongoing human social behavior. While the model does not represent reciprocal interactions it is not incompatible with this perspective, and could constitute a component of an interactional model.

<u>Conclusions from evaluating Rosenthal's model</u>

At present the model appears to be well suited to the type of research that has been conducted on judgments of pain based on viewing facial behavior. It has served well in structuring the review of the literature and suggests several interesting questions. It would appear that the limitations of the model have not yet become a serious issue as, other than the possibility of connecting the model to the physiological data available, the necessity for decomposing each component is not yet critical. Also, since the research to date has not been able to deal completely with the whole model it is unlikely that the model will presently become eclipsed by a more reciprocal model. Therefore the model will be used for describing the design of the present studies and in discussing the results.

HYPOTHESES

Prior research has pinpointed the occurrence of certain AUs as important with regards to subjects' facial displays of pain (Craig & Patrick, 1985; Hyde, 1985; LeResche, 1982; Patrick et al., in press) as well as judges' ratings of pain (Patrick et al., in press). The two studies reported here examined a number of critical features of facial expressions of pain which could play significant roles in leading judges to rate a subject as experiencing pain.

In these studies those AU configurations that have been found related to pain were systematically varied in order to measure the impact of these AU configurations on judges' ratings of the encoders' facial expressions of pain. Within the judgment model this research was concentrating on the B-C relationship. As discussed in the literature review, the research to date studying the effects of nonverbal facial behavior on judges' ratings has been solely correlational and wanting of tight control over the stimuli viewed by the judges. These studies were the first attempts to experimentally control facial behavior of pain at a micro-behavioral level.

The major factors addressed were 1) the impact of increasing the complexity of specific AU configurations on judges' ratings of pain; 2) the effect of overlapping versus non-overlapping AU configurations on judges' ratings; and 3) the role of upper and lower face AUs as cues of encoders' attempts to control their facial expressions of pain.

Previous research has shown there to be no sex of judge effect with regards to rating levels of pain (Kleck et al., 1976; Prkachin et al., 1983). However, the lack of experimental control over the facial expressions viewed by the judges still leaves open the question of whether there are sex differences between judges that are specific to certain types of facial configurations. These findings in pain judgments are also in contradiction to general judgment research using facial expressions of emotion which consistently have found sex of judge effects (see Hall, 1979, 1984 for reviews). Thus sex of judge was included as a subsidiary factor in the present studies.

Due to limitations on the sample size of encoders, which will be discussed shortly, no sex of encoder factor will be examined as it is not appropriate to construe the small sample as representative of either sex (Stark-Adamec & Kimball, 1984).

<u>Hypothesis 1 - Impact of the Complexity of AU Configurations</u>

The initial question investigated was concerned with the effects of increasing the complexity of the AU configurations (i.e. by increasing the number or AUs per configuration) on judges' ratings of the encoders' expressed level of pain. From previous research it has been demonstrated that judges can discriminate amongst facial expressions of subjects receiving various levels of painful stimulation (see the A-C research in Table 1). It has also been shown that, at a global level, it is possible to manipulate facial expressions in ways that alter judges' ratings (see the B-C research in Table 1). However, this research has made very few inroads into the question of what specific facial movements judges use in making their discriminations.

Despite large intra- and inter-individual variations in AUs displayed, research by Craig and Patrick (1985), Hyde (1985), and Patrick et al. (in press) has revealed a consistent set of pain-related AUs. From a total of 44 possible AUs the only Action Units that have discriminated pain and non-pain states are AUs 4, 6, 7, 10, 12, 25, 43, and 45, with AUs 4, 6, 7, 10, and 45 appearing as a core constellation of AUs since they all appeared in each of the above three studies which focused on cold pressor, shock, and low back pain.

This core constellation of AUs also contibuted to judges' attributions of pain. Patrick et al. (in press) found that a multiple regression equation containing AUs 4, 6, 10, and 45 had a mean multiple R = 0.78 in predicting judges' overall ratings of pain. This correlation between the core AUs and judges' ratings does not indicate a clear causal relationship between the core AUs and judges' ratings. A possible causal relationship could not be investigated by Patrick et al. as the core AUs were embedded in an ongoing stream of other AUs which did not allow for direct manipulation of the these specific AUs. The complexity of the AU configurations and lack of knowledge regarding how judges made their ratings leaves open the possiblility for third variable factors which could account for the correlation between the core AUs and judges' ratings. Factors which could possibly account for the multiple correlation between the core AUs and judges' ratings are 1) the core AUs occurred with other AUs which could have been important cues for interpreting the significance of the core AUs; 2) there may be subtle changes in AU configuration patterns which were not detected by analysis using only the frequency of AU occurrence; and 3) there were uncoded and uncontrolled head and body movements which could have been used as cues by the judges. To eliminate these sources of ambiguity in interpreting the role of facial actions as determinants of judgments of pain AUs 4, 6, 7, 10, and 45 were isolated from all other AUs and presented systematically to judges in order to investigate their causal effects on judges' ratings. It was hypothesized that the more of these core AUs contained in a facial configuration the higher would be the judges' ratings of the encoders' pain.

Hypothesis 2 - Overlapping versus Non-Overlapping AUs

The above mentioned hypothesis is based on the assumption that the simple occurrence of pain-related AUs allows judges to discriminate among facial expressions related to different levels of pain. Ekman (1982b), and Ekman and Friesen (1978) asserted that the critical factor that judges use in rating an AU configuration is not the occurrence of particular AUs per se but whether the AUs have a temporal overlap or not. In particular, Ekman and Freisen contend that the AUs must have overlapping apices (an apex is the span of time during which an AU is at its most intense level for that particular occurrence) before the AUs will be perceived as an unitary global expression. Following Ekman and Friesen's argument it is suggested that it is necessary for potentially relevant AUs to be perceived by judges as unitary expressions in order for the judge to rate the expression. In describing the occurrences of core AUs Patrick et al. (in press) indicated that the core pain-related AUs tended to overlap in various combinations during painful stimulation. It was not possible to test the importance of the AUs having overlapping apices with regards to judges' ratings because of the concurrent presence of other AUs, lack of comparable segments containing AUs with and without overlapping appices, and insufficient coder reliability in measuring AU apex onset and offset. It was hypothesized that video segments in which the AUs had overlapping apices would be rated as displaying more pain than segments where the same AUs did not overlap.

Hypothesis 3 - Hidden vs Non-Hidden AU Configurations of Pain

The question of how well people can control their facial expressions and how observers deal with such expressions has been of interest with regards to emotional expressions (Ekman & Friesen, 1969, 1975) and pain (Lanzetta et al. 1976; LaRusso, 1978). In this area of research a distinction has been made between voluntary (controlled) and involuntary (uncontrolled) movements. Clinically this distinction is relevant to studying deceptive and stoical behavior. Since both types of behavior are dependent upon whether people can or cannot control their behavior it is likely that research on either stoical or deceptive behavior will relate to the other type of behavior.

A distinction between hidden and non-hidden (voluntary and involuntary facial movements) expressions can be made on a physiological as well as a clinical, observational basis. The control of facial muscles of expressions is through the seventh cranial nerve or facial nerve (Diamond, 1979; Hjortsjo, 1969). The branches of the facial nerve to the upper and lower face are different. The upper face is primarily innervated by the temporal and zygomatic branches while the lower is innervated mainly by the buccal and mandibular branches of the facial nerve (Anderson, 1983; Diamond, 1979). In addition, the central nervous system substrates of the upper branches differ from that of the lower branches. The temporal and zygomatic branches are subject to bilateral brain innervation while the lower face branches receive contralateral innervation (Diamond, 1979; Rinn, 1984). Contralaterally innervated muscles tend to be under greater voluntary control than bilaterally innervated muscles (Rinn, 1984). It would seem reasonable to assume that upper face movements are under less voluntary control (ie cannot be hidden or controlled easily) than lower face movements.

Ekman and Friesen (1975) in discussing facial movements that can be more or less controlled during deception repeatedly pointed to upper face actions as the least controllable actions. In describing where to find cues of the person's actual emotional state when they are trying to be deceptive, Ekman and Friesen suggested that the upper face is the most consistently revealing area to look for cues of the person's actual emotional state.

The above arguments indicate that upper and lower face movements can be classified with regards to their ease of control by the encoder. It was not known if naive judges use this distinction to differentially rate upper and lower face configurations as indicating attempts by the encoders to hide their facial expressions. It was hypothesized that judges would discriminate upper face configurations as more often indicating attempts by the encoders to hide their facial expressions of discomfort than lower face configurations.

Hypothesis 4 - Effects of Judgment Context on Judges' Ratings

The purpose of the fourth hypothesis was to investigate the effects of the judgment context given to the judges regarding the state of the encoders. From prior research studying the A-C relationship (Patrick et al., in press; Prkachin et al., 1983) it was found that information given to the judges regarding the subjects' state affected the judges' ratings of the subjects' pain. Lanzetta et al. (1976) demonstrated that judges easily discriminated between facial expressions of the subjects when they were trying to pose "no pain" or "intense pain" while receiving various levels of shocks. However, as is the case with A-C research no information was available on the exact nature of the facial movements which may have led judges to make differential ratings.

As will be explained below two studies were conducted. These studies contained a subset of identical AU configurations but differed in the instructions given to the judges. This situation provided an opportunity to assess the impact of the judgment context given to the judges on their pain ratings of specified AU configurations. It was hypothesized that the ratings would be significantly different between the two sets of instructions. The lack of directionality in this hypothesis was due to the exploratory nature of the hypothesis.

GENERAL METHODS

Two separate studies were conducted. The first study investigated the first two hypotheses while the second study focused on the last two hypotheses. Both studies shared a number of common procedures regarding the development of the stimulus tapes. The following sections give details on the encoders who made the facial configurations, the rules for selecting stimulus segments, and the reliabililty of the facial configurations on the final stimulus tapes.

Construction of Video Segments

In order to focus solely on the core pain-related AUs it was necessary to use short duration video segments to minimize the intrusion of extraneous factors, such as inadvertent AUs, head movements, and facial skin tone changes, which could alter the judges' interpretation of the manipulated AU configurations (Ekman, 1978; O'Sullivan, 1982). All segments showed only the encoders' face and shoulders.

An initial issue was setting the duration of the video segments. The duration of the segments was important for controlling extraneous factors and setting the visual context for the expressions. Due to the widely differing research thrusts on studying facial expressions no standards for video segment length have been established. In the literature on facial expressions of pain the segments rated by judges have ranged in duration from three seconds (Patrick et al., in press) to 10 minutes (Keefe & Block, 1982). This variability is also apparent in studies of facial expressions of emotion. Rosenthal and DePaulo (1979) reported a study using video segments of 250 milliseconds and two seconds, while Ekman et al. (1980) used two minute segments. The segment duration for study 1 was five seconds, and for study 2 it was three seconds. The difference was due to the fact that the second study had simpler AU configurations which the encoders could do faster than many of the configurations in the first study.

Encoders

The major criterion for selection of the encoders was that they be able to produce the complete set of required AU configurations upon demand. Due to this constraint only four encoders, two male and two female were recruited. One male and one female encoder had extensive experience with FACS while the other two encoders had some familiarity with the system. The mean age of the encoders was 26.0 yrs.

Encoder Training

As all four encoders were familiar with FACS it was possible to use the FACS coding manual for describing the AU configurations to the encoders. The author, who is a certified FACS coder, gave continuous feedback to the encoders during the training for each configuration and throughout the taping sessions leading to the production of the stimuli segments. All training and taping was done on an individual basis. The individualized feedback and training over and above the information from the coding manual was necessary due to the individual differences exhibited by the encoders in mastering the various AU configurations.

<u>Rules for Selecting Video Segments</u>

While posed expressions allow for increased experimental control they do so by increasing the risk of stimulus artifacts that confound the interpretation of the effects of the independent variables and may also decrease the external validity of the study (Ekman et al., 1982b; O'Sullivan, 1982; Rosenthal, 1982). The two issues central to this general concern are stimulus measurement ability (Ekman, 1982b; O'Sullivan, 1982), or the ability to accurately measure the independent variables, and stimulus criterion validity (Archer & Akert, 1977), which focuses on the relationship between the parameters of posed and spontaneous expressions. Both of these issues are addressed in the rules for selecting video segments.

<u>Rules for equating posed with spontaneous expressions</u>

By reviewing the prior research which identified the core pain relatd AUs (Craig & Patrick, 1985; Hyde, 1985; Patrick et al., in press) it was possible to describe the spontaneously occurring patterns of AUs. From this review the following two selection criteria were used.

1. The apex for isolated AUs and for AU configurations could not exceed one second in duration. Thus if an AU occurred by itself its apex could not be more than one second. If two or more AUs had overlapping apices then the period when their apices overlapped could not exceed one second.

2. No AUs could be at the extreme intensity level using the

FACS coding method for assessing intensity.

<u>Rules for eliminating movement artifacts</u>

In addition to the above two rules each segment must also have passed the following two rules:

3. No movement other than the ones specified were allowed.

4. A minimum of two and one-half seconds of neutral expression was required before and after the completion of the configuration to ensure that there was enough segment duration to standardize the segment length and have the complete configuration in the final editted form.

<u>Rule for standardizing segment context</u>

O'Sullivan (1982), Kendon (1982), and Tagiuri (1969) pointed out that behavior can only occur within a context and that variations in context may affect judges' perceptions of the behavior. To control for contextual artifacts the following general rule was used.

5. Any segment which had a variation in context (eg. clothing, make-up, hair-style, lighting, etc.) was eliminated. This rule was necessary to eliminate variations due to the encoders needing several days to complete the set of posed configurations. Each encoder chose his or her own appearance and remained with that appearance for all their segments.

One context which was systematically varied over encoders and facial configurations was head position. This was done to produce a more realistic nature for the stimulus tape. Four head positions were used. The encoders looked near-off camera right and left (30-40cm), and far-off camera right and left (70-90 cm). These four head positions were fully balanced across encoders and configurations. The far-off camera positions still allowed for a full view of the face.

How rules were applied

An independent, certified FACS coder applied all five rules. To minimize the amount of coding required the author used rules 3, 4, and 5 to eliminate the more obviously unacceptable segments prior to giving the independent coder the list of potentially acceptable segments. If two or more segments contained acceptable configurations by the same encoder one was randomly selected for inclusion into the study.

Equipment

The encoders were recorded using an RCA CC007 color camera and Panasonic AG6300 VHS video cassette recorder. Editing was done using two Panasonic AG6300 VHS recorders controlled by a Panasonic A500 edit control board. The stimulus segments were played to the judges through a Panasonic AG6300 recorder via a RCA JD975WV 19 inch color monitor.

Reliability of Encoders' AU Configurations

The reliability of the encoders' configurations was assessed simultaneously for all configurations used in studies 1 and 2. An independent, certified FACS coder who was blind to the configurations required for each hypothesis coded the occurrence of all AUs per segment, the time at which each AU onset, the apex onset, the apex offset, and the AU offset, and the intensity of each AU. The timing data was used to assess if the configurations were overlapping or non-overlapping. A time-date generator which superimposed the time down to 1/60 second on the monitor screen was used to assess the onset and offset times. The author acted as the reliability coder and coded the occurrence of all AUs in every segment, their intensity, as well as assessing whether they were in overlapping or non-overlapping patterns. A random 25% of all final segments were selected to assess the reliability of onset and offset times of all AUs in those segments.

There was 100% agreement regarding the configurations as overlapping or non-overlapping. For calculating the reliability of AU occurrences the formula presented in the FACS manual was used. The formula is:

$$\frac{\text{Total number of agreements X 2 coders}}{\text{Total number of AUs coded}} \text{ X } 100$$

The overall percent agreement for AU occurrence was 98.5%. The overall percent agreement for AU intensity was 67%.

As the timing for the AU configurations was critical for these studies a more stringent than usual margin of error was used to calculate the agreements for onset and offset times. Instead of the margin being $\pm$0.2 seconds, or 12/60 of a second on the time-date generator (Craig & Patrick, 1985; Hyde, 1985; Patrick et al., in press) the margin was set at

±0.17 seconds or 10/60 of a second. The mean percent agreement for AU onset was 94.8%, for apex onset 80.2%, for apex offset 87.8%, and for AU offset 82.5%. The reliability scores for the AU occurrences and for the timing of the AU and apex onsets and offsets were acceptable for the purposes of these studies. The reliability of the intensity scores was deemed too low for inclusion of intensity scores as measures in the present research.

STUDY 1

Study 1 tested the hypotheses regarding the impact of increasing the number of core AUs per facial expression and the role of overlapping apices on judges' ratings of the encoders' portrayed pain. The rationale for combining both hypotheses in the same data collection process was that it was possible to achieve, with the greatest methodological parsimony, the maximum number of comparisons using the largest sample size. These two hypotheses were combined in a single data collection method but have been separated for individual data analysis.

Method

Subjects

Judges

In keeping with the research which defined the AUs investigated in this study, subjects were selected from the university student population. Twenty-five male and 25 female judges were selected from the psychology department's subject pool. The judges were all from first year psychology courses. Six judges were replaced as five missed at least one rating and one misunderstood the instructions. The mean age for males was 18.9 yrs. and for females 18.3 yrs. All judges received either course credit or $3.00 if they had their quota of credits.

The size of the sample of judges was determined according to the formula for estimating the effective reliability (i.e. the reliability of the mean of the judges' ratings) of judges' ratings proposed by Rosenthal (1973. Effective reliability focuses on the reliability of the judges' ratings rather than the items being rated. It is an adaption of the Spearman-Brown formula and is used for estimating the reliability of the mean ratings for different sample sizes of judges. To estimate the mean effective reliability, a mean correlation for a sample of 72 independent pairs of judges was calculated from the data on judges' ratings from Patrick et al. (in press). The mean reliability was *r* = Ø.46. The estimated effective inter-judge reliability for a sample of 5Ø judges with an estimated mean reliability of *r* = Ø.46 was *R* = Ø.98 (*R* = effective reliability).

Encoders

Description of the four encoders is given above in the General Methods.

Procedures for Developing Stimulus Tape

Selection of AU configurations to be tested

Table 2 presents the 23 possible AU configurations ranging from single AUs to combinations of four AUs (AUs 6 and 7 cannot be coded simultaneously so there were no combinations containing both together). With the exception of configurations in which AU 6 occurs alone or with AU 45 all combinations were tested with the AUs in overlapping and non-overlapping apex configurations. Configurations in which AU 6 occurred alone or simply combined with AU 45 were excluded from testing as AU 6 is extremely difficult to

Table 2

Possible AU Configurations

Level of AU Configuration	Possible AU Configurations
Level 1	4[a] 6 7 10 45
Level 2	4+6 4+7 4+10 4+45 6+10 6+45 7+10 7+45 10+45
Level 3	4+6+10 4+6+45 4+7+10 4+7+45 4+10+45 6+10+45 7+10+45
Level 4	4+6+10+45 4+7+10+45

[a] Description of the AUs are as follows:

AU	Muscle	Movement
4	Corrugator	brow lowerers
6	Orbicularis Oculi	cheek raiser
7	Orbicularis Oculi	lid tightener
10	Lavator Labbi Superioris	upper lip raiser
45	Lavator Palebrae	blink

produce voluntarily by itself (Ekman & Freisen, 1978). The configuration 6+45 was excluded as AU 45 (eye blink) does not facilitate the voluntary control of AU 6. That configurations where AU 6 occurred independently were not tested was not a major concern as such configurations appear very infrequently during spontaneous expressions of pain. Of the 360 ten second segments coded in Craig and Patrick (1985) and the 450 three second segments coded in Patrick et al. (in press) AU 6 and 6+45 occurred a total of seven times. This very low frequency would suggest that these configurations are not readily produced by subjects experiencing painful stimulation and thus their exclusion does not hamper generalizations of the results to spontaneous expressions of pain.

To clarify the configurations that were the focus of each hypothesis Tables 3 and 4 list the configurations for hypothesis 1 and 2, respectively.

Construction of stimulus tape

A total of 148 segments were viewed by the judges. There were 4 segments with single AUs, 17 with overlapping apex configurations, 11 with non-overlapping apex configurations, and 5 randomly chosen segments per encoder. Thus there were 4 + 17 + 11 + 5 segments times 4 encoders equalling 148 segments. The randomly chosen segments were used to check intra-judge reliability. The order of presentation of the segments was randomized across encoders and configurations with the limitations that the same encoder was not seen more than three times consecutively and that no

Table 3

AU Configuration for Hypothesis 1

Level of AU Configuration	AU Configurations			
Level 1	4	7	10	45
Level 2	4+6[a]	4+7	4+10	4+45
	6+10	7+10	7+45	10+45
Level 3	4+6+10	4+6+45	4+7+10	
	4+7+45	4+10+45	6+10+45	
	7+10+45			
Level 4	4+6+10+45	4+7+10+45		

[a] A "+" mark indicates that the apices of the AUs overlapped.

Table 4

AU Configurations for Hypothesis 2

Level of AU Configuration	Type of AU Configurations: Overlapping Apices[a]		Type of AU Configurations: Non-overlapping Apices[b]	
Level 2	4+7	4+10	4-7	4-10
	4+45	7+10	4-45	7-10
	7+45	10+45	7-45	10-45
Level 3	4+7+10		4-7-10	
	4+7+45		4-7-45	
	4+10+45		4-10-45	
	7+10+45		7-10-45	
Level 4	4+7+10+45		4-7-10-45	

[a] These configurations are identical to their counterparts listed in Table 3.

[b] A "-" mark indicates that the apices of the AUs did not overlap.

identical configurations were seen consecutively. The segments were five seconds in duration and were separated by a six second blank period during which the judges recorded their ratings. The segments were sequentially numbered by a voice on the video which announced the beginning of each new segment at the end of each blank period.

Procedures for Judges

From one to three judges participated in each session. If three judges were used the third judge sat behind the first two with a clear view of the monitor between the front two judges. This arrangement was used to ensure that the judges had views of the encoders undistorted by the bias of the monitor screen when viewed at extreme angles. The distance of the monitor from the front row was 150cm and the back row 270cm. The seating arrangements were allocated on a first come basis.

Each judge was greeted by the author at the main entrance to the lab and escorted into the viewing room. The room was approximately 3m X 5m. Following the arrival of the last judge the participants were given the rating forms (Appendix A) so that they could follow the rating instructions. The instructions (Appendix B) were then read to the participants. The instructions contained a bogus description of a pain study in which the encoders were purported to have been subjects. Following this description a description of the stimulus tape was given along with instructions on how to use the rating form. Following clarification of any questions a five segment practice session was conducted to ensure that the judges properly understood the instructions. These practice trials were of three encoders who were not on the actual stimulus tape. Following any further clarification the stimulus tape was presented. This presentation lasted approximately 26 minutes. During the viewing of the practice and stimulus tape the incandescent lights were dimmed slightly to eliminate any glare on the monitor screen. The author remained in the room throughout the viewing.

Upon completion of the stimulus tape the rating forms were collected and the judges were given the post-experimental questionnaire (Appendix C). Following the completion of the questionnaire the judges were debriefed (Appendix D).

Measures

As discussed in the literature review there was a question regarding the importance of matching the subjects' self-reported level of pain to the judges' ratings of the subjects' pain. It was suggested at that time that it would be beneficial to use equivalent scales for self-report and judges' ratings to encourage comparisons between these two measures. In particular this would help in the integration of research on the component relationships of the judgment model. In keeping with this suggestion the Gracely pain rating scale of Affective Discomfort (Gracely et al., 1978) was used. This version as well as the 1979 version have been used as self-report measures in studies on facial expressions of pain (Craig & Patrick, 1985; Hyde, 1985; Patrick et al., in press). Keefe and Block (1982) used the Sensory Intensity pain descriptor scale (Gracely et al., 1978) as a judges' rating scale. It was thought that staying with the Affective Discomfort scale offered more opportunity for integration between the prior studies using the Affective Discomfort scale and the present research than between the Keefe and Block study and the present research. In addition to the 15 standard adjectives, which ranged from "distracting" to "excruciating", a baseline item labelled "no discomfort" was added and scored as zero. The adjective list and scaled scores are contained in Appendix E.

Design

The design for hypothesis 1, regarding the impact of increasing the complexity of the facial configurations, was a 2 X 4 factorial design with the first factor as a between subjects factor and the second as a within subjects factor. The factors were sex of judge and level of complexity for overlapping AU configurations, respectively.

The design for hypothesis 2, concerning the role of overlapping AU apices, was a 2 X 2 X 3 factorial design with the first factor as a between subjects factor and the last two as within subjects factors. The factors were sex of judge, type of apex configuration (overlapping vs. non-overlapping), and level of complexity.

Results

Reliability of Judges

The standard inter-judge as well as intra-judge reliabilities were calculated. The inter-judge reliability was calculated using Winer's (1971) formula for estimating inter-judge reliability from ANOVA data. The inter-judge reliability was estimated to be $\underline{r}$ = Ø.94.

The intra-judge reliability was calculated using the 2Ø pairs of identical repeated AU configurations. Three types of reliability estimates were computed. A product moment correlation, percent perfect agreements, and percent agreements ±2 scale unit were calculated for each judge. The correlation coefficient, which is commonly used on such data, must be interpreted with caution in this instance as the pairs of configurations are non-independent. The mean intra-judge reliability coefficient was $\underline{r}$ = Ø.47. A possible reason for the difference between the inter- and intra-judge reliability coefficients is the difference in sample sizes used for computing the correlations. The inter-judge reliability was computed using a sample size of 128 ratings while the intra-judge reliability was based on 2Ø ratings.

The mean percentage and standard deviation of pairs with perfect agreements within judges was 26.4% ± 13.4%. The percent agreement ±2 scaled score was used because the Gracely scale, although built on an interval metric, does not have the adjective descriptors evenly distributed on the scale. There are a number of very fine discriminations between several adjectives which may lead the perfect agreement criterion to misrepresent the judges general consistency. The mean percentage and standard deviations of pairs with ±2 scale unit agreement was 69.2% ± 20.7%.

Post-Experimental Questionnaire Data

This brief questionnaire contained five items, two of which were manipulation checks. The summary data for the first four items are presented in Table 5. Results from question 4, which was the main manipulation check, indicated that judges viewed the configurations as acceptable representations of the expressions they expected for the range of ratings they used. The means for both male and female judges were very similar to those found in Hyde et al. (1985) who used stimulus segments of actual low back pain patients experiencing pain during a range-of-motion test. The fifth question, which was completed contingent upon the judges' response to question 4, indicated that even when judges rated the facial configurations as at least somewhat unrepresentative they usually identified factors other than the configurations themselves as the issue. The most common issues were problems in discriminating between adjacent items on the rating scale, the lack of body movement or sound from the encoders, and having difficulty discriminating amongst certain configurations. The results of question 4 and 5 suggest that the manipulations of the facial configurations were successful in producing stimuli that the judges' found believable.

To evaluate whether any of the items in the questionnaire were related to judges' ratings a multiple regression was conducted using sex of judge and the first

Table 5

Study 1 - Analysis of Post-Experimental Questionnaire Data

Questions		Males	Females
1. General accuracy	Mean	4.20	4.12
2. Types of pain interactions (frequencies per category)	Acute	9	14
	Chronic	2	0
	Both	13	9
	None	1	2
3. Easier to judge males or females (frequencies per category)	Females	6	8
	Males	6	5
	No difference	13	12
4. Representativeness of facial expressions	Mean	3.32	3.88

four questions to predict judges' overall mean ratings. The two factors which predicted judges' overall mean ratings were sex of judge and question 3, whether judges found male or female subjects easier to rate. The multiple *R* was Ø.48. As it was not appropriate to analyze the effect of sex of encoder in this study, and the fact that question 3 only accounted for 1Ø.7% of the variance of judges' overall mean rating, question 3 was not dealt with in further analyses.

<u>Analysis of Hypothesis 1</u>

Hypothesis 1 asserted that increasing levels of configurational complexity would lead to increased judges' ratings of encoder displayed discomfort. Table 6 presents the mean rating and standard deviations for each level of configuration complexity for each sex of judge. The analysis was a 2 X 4 ANOVA (sex of judge X level of complexity) with the mean Gracely scaled score for each level as the dependent measure (Appendix F). Significant main effects were found for sex of judge $F(1,48) = 8.08$, $p < .005$, and level of complexity $F(3,144) = 75.89$, $p < .0001$. The sex of judge X level of complexity interaction was also significant $F(3,144) = 6.01$, $p < .0005$. Figure 2 portrays the interaction.

Due to the significant interaction effect both main effects were reduced to simple effects (Appendix G). The simple effects for sex of judge revealed that female judges rated the third and fourth levels of complexity significantly higher than males. For level 3 the result was $F(1,192) = 11.85$, $p < .001$, and for level 4 $F(1,192) = 14.78$, $p < .001$. The simple effects for level of complexity were significant

Table 6

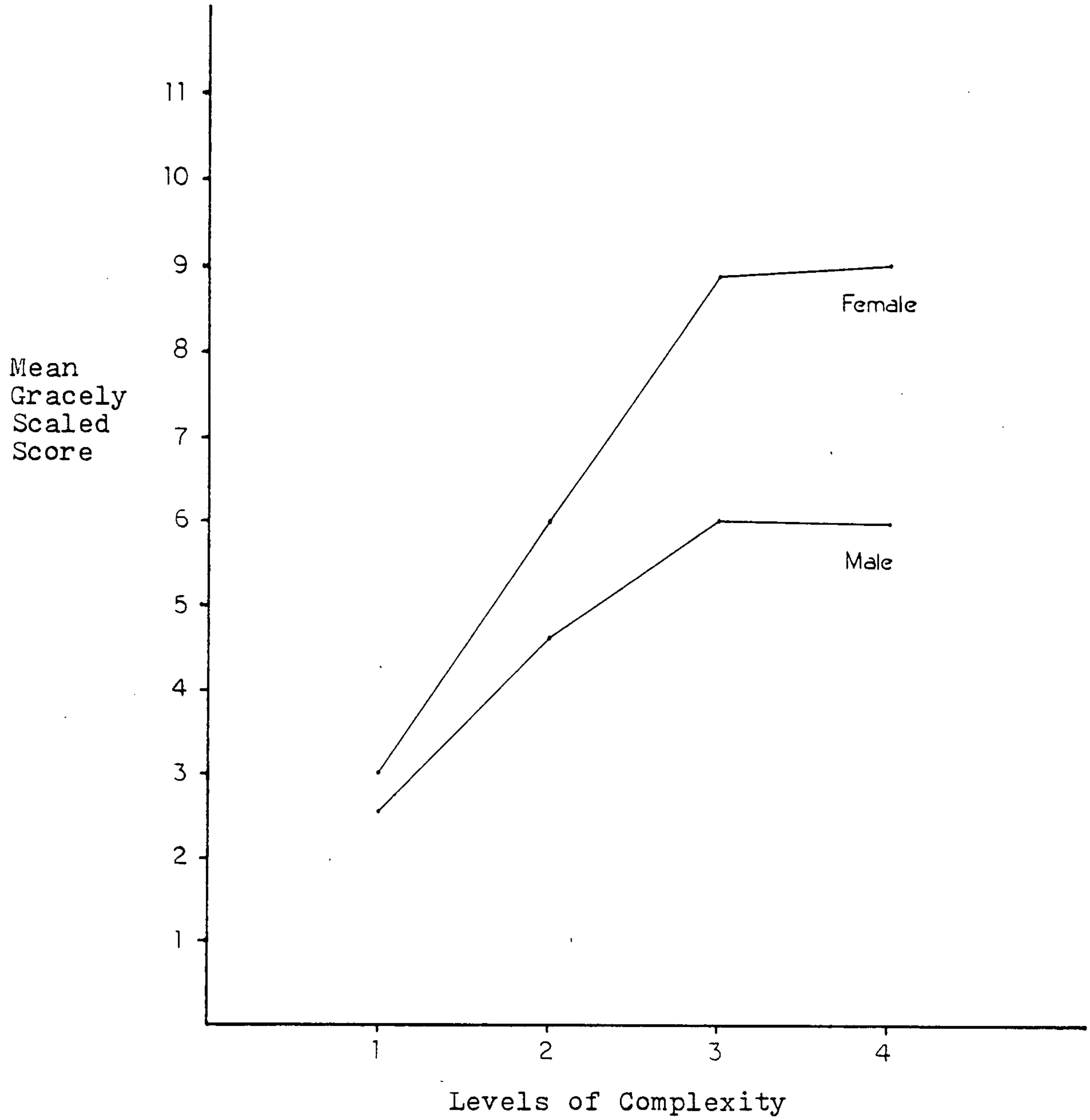
Mean Gracely Ratings for Each Level of AU Configuration by Sex of Judge

	Level of Configuration			
	1	2	3	4
Sex of Judge				
Male				
M	2.59	4.65	6.Ø8	5.93
SD	Ø.69	1.38	2.85	2.5Ø
Female				
M	3.Ø7	6.Ø6	8.91	9.Ø9
SD	Ø.79	2.25	4.52	4.97

Figure 2

Interaction of Sex of Judge

and Levels of Complexity



for both male and female judges. The result for male judges was $F(3,144) = 44.4$, $p < .0001$, and for female judges $F(3,144) = 39.3$, $p < .0001$.

As the simple effects for level of complexity were significant, post hoc comparisons, using Newman-Keuls, were done for each simple effect (Appendix H). The post hoc analyses indicated that for male and female judges all levels of complexity differed significantly from each other except levels 3 and 4. With a combined total of eight simple and post hoc comparisons care should be exercised in accepting the post hoc comparisons suggesting that for male judges level 2 was rated significantly lower than levels 3 and 4 as a more conservative alpha level, which would protect the family-wise error rate, would eliminate the difference.

Figure 3 graphically represents the contribution of each unique AU configuration to the mean level scores. As can be seen the lack of effect between levels 3 and 4 could be due to either the configurations 4+6+10, 4+6+45, and 6+10+45, or some lack of effect of the level 4 configurations.

Analysis of Hypothesis 2

Hypothesis 2 predicted that overlapping AU configurations would be rated as expressing more discomfort than non-overlapping configurations. This was tested over the three levels of configurations which contained multiple AUs. Table 7 presents the mean ratings and standard deviations for each type of configuration across the three levels of complexity for each sex of judge. The analysis was

Figure 3

Distribution of Overlapping AU Configurations

within Levels of Complexity

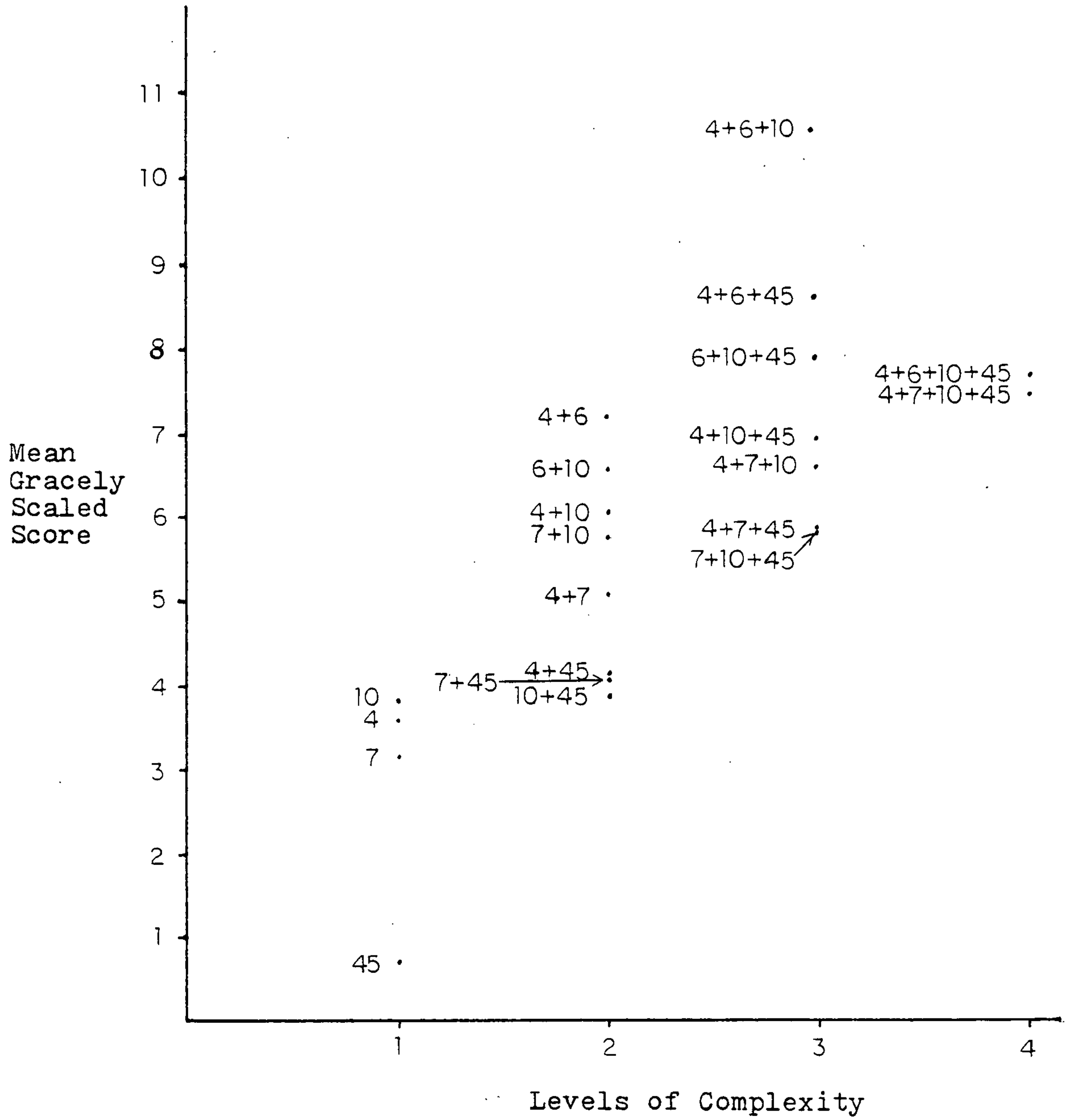


Table 7

Mean Gracely Ratings for Type of Facial Configuration across Levels of Complexity for each Sex of Judge

		Configuration				
	Overlapping[a]			Non-Overlapping		
Level	2	3	4	2	3	4
Male						
M	4.22	5.24	5.89	3.85	4.30	4.52
SD	1.00	2.11	2.77	0.95	1.35	1.65
Sex of Judge						
Female						
M	5.44	7.38	8.96	4.60	5.00	5.68
SD	1.92	3.64	5.06	1.27	1.36	1.98

[a] The means for overlapping configurations do not match those of Table 5 as all configurations containing AU 6 were excluded.

a 2 X 2 X 3 ANOVA (sex of judge X type of configuration X level of complexity) with the mean Gracely scaled score as the dependent measure (Appendix I). Significant main effects were found for sex of judge $F(1,48) = 7.65$, $p < .005$, level of complexity $F(2,96) = 36.85$, $p < .0001$, and type of configuration $F(1,48) = 33.29$, $p < .0001$. All first order interactions were also significant. The results for the first order interactions were configuration X level, $F(2,96) = 14.51$, $p < .0001$, sex of judge X configuration $F(1,48) = 5.84$, $p < .025$, and sex of judge X level of complexity $F(2,96) = 3.94$, $p < .025$. Figures 4, 5, and 6 demonstrate these interactions.

Following the results of this main analysis all main effects were separated into their simple effects for each interaction.

Testing the simple effects within the configuration X level of complexity interaction revealed significant effects for configuration and level of complexity (Appendix J). Configuration was significant at level 3 $F(1,144) = 26.27$ $p < .001$ and level 4 $F(1,144) = 51.51$, $p < .001$. These results indicate that overlapping configurations were rated as demonstrating significantly more discomfort than non-overlapping configurations when the configurations contained three or four AUs. The simple effects for level of complexity were significant for both configurations. For overlapping configurations the effect of levels of complexity was $F(2,96) = 50.66$, $p < .001$ and for non-overlapping configurations $F(2,96) = 5.71$, $p < .025$.

Figure 4

Interaction of Type of Configuration

and Level of Complexity

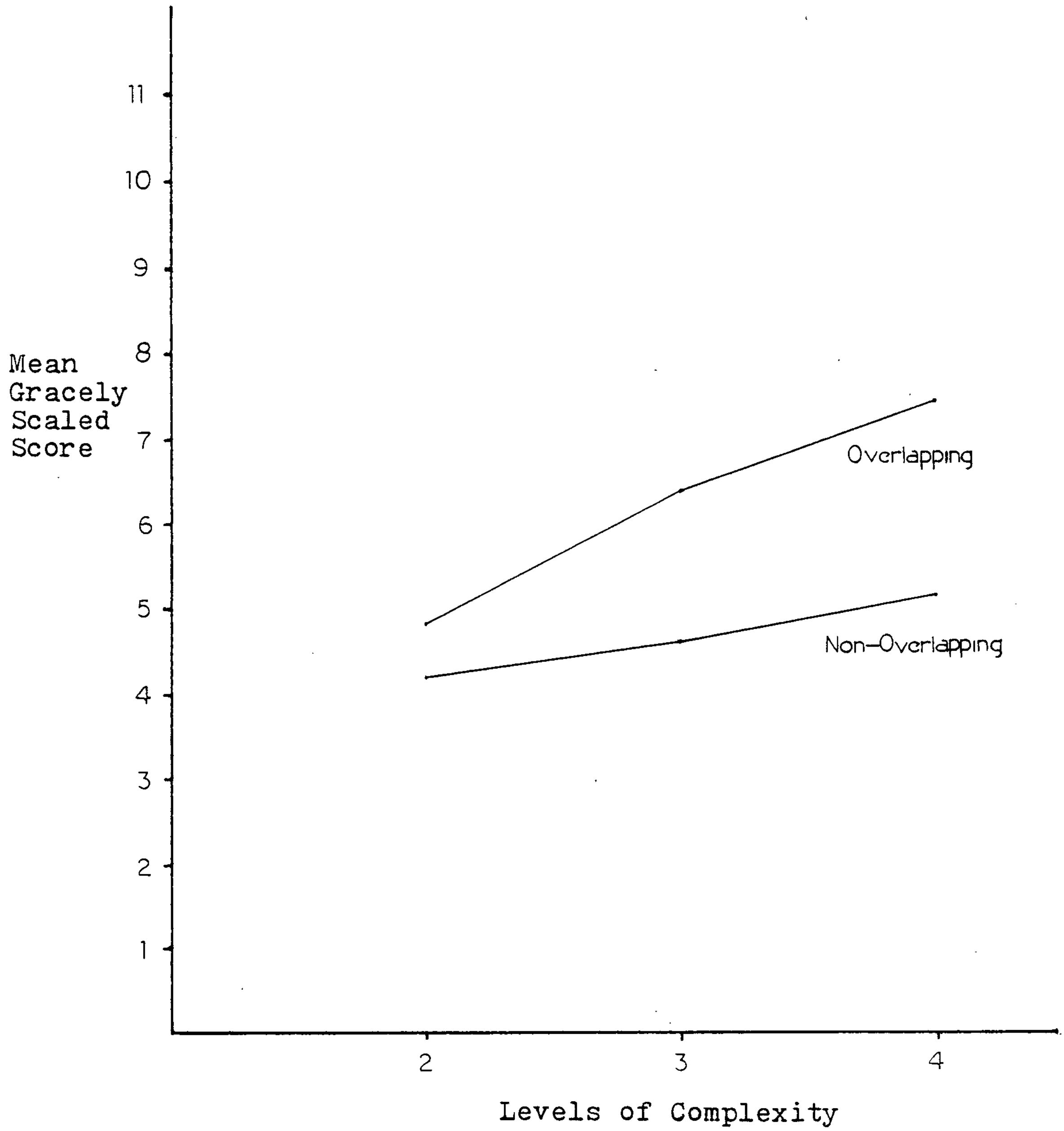


Figure 5

Interaction of Sex of Judge

and Type of Configuration

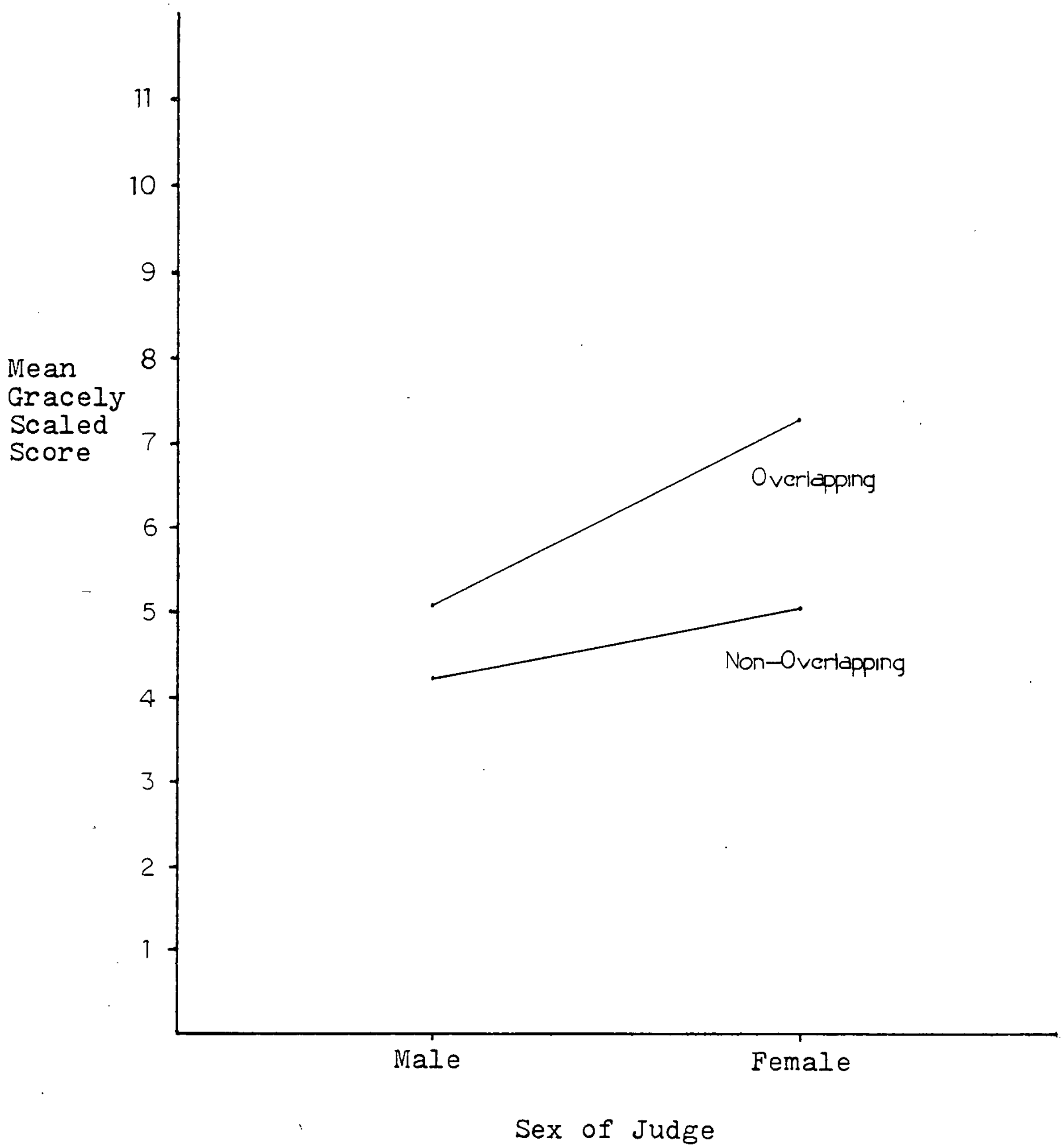
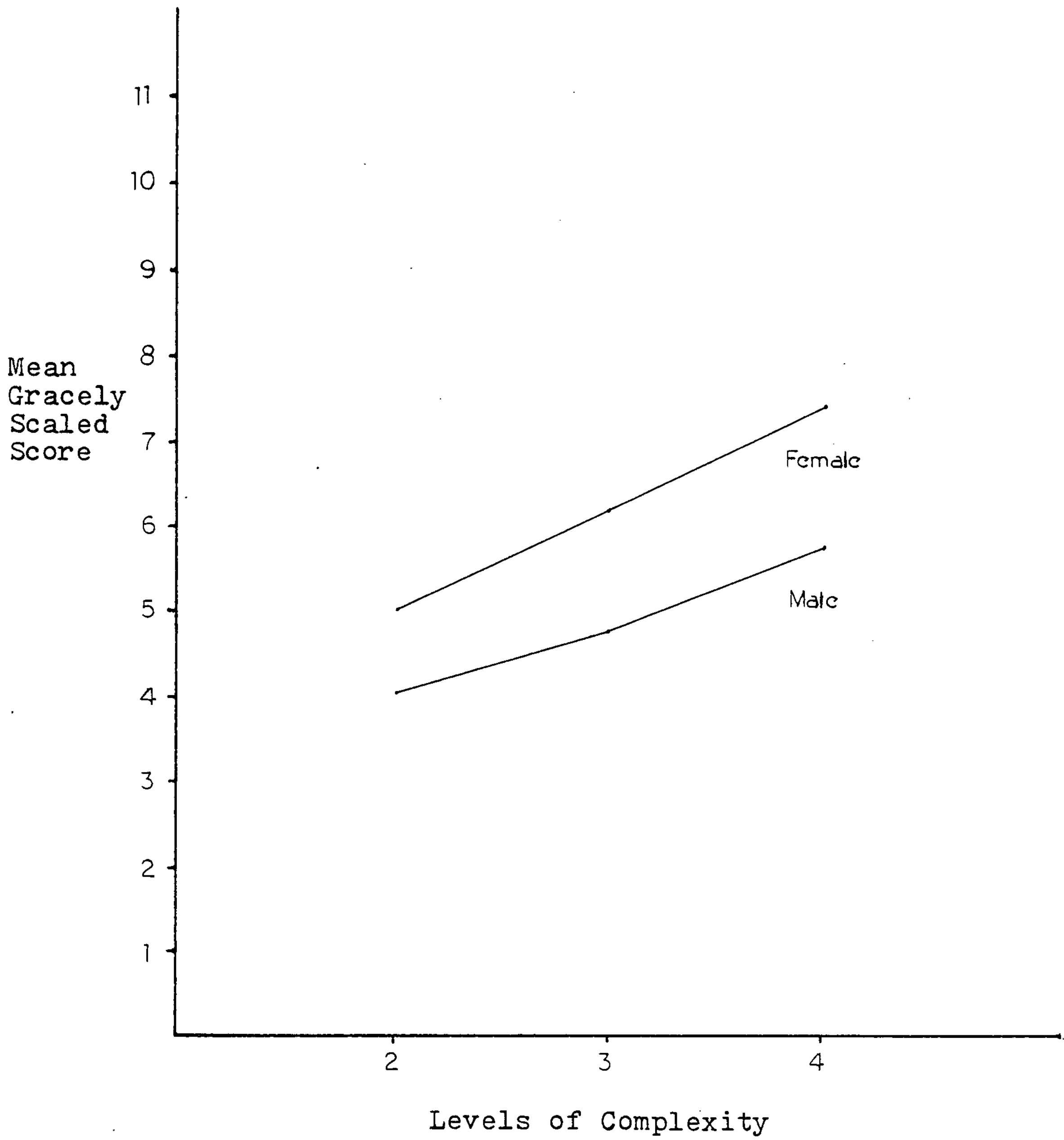


Figure 6

Interaction of Sex of Judge and Level of Complexity



To further explore the effects of level of complexity within each type of configuration post hoc comparisons were computed (Appendix K). For overlapping configurations all levels were significantly different from each other. For non-overlapping configurations it was found that only levels 2 and 4 differed significantly.

Tests on the simple effects for the sex of judge X configuration interaction produced three significant effects (Appendix L). Sex of judge was significant for overlapping configurations $\underline{F}(1,96) = 12.56$, $\underline{p} < .001$. This indicates that the female judges rated overlapping configurations significantly higher than male judges. As the configurations tested here were also tested in hypothesis 1 the result is redundant. It does suggest however, that the differences between male and female judges are not accounted for solely on the basis of configurations containing AU 6 as similar results were obtained with and without this AU being present. The absence of an effect for sex on non-overlapping configurations suggests that neither sex attached as much importance to these configurations as they did to the overlapping configurations.

The simple effects for configuration were significant for each sex of judge. For male judges the result was $\underline{F}(1,48) = 5.62$, $\underline{p} < .025$ and for female judges $\underline{F}(1,48) = 33.50$, $\underline{p} < .001$. Both effects indicate that judges rated overlapping configurations higher than non-overlapping configurations with the effect being much more pronounced for female judges.

The simple effects for the sex of judge X level of complexity interaction demonstrated significant effects for both factors (Appendix M). Sex of judge was significant at level 3 $F(1,144) = 5.73$, $p < .01$ and level 4 $F(1,144) = 12.72$, $p < .001$. The simple effects for level were significant for male judges $F(2,96) = 8.56$, $p < .001$ and female judges $F(2,96) = 32.22$, $p < .001$.

The post hoc comparisons for level of complexity indicated that for male judges level 2 and 4 were significantly different while for female judges all levels were significantly different (Appendix N).

As was the case with hypothesis 1 a large number of simple and post hoc comparison analyses were needed to investigate hypothesis 2. Thus the same caution should be noted regarding the acceptance of differences which show significance at the more liberal alpha levels. The effect of a more conservative error rate would be to reduce the number of significant effects due to level for non-overlapping configurations and for sex of judge effects relating to the second level of complexity.

Summary

The main hypotheses of study 1 focused on the effects of increasing the number of AUs per configuration and whether the type of configuration, overlapping apices or not, would affect judges' ratings of the amount of discomfort portrayed by the encoders. The sex of judge was added as a secondary area of investigation.

Despite the large number of interactions found in testing the two hypotheses several clear results were obtained. First, the level of complexity, as defined by the number of AUs present in any configuration, had a strong effect on judges' ratings. The times when differences between levels were not discriminated were confined to adjacent pairs of levels. Second, when comparing levels of complexity within and between the types of configurations it is apparent that the effect of increasing the complexity is largely confined to overlapping configurations. Overall, the first hypothesis was supported.

It is evident from the differential effect of level of complexity found between types of facial configurations that male and female judges reacted quite differently to the two configurations. Females were found to rate overlapping configurations much higher than males while there was no difference between judges on non-overlapping configurations. In addition to these findings a large difference was found between the male and female judges' discrimination between overlapping and non-overlapping facial configurations. While female judges rated overlapping configurations much higher than non-overlapping configurations, male judges rated overlapping configurations only minimally, but still significantly higher, than non-overlapping configurations. Thus, in general, hypothesis 2 has been supported in that non-overlapping configurations were judged as displaying significantly less discomfort than overlapping configurations for the more complex configurations. However, this general result was less the case for male judges.

Finally, as is apparent in all the comparisons between judges the trend and all significant differences indicate that female judges rated all types and levels of facial configurations higher than male judges. Non-significant differences between sex were only found for the simpler and non-overlapping configurations.

STUDY 2

Study 2 investigated the third and fourth hypotheses. The third hypothesis was whether judges would more often indicate that upper face versus lower face configurations revealed attempts by the encoders to hide their facial reactions to painful stimuli. The fourth hypothesis evaluated the impact of the judgment conditions which differed between the two studies on the judges' ratings. As the procedures were as identical as possible between the two studies the method section is focused mainly on the differences between the two studies.

Method

Subjects

Judges

Twenty-five male and 25 female first year psychology students served as judges. Seven judges were replaced as two indicated they thought the expressions were posed and five others missed at least one rating. The mean age for males was 18.3 yrs. and for females 18.5 yrs.

Encoders

Descriptions of the four encoders were provided in the General Methods.

Procedures for Developing Stimulus Tape

Selection of AU configurations to be tested

In order to have adequate frequencies of upper and lower face AUs for testing it was necessary to use more than just the core-pain related AUs, as AU 10 is the only core pain-related AU that occurs in the lower face. Two more lower face AUs were added from the larger pool of pain-related AUs from which the core AUs were identified as a special subset. AUs 12 and 25 were included in this study. The AUs tested were 4, 6, 7, 1Ø, 12, 25, and 45. Table 8 lists the seven upper and six lower face configurations that were tested.

Construction of stimulus tape

The stimulus tape consisted of 6Ø segments. There were 13 segments X 4 encoders + 8 repeated segments (two per encoder) randomly selected to check intra-judge reliability. The order of presentation of the segments was randomized as in study 1. The segments were three seconds in duration and were separated by a six second blank period. The duration of the stimulus segments was reduced to three seconds to reduce the amount of neutral expression viewed. In study 1 the more complex expressions often took most of the segment duration which kept the tape as a whole looking natural. Equating the two studies on the basis of segment duration was sacrificed to ensure the naturalness of the tape used in the present study. All configurations, however, were seen in their entirety during this study. The segments were sequentially numbered by a voice on the video tape.

Procedures for Judges

The basic procedure was identical to that of study 1. The instructions (Appendix O) contained the following alterations. The description of the bogus pain experiment was changed. The description now was of a two part

Table 8

Upper and Lower Face AU Configurations

Upper Face Configurations

Level 1	4[a]	7	45	
Level 2	4+6	4+7	4+45	7+45

Lower Face Configurations

Level 1	10	12	25
Level 2	10+12	10+25	12+25

[a] Description of the AUs are as follows:

AU	Muscle	Movement
4	Corrugator	brow lowerers
6	Orbicularis Oculi	cheek raiser
7	Orbicularis Oculi	lid tightener
10	Lavator Labbi Superioris	upper lip raiser
12	Zygomatic Major	lip corner pull
25	Depressor Labi	lips part
45	Lavator Palebrae	blink

experiment where, in the first part, the subjects spontaneously responded to the pain stimuli while, in the second part, subjects had been instructed to try and hide their facial expressions to an identical set of stimuli. The rating instructions were also altered. They now detailed how to first make a judgment of whether the encoders were or were not trying to hide their facial expressions and then how to rate the configuration using the Gracely pain rating scale of Affective Discomfort. The rating form is contained in Appendix P. The judges were given five practice trials. Presentation of the stimulus tape lasted approximately 9 minutes.

Upon completion of the stimulus tape the rating forms were collected and the judges were given the post-experimental questionnaire. Following the completion of the questionnnaire the judges were debriefed (Appendix Q).

Measures

For each segment the judges classified each configuration as being either a spontaneous expression (not hidden) or as an attempt by the encoder to hide their facial expression (hidden). The judges then used the Gracely pain scale of Affective Discomfort to rate the amount of discomfort exhibited by each encoder.

Design

The design for the third hypothesis was a 2 X 2 X 2 factorial with the first factor as a between subject factor and the last two as within subject factors. The factors were sex of judge, area of AU configuration (upper or lower face), and level of configuration complexity. For the fourth hypothesis the design was a 2 X 2 X 2 factorial with the first two factors as between subject factors and the last as a within subject factor. The factors were sex of judge, type of judgment condition (study 1 or study 2), and level of configuration complexity.

Results

Reliability of Judges

Inter-judge and intra-judge reliabilities were calculated for both the judging of the configurations as hidden or not hidden as well as the ratings of discomfort. The inter-judge reliability estimate for the judgment of configurations as hidden or not hidden was r = 0.84. The inter-judge reliability estimate for the Gracely ratings was r = 0.90. This inter-judge reliability for the Gracely ratings was close to that found in study 1.

For calculating the intra-judge reliability for the hidden/not hidden judgments it was not advisable to calculate the correlation type reliability coefficient. Having a small sample size (eight pairs of judgments) and dichotomous data produced very skewed distributions in the data which makes correlation coefficients such as phi or kappa quite biased (House, House & Campbell, 1981). Thus only the mean percent agreement was used. The mean percent agreement was 67.5%.

The mean correlation, mean percent perfect agreement, and mean percent agreement ± 2 scaled score were computed for the intra-judge reliability of the Gracely ratings. The mean correlation was r = Ø.45. The mean percentage scores and standard deviations were 23.2% ± 19.2% for perfect agreements and 65.Ø% ± 2Ø.Ø% for agreements ±2 scaled score. The reliability coefficient and percent agreements are similar to the those reported in study 1.

Post-Experimental Questionnaire Data

The questionnaire was identical to that used in study 1. Summary data for the first four items are presented in Table 9. Results from questions 4 and 5 which were manipulation checks again supported the position that the judges thought the facial configurations were believable. In this study the same comments were found in question 5 as in study 1. In addition, several judges indicated that they felt rushed in making their ratings.

As in study 1 a multiple regression was conducted, using sex of judge and the first four questions as predictors. In this study the dependent measure was the total frequency of judges' rating a configuration as not hidden. Only sex of judge proved to be predictive. The multiple R was Ø.33. Due to the lack of predictive power of the questionnaire items they were not included in any further analyses.

Analysis of Hypothesis 3

It was proposed in hypothesis 3 that more upper face than lower face configurations would be judged as attempts by the encoders to hide their facial expressions. Table 1Ø gives the mean proportion of times the upper and lower face configurations were judged as hidden and their standard deviations broken down by level of complexity and sex of

Table 9

Study 2 - Analysis of Post-Experimental Questionnaire Data

Questions		Males	Females
1. General accuracy	Mean	3.92	3.36
2. Types of pain interactions (frequencies per category)	Acute	11	13
	Chronic	1	0
	Both	6	4
	None	7	8
3. Easier to judge males or females (frequencies per category)	Females	10	11
	Males	6	4
	No difference	9	10
4. Representativeness of facial expressions	Mean	3.96	4.24

Table 1Ø

Mean Proportion of Times Configurations of the Upper and Lower Face were Judged as Attempts to Hide

	Areas of Face			
	Upper		Lower	
Level	1	2	1	2
Male				
M	Ø.38[a]	Ø.54	Ø.52	Ø.72
SD	Ø.17	Ø.18	Ø.2Ø	Ø.19
Sex of Judge				
Female				
M	Ø.48	Ø.64	Ø.58	Ø.73
SD	Ø.25	Ø.16	Ø.19	Ø.23

[a] The lower the proportion the more the configurations were judged as attempts by encoders to hide their expressions.

judge. The lower the score the more frequently the configurations were rated as hidden. The analysis was a 2 X 2 X 2 ANOVA (sex of judge X facial area X level of complexity) with the mean proportion of times the configurations were rated as hidden as the dependent variable (Appendix R). All three main effects were significant with no significant interactions. The main effects were for sex of judge $F(1,48) = 5.87$, $p < .025$, for area of the face $F(1,48) = 10.07$, $p < .005$, and for level of complexity $F(1,48) = 58.04$, $p < .0001$.

Using these results to interpret Table 10 indicates that at each level of complexity the upper face configurations were more often judged as portraying attempts to hide facial expressions than lower face configurations. Male judges scored more configurations as hidden than female judges.

Analysis of Hypothesis 4

Hypothesis 4 served a dual purpose. It provided an opportunity for a limited replication of hypothesis 1 and allowed for exploratory testing of the possible effects of differing judgment conditions on judges' ratings of the encoders' expressed discomfort. As detailed in study 1 the judges were led to believe that the encoders were receiving painful stimuli and the judges were instructed to rate the encoders' portrayed discomfort. In study 2 the judges were led to believe that the encoders were either being spontaneous or were trying to hide their facial expressions. Judges were then asked to make a judgment as to whether the encoder was hiding their expression or not and then rate the encoders' level of discomfort. Thus the judgment conditions (instructional context and required judgments) differed between the studies. The segment durations were also different. They were five seconds in study 1 and three seconds in study 2.

Table 11 lists the eight facial configurations common to both studies. Table 12 presents the mean Gracely ratings and standard deviations for each level of complexity by sex of judge and type of judgment condition (study 1 or 2).

The analysis was a 2 X 2 X 2 ANOVA (sex of judge X type of judgment condition X level of complexity) with the mean Gracely scale scores for each level as the dependent measure. (Appendix S). Significant main effects were found for judgment condition $\underline{F}(1,96) = 17.36$, $\underline{p} < .0001$ and level of complexity $\underline{F}(1,96) = 123.20$, $\underline{p} < .0001$. The sex of judge X judgment condition interaction was also significant $\underline{F}(1,96) = 8.29$, $\underline{p} < .005$. Figure 7 illustrates this interaction.

The sex of judge X judgment condition interaction was broken into its simple effects (Appendix T). Two significant simple effects were found. The judgment condition for male judges was significant $\underline{F}(1,96) = 12.41$, $\underline{p} < .001$ and the sex of judge effect was significant for the study 2 judgment condition $\underline{F}(1,96) = 6.63$, $\underline{p} < .025$. These effects reveal that male judges rated configurations significantly higher in study 2 than study 1. As well, male judges rated configurations within study 2 significantly higher than female judges.

Table 11

AU Configurations Common to Studies 1 and 2

Level of Configuration Complexity	AU Configurations			
1	4	7	1Ø	45
2	4+6	4+7	4+45	7+45

Table 12

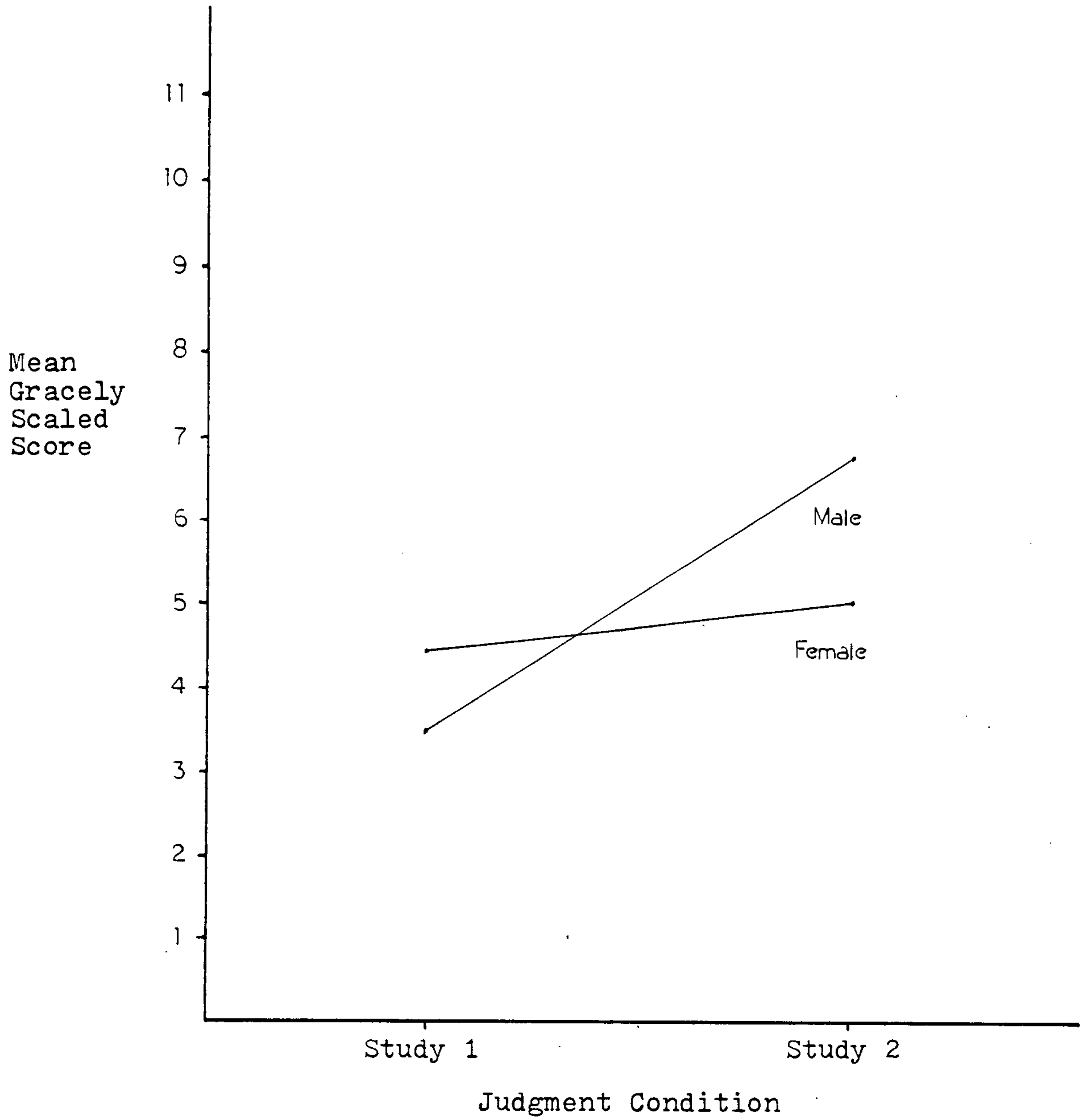
Mean Gracely Ratings for each Level by Sex of Judge and Judgment Condition

Sex of Judge	Judgment Condition		Level 1	Level 2
Male	Study 1	M	2.59	4.41
		SD	Ø.69	1.34
	Study 2	M	5.52	7.89
		SD	2.93	4.36
Female	Study 1	M	3.Ø7	5.85
		SD	Ø.79	2.48
	Study 2	M	3.96	6.14
		SD	1.33	3.4Ø

Figure 7

Interaction of Judgment Condition

and Sex of Judge



Summary

The results dealing with hypothesis 3 clearly support the position that judges view the upper rather than the lower face configurations as indicators of an encoder's attempt to hide their facial expression. Furthermore, the data suggests that single AUs are more likely to be judged as indicative of an encoder trying to hide their expression than configurations with two AUs. This may imply that judging a configuration as indicating control over the expression by the encoder is somewhat dependent upon the amount of facial movement exhibited. Of exploratory interest is the result that male judges perceived more configurations as suggesting attempts to hide than female judges.

Results for hypothesis 4 were consistent with the research on the A-C relationship which indicated that instructions to judges led them to alter their ratings of the subjects' pain (Patrick et al., in press; Prkachin et al., 1983). The partial replication of hypothesis 1 was also successful. The level of complexity of facial configurations would appear to be an imporant factor affecting judges' ratings of an encoder's facial expression of discomfort. Finally, the finding that the type ofjudgment condition had a strong differential effect on the ratings of male and female judges, with females rating higher under the straight rating condition but males rating higher under the dual judgment condition, may indicate that the sex of judge effect is open to various environmental manipulations.

GENERAL DISCUSSION

The aim of this research was to study the effects of specific facial configurations on judges' ratings of encoder displayed discomfort. The findings from the two studies revealed several factors which seem to be important in determining a judge's rating of a person's discomfort from viewing their facial expressions.

Level of Complexity

There were two very consistent findings. The most consistent factor was the level of complexity of the facial configuration portrayed. In all cases, except with non-overlapping configurations, there was a steep increase in the ratings of discomfort related to increasing levels of complexity. Only when the configurations were quite complex did the ratings of discomfort not discriminate between levels of configuration complexity. The limited replication of the level of complexity effect done in study 2 supports the robustness of this effect. By comparing the results of the overlapping and non-overlapping configurations it is evident, due to the relatively small effect of non-overlapping configurations, that the critical issue is the judges' interpretation of the configuration pattern (i.e. overlapping apices or not) and secondarily the absolute number of AUs contained in each video segment.

This finding has two important implications. The first is that attention must be paid to the pattern of the facial expressions related to experiencing pain. Patrick et al. (in press) used frequency of occurrence of individual AUs for most of their data analysis but also included information on the correlation of AUs within segments and the frequency of co-occurrences of the significant AUs. Hyde (1985) used factor analysis and also addressed the issue of AU co-occurrence when analyzing her data on low back pain patients. The extensive analyses of these prior studies present a much clearer picture of the complex relationship between facial activity and the experience of pain than simply focusing on one measure such as frequency of AU occurrence.

The second implication is concerned with the possible role of other AUs. As noted in prior research (Craig & Patrick, 1985; Hyde, 1985; Lanzetta & Kleck, 1970; Patrick et al., in press) there were large inter-individual differences in facial behavior even within similar levels of self-reported discomfort. While a small subset of AUs have been identified as important indicators of discomfort, additional variance in self-reported discomfort may be accounted for by studying patterns of configurations that may or may not contain some or all of the AUs studied in the present research. Thus seeing an overlapping configuration of AUs 4 + 9 (nose wrinkle) + 23 (lip tightens) may indicate more discomfort than a non-overlapping configuration of AUs 4-10-45, even though the latter configuration has more pain related AUs.

Sex of Judge

The sex of judge effect was the other consistent factor throughout both studies. It was consistent in that male and female judges' ratings were found to be significantly different at some point for all major analyses. However, there was little consistency in the direction of the difference. As indicated in the results from testing hypothesis 4 there was a dramatic reversal in the sex of judge effect. Female judges rated facial configurations higher in study 1 but males rated them higher in study 2. Of special note is that the sex of judge effects are inconsistent with some prior research (Kleck et al., 1976; Prkachin et al., 1983) which found no difference between male and female judges' ratings. The difference between the prior research and the present studies cannot be attributed to the selective nature of the configurations used in the present case as sex of judge effects were also found in a recently completed study (Hyde et al., 1985) in which judges' ratings of back pain patients' spontaneous facial expressions of pain were examined.

Stark-Adamec and Kimball (1984) in their discussion of research on sex differences and Hall (1979) in her review of sex differences found in judgment studies clearly pointed out that it is unlikely that it is the sex of the judge which is the issue. Rather it is most likely one or more of a number of developmental or social factors which are the important determinants and the use of a sex of judge variable is a general shorthand method of tapping these underlying factors.

Hall (1984) compared six hypotheses that have been developed to account for the sex differences in judgment accuracy. Of the six hypotheses only the practice/attention and accommodation hypotheses seem promising. The accommodation hypothesis, which was developed by Rosenthal and DePaulo (1979), appears to offer an explanation of the sex of judge findings in the present research. As stated by Rosenthal and DePaulo:

> If it is disruptive of smooth interpersonal functioning for a participant to 'know too much' about the state of the other, we would expect women to show relatively less advantage over men in decoding nonverbal cues when those cues are under less control of the sender and more likely to be unintended than intended cues. (p.69).

One of the major differences between study 1 and 2 were the instructions. In study 1 the focus was on trying to accurately rate the encoders' discomfort from viewing the encoders' facial expressions. The assumption behind such instructions was that the encoders' were openly communicating their actual state. In study 2 the instructions clearly focused on the possibility that the encoders' may have been trying to hide facial cues about their state. In such circumstances any facial activity could have been viewed as unintended and not indicative of what the encoder was trying to communicate, even if it did suggest something about the state of the encoder. Following the accommodation hypothesis it would have been predicted that female judges would have attended more than male judges to what the encoder attempted to communicate (i.e. no pain) than to leaked cues that were contradictory to the attempted nonverbal communication. Thus in study 1 there would have been no disruption in the possible flow of an interpersonal relation with the encoder due to the judge accurately assessing the encoders' state. However, in study 2, accurate perception of the encoders' state might have been disruptive as attention would have had to be paid to the perceived contradiction and the judge would have had to ask for clarification to better understand the communicative intent of the encoder.

The line of reasoning offered by the accommodation hypothesis suggests that the reversal of the sex of judge effect is open to a consistent explanation. However, as Hall (1984) pointed out much more work is needed on the accommodation, as well as other hypotheses on the differences between male and female judges before confidence can be taken in any possible explanation.

Type of Configuration

When initially postulating an effect of configuration type (overlapping vs. non-overlapping) there were two possible positions to take. The first position assumed that only "facial expressions" would have been rated by judges as displaying discomfort. The key element to defining a configuration as an expression was that it contained AUs with overlapping apices. Non-overlapping configurations would thus not have been perceived as expressions and would have been rated as indicative of little or no pain. The second position assumed an additive function which judges would have used to add up the overall effect of seeing several AUs. The only difference between two segments with identical AUs would have been the difference between seeing the AUs in an overlapping or non-overlapping pattern. Thus a strong level effect would have been proposed for both types of configurations with the overlapping configurations possibly being rated higher at each level due to the added effect of overlapping apices.

The actual data from examining the types of configurations does not unequivocally support either position. There was only a very weak level of complexity effect for non-overlapping configurations. However, this weak effect plus the fact that all mean levels of non-overlapping configurations were higher than the mean rating of single AU configurations suggests that there was some effect due to simply seeing more AUs. It is apparent that overlapping configurations are much easier for judges to discriminate and also lead to much higher ratings of discomfort.

While the present research found that the pattern of the facial configuration was an important factor in judges' ratings, particularly for more complex configurations, only two of many possible patterns were tested. Besides other variations in overlapping configurations research on the duration and intensity of the AUs within various configurations needs to be done to evaluate their role as cues for observers.

<u>Hidden vs Not Hidden Facial Configurations</u>

It was found that judges made discriminations between perceived attempts by the encoders to hide or not hide their facial expressions of pain. This discrimination was based on the area of the face containing the facial activity. Judges discriminated between upper and lower face configurations when deciding if the encoder was attempting to hide their expression of discomfort. The rating of upper face activity as more often indicating an attempt by the encoder to hide their state was maintained across both levels of complexity that were tested.

While the results are consistent with the notion of upper face movements being less controllable, and thus are more likely to be cues available to judges regarding a person's state, the present research cannot directly support the idea that the distinction is based solely on the area of the face. The reason there is no direct support is that it was not possible, in the present research, to separate the effects of facial area from the effects of the different AUs in the upper and lower face. What is known is that judges make discriminations between hidden and not hidden configurations consistent with an upper and lower face discrimination.

Ekman and Friesen (1969a; 1969b; 1975) described four methods which people may use to control their facial behavior: 1) intensifying, 2) deintensifying, 3) neutralizing, and 4) masking. Prior work has focused on masking (i.e. dissimulation) facial expressions of emotional (Ekman & Friesen, 1969b; 1974; Ekman et al., 1980) and painful states (Lanzetta et al., 1976). Lanzetta et al. also studied the neutralizing method. Study 2 investigated the effects of telling judges the encoder used the neutralizing method in order to study what cues the judges used to decide when the encoders had attempted to neutralize their expressions. As outlined above it was found that judges more often used upper than lower face actions as cues of an encoders' attempt to hide (neutralize) their reactions to painful stimuli. This finding, and those on dissimulation which looked at distinguishing areas of the face, are consistent with the physiological data (Rinn, 1984) which indicated that upper face actions are under less voluntary control than lower face actions and thus may be more useful as cues of the actual state of the encoder.

The work on the masking and neutralizing methods indicate that the critical issue is the degree of voluntary control of the facial muscles. Thus if research was to be done on the intensifying and deintensifying methods it would seem likely that the results might indicate that the methods work best on lower face muscles, leaving the upper face movements as clearer indicators of the encoder's actual state.

<u>Placing the Present Research within the Judgment Model</u>

As discussed earlier the present research focused on the B-C relationship in Rosenthal's model. This type of research is used to study the relationship between an encoder's nonverbal behavior and the judge's rating of the encoder's state. A comparison of the present research with the B-C research listed in Table 1 indicates that this is the first study to systematically manipulate known and verifiable facial behaviors to see how various manipulations affected judges' ratings of the encoders' displayed painful discomfort.

A tentative connection can be drawn between the present experiments and prior A-C research. Patrick et al. (in press) obtained a multiple regression equation which contained most of the core AUs used in the first study. The results from hypothesis 1 provide experimental support for Patrick et al.'s equation which suggested that AUs 4, 6, 10, and 45 were associated with judgments of increased pain. The present research suggests that this equation should be weighted to give more weight to the AUs when they overlap than to their simple occurrence within any segment of time.

The present research also indicates the importance of using sophisticated coding systems when studying the A-B relationship (Craig & Patrick, 1985; Hyde, 1985; LeResche, 1982; Patrick et al., in press). Without such research it would not have been possible to investigate the B-C relationship with the specificity demonstrated in the present study. This inter-dependence of one type of research on another highlights the utility of Rosenthal's model for clarifying issues related to conducting judgment studies and the role of sophisticated coding systems, such as FACS, in research on judgments and facial expressions.

Facial Activity as a Nonverbal Measure of Pain

Craig and Patrick (1985), Hyde (1985), and Patrick et al. (in press) have all demonstrated one method for measuring facial expressions of pain. These studies all used FACS and found, using highly trained coders, that a small subset of AUs were associated with various types of self-reported discomfort. However, while effective, using FACS is a very costly and time consuming procedure which all but limits its use to research purposes only. A more practical method would involve the use of judges (i.e. minimally trained observers) rather than highly trained coders. Ganchrow et al. (1978) and Keefe and Block (1982) devised coding systems that are more practical in applied settings. The present study can be seen as a first move towards synthesizing the highly technical coding of FACS with the practical limitations of applied settings.

To give health care staff the knowledge about empirically demonstrated nonverbal behavior (e.g. facial expressions) related to pain discomfort and the skill to observe such behavior is recognize as an important goal (Keefe et al., 1984; LeResche & Dworkin, 1984; Teske et al., 1983). It is now clear from the present studies that naive judges can make clear discrimination amongst an array of subtly different facial configurations that are known to relate to different levels of self-reported pain. Given this relationship the ground work has been established for researchers to investigate the possibility of teaching judges to accurately observe these facial configurations.

The reliability data for both studies indicated that when judges were given numerous occassions for making ratings that their mean ratings tended to be quite reliable. These results would suggest that if judgment of facial activity was to be used as a measure of painful discomfort then the measure should be averaged over several judges or over one judge's repeated ratings.

Limitations of the Present Studies

There are a number of limitations to the present study. The two most obvious ones are that only four encoders were used and there were no comparisons made between pain-related AU configurations and other types of facial configurations. Both of these limitations were due, in part, to the large number of configurations needed to test all the combinations of pain-related AUs. Obtaining skilled encoders was also a problem. To increase the number of encoders would require placing limitations on the number of configurations used and possibly require the selection of easily reproducible AUs. While prototypic expressions for a number of emotions have been described (Ekman & Friesen, 1978a; Izard, 1979) they generally lack the repeated empirical research which supports the validity of the described configurations. This would leave a comparative study between the pain-related AUs and other facial configurations on difficult ground.

There are also various other types of configurations which could be tested. The present study focused on two extreme cases. More subtle alterations of the overlapping and non-overlapping configurations could lead to a fuller understanding of the important elements which judges used when making their differential rating of overlapping and non-overlapping configurations. Furthermore, the issues of intensity and timing of the AUs and facial configurations need to be more fully explored. Although both factors can be measured there is no research on their impacts on judges' ratings or in relation to subjects' self-report. Both factors were measured for reliability purposes in the present research but were not included in further data analysis. Until the reliability of coding intensity increases to an acceptable level intensity will remain an unstudied area. While the reliability for coding the timing of the onset and offset of AUs and their apices was acceptable there were problems in deciding how to use such measures in analyzing the data. For the two types of configurations used in the present research there were five different aspects of the facial configurations for which measures of duration could be derived. These measures have not been investigated and so their utility in explaining the role of facial expressions is unknown. The designs for the present studies did not lend themselves to investigating the different measures of timing.

Finally, there is the limitation of the type of judges used. For the purpose of moving towards the development of a nonverbal measure of pain an important issue is what AU configurations do health care personnel use as indicators when rating facial expressions of pain. The present studies have demonstrated that it is possible to systematically investigate pain-related AUs but did not address the issue of how health care personnel rate such AU configurations. With the strong but inconsistent sex of judge effects and the effect of judgment conditions it would not appear safe to generalize the present results to a very different population of judges or judgment conditions.

Conclusion

To conclude three points will be summarized. First, the present research succeeded in demonstrating that direct manipulation of facial configurations can be used to study a variety of hypotheses regarding the communicative value of nonverbal facial movements. Such research is totally dependent upon prior research which identifies the facial movements which are important for a particular area of investigation. Second, the two studies established that the level of configuration complexity, type of configuration, and judgment condition all affected the judges' ratings of the encoders' portrayed discomfort. Furthermore, judges tended to use upper face AUs as indicators of an attempt to hide an expression when the judges believed some of the configurations were attempts by the encoders to hide their expressions. Finally, due to the present study being the first to use direct manipulation of pain-related facial movements further research will be needed to assess whether the encouraging results of these studies prove useful in the ongoing attempts to develop better measures of human pain.

REFERENCES

Anderson, J. E. (1983). Grant's atlas of anatomy (8th ed.). Baltimore: Williams & Wilkins.

Archer, D., & Akert, R. M. (1977). Words and everything else: Verbal and nonverbal cues in social interpretation. Journal of Personality and Social Psychology, 35, 443-449.

Bandler, R. J. Jr., Mararas, C. R., & Bem, D. (1986). Self-observation as a source of pain perception. Journal of Personality and Social Psychology, 9, 205-209.

Bandura, A. (1977). Social learning theory. Englewood Cliffs, N.J.: Prentice-Hall.

Bandura, A., & Rosenthal, T. L. (1966). Vicarious classical conditioning as a function of arousal level. Journal of Personality and Social Psychology, 3, 54-62.

Beecher, H. K. (1959). Measurement of subjective responses: Quantitative effects of drugs. New York: Oxford University Press.

Birdwhistell, R. (1952). Introduction to kinesics. University of Louisville Press.

Blanchard, E. B. (1981). Behavioral assessment of psycho-physiological disorders. In D. H. Barlow (Ed.), Behavioral assessment of adult disorders (pp. 239-270). New York: Guildford Press.

Block, A. R. (1981). An investigation of the response of the spouse to chronic pain behavior. Psychosomatic Medicine, 43, 415-422.

Block, A. R., & Boyer, S. L. (1984). The spouse's adjustment to chronic pain: Cognitive and emotional factors. Social Science and Medicine, 19, 1313-1317.

Bonica, J. J. (1983). Pain research and therapy: Achievements of the past and challenges of the future. In J. J. Bonica, U. Lindblom, & A. Iggo (Eds.), Advances in pain research and therapy: Proceedings of the third world congress on pain: (Vol. 5, pp. 1-36). New York: Raven Press.

Boucher, J. D. (1969). Facial displays of fear, sadness, and pain. Perceptual and Motor Skills, 28, 239-242.

Brena, S. F. (1978). The staggering cost of chronic pain. In S. F. Brena (Ed.), Chronic pain: America's hidden epidemic (pp. 1-10). New York: Atheneum.

Bresler, D. E. (1979). Free yourself from pain. New York: Simon & Schuster.

Cacioppo, J. T., Marshall-Goodell, B. S., & Gormezano, I. (1983) Social psychophysiology: Bioelectrical measurement, experimental control, and analog-to-digital data acquistion. In J. T. Cacioppo, & R. E. Petty (Eds.), Social psychophysiology: A sourcebook (pp. 666-690). New York: Guilford Press.

Chambers, W. G., & Price, G. (1967). Influence of nurse upon effect of analgesic administered. Nursing Research, 16, 228-233.

Chapman, C. R., Casey, K. L., Dubner, Foley, K. M., Gracely, R. H., & Reading, A. E. (1985). Pain measurement: An overview. Pain, 22, 1-31.

Chapman, W. P., & Jones, C. M. (1944). Variations in cutaneous and visceral pain sensitivity in normal subjects. Journal of Clinical Investigations, 23, 81-91.

Chudler, E. H., & Dong, W. K. (1983). The assessment of pain by cerebral evoked potentials. Pain, 16, 221-244.

Cline, V. B., Atzel , J., & Holmes, E. (1972). Assessing the validity of verbal and nonverbal cues in accurately judging others. Comparative Group Studies, 3, 383-394.

Craig, K. D. (1978). Social modeling influences on pain. In R. A. Sternbach (Ed.), The psychology of pain (pp. 73-109). New York Raven Press.

Craig, K. D. (1980). Ontogenetic and cultural influences on the expression of pain in man. In H. W. Kosterlitz, & L. Y. Terenius (Eds.), Pain and Society (pp. 39-52). Weinheim: Verlag Chemie Gmbh.

Craig, K. D. (1984). Emotional aspects of pain. In P. D. Wall, & R. Melzack (Eds.), Textbook of Pain (pp. 153-161). Edinburgh: Churchill/Livingstone.

Craig, K. D., & Patrick, C. J. (1985). Facial expression during induced-pain. Journal of Personality and Social Psychology, 48, 1080-1091.

Craig, K. D., & Prkachin, K. M. (1980). Social influences on public and private components of pain. In I. G. Sarason, and C. D. Spielberger (Eds.), Stress and anxiety (Vol. 7 pp. 57-72). New York: Hemisphere.

Craig, K. D. & Prkachin, K. M. (1983). Nonverbal measures of pain. In R. Melzack (Ed.), Pain measurement and assessment (pp. 173-179). New York: Raven Press.

Darwin, C. (1872). The expression of the emotions in man and animals. London: John Murray.

Diamond, C. (1979). Anatomy and physiology. In C. Diamond, & I. Frew (Eds.), The facial nerve (pp. 1-36). Oxford University Press.

Dimatteo, R. R., Friedman, H. S., & Taranta, A. (1979). Sensitivity to bodily nonverbal communication as a factor in practitioner-patient rapport. Journal of Nonverbal Behavior, 4, 18-26.

Dunlap, K. (1927). The role of eye-muscles and smooth muscles in the expression of the emotions. Genetic Psychology Monographs, 2, 199-233.

Ekman, P. (1965). Communication through nonverbal behavior: A source of information about an interpersonal relationship. In S. S. Tomkins, & C. E. Izard (Eds.), Affect, cognition, and personality (pp. 390-442). New York: Springer.

Ekman, P. (1972). Universals and cultural differences in facial expressions of emotion. In J. Cole (Ed.), Nebraska symposium on motivation (Vol. 19, pp. 207-281). Lincon: University of Nebraska Press.

Ekman, P. (1978). Facial signs: Facts, fantasies, and possibilities. In T. A. Seboeok (Ed.), Sight, sound, and sense (pp. 124-156). Bloomington: Indiana University Press.

Ekman, P. (1982a). Methods of measuring facial action. In K. R. Scherer, & P. Ekman (Eds.), Handbook of methods in nonverbal behavior research (pp. 45-90). Cambridge University Press.

Ekman, P. (Ed.). (1982b). Emotion in the human face (2nd ed.). Cambridge University Press.

Ekman, P., & Friesen, W. V. (1968). Nonverbal behavior in psychotherapy research. In J. Shlien (Ed.), Research in psychotherapy (Vol. 3 pp. 179-216). Washington, D. C.: American Psychological Association.

Ekman, P., & Friesen, W. V. (1969a). The repertoire of nonverbal behavior: Categories, origins, usage and coding. Semiotica, 1, 49-98.

Ekman, P., & Friesen, W. V. (1969b). Nonverbal leakage and clues to deception. Psychiatry, 32, 88-106.

Ekman, P., & Friesen, W. V. (1975). Unmasking the face: A guide to recoginzing emotions from facial cues. Englewood-Cliffs, N. J.: Prentice-Hall.

Ekman, P., & Friesen W. V. (1978a). Facial Action Coding System (FACS): A technique for the measurement of facial action. Palo Alto, Ca.: Consulting Psychologists Press.

Ekman, P., & Friesen, W. V. (1978b). Investigators guide to the Facial Action Coding System. Palo Alto, Ca.: Consulting Psychologist Press.

Ekman, P., & Friesen, W. V., & Ellsworth, P. (1982a). Conceptual ambiguities. In P. Ekman (Ed.), Emotion in the human face (2nd ed. pp. 7-21). Cambridge University Press.

Ekman, P., Friesen, W. V., & Ellsworth, P. (1982b). Does the face provide accurate information. In P. Ekman (Ed.), Emotion in the human face (2nd ed. pp. 56-97). Cambridge University Press.

Ekman, P., Friesen, W. V., O'Sullivan, M., & Scherer, K.

(1980). Relative importance of face, body, and speech in judgments of personality and affect. Journal of Personality and Social Psychology, 38, 270-277.

Ekman, P., Friesen, W. V., & Scherer, K. (1976). Body movement and voice pitch in deceptive interaction. Semiotica, 16, 23-27.

Ekman, P., Friesen, W. V., & Tomkins, S. S. (1971) Facial Affect Scoring Technique (FAST): A first validity study. Semiotica, 3, 37-58.

Ekman, P., Levenson, R. W., & Friesen, W. V. (1983). Autonomic nervous system activity distinguishes among emotions. Science, 221, 1208-1210.

Ekman, P., & Oster, H. (1982). Review of research: 1970-1980. In P. Ekman (Ed.), Emotion in the human face (2nd ed. pp. 147-173). Cambridge University Press.

Fordyce, W. E. (1976). Behavioral methods for chronic pain and illness. St. Louis: C. V. Mosby.

Ganchrow, J. R., Steiner, J. E., Kleiner, M., & Edelstein, E. L. (1978). A multidisciplinary approach to the expression of pain in psychotic depression. Perceptual and Motor Skills, 47, 379-390.

Gracely, R. H. (1983). Pain language and ideal pain assessment. In R. Melzack (Ed.), Pain measurement and assessment (pp. 71-77). New York: Raven Press.

Gracely, R. H. McGrath, P., & Dubner, R. (1978). Ratio scales of sensory and affective verbal pain descriptors. Pain, 5, 5-18.

Gracely, R. H., McGrath, P., & Dubner, R. (1979). Narcotic analgesia: Fentanyl reduces the intensity but not the unpleasantness of painful tooth pulp sensations. Science, 2Ø3, 1261-1263.

Gubar, G. (1966). Recognition of human facial expressions judged in a laboratory setting. Journal of Personality and Social Psychology, 4, 1Ø8-111.

Hall, J. A. (1979). Gender, gender roles, and nonverbal communication skills. In R. Rosenthal (Ed.), Skill in nonverbal communication: Individual differences (pp. 32-67). Cambridge, Mass.: Oelgeschlager, Gunn & Hain.

Hall, J. A. (1984). Nonverbal sex differences: Communication accuracy and expressive style. Baltimore, ML.: The John Hopkins University Press.

Hanawalt, N. G. (1944). The role of the upper and lower parts of the face as the basis for judging facial expressions: II In posed expressions and "candid camera" pictures. Journal of General Psychology, 31, 23-36.

Harper, R. G., Wiens, A. N., & Mararazzo, J. D. (1978). Nonverbal communication. New York: Wiley.

Harris, G., & Rollman, G. B. (1983). The validity of experimental pain measures. Pain, 17, 369-376.

Hjortsjo, C. H. (1969). Man's face and mimic language. Lund, Sweden: Student Litterateur.

House, A. E., House, B. J., & Campbell, M. B. (1981). Measures of interobserver agreement: Calculation, formulas and distribution effects. Journal of Behavioral Assessment, 3, 37-57.

Hyde, S. (1985). Facial expressive behavior associated with acute pain in a chronic low back pain population. Unpublished doctoral dissertation. University of British Columbia, Vancouver, Canada.

Hyde, S., Gerhard, B., Swalm, D., Lee, D. S., Prkachin, K.M., & Craig, K. D. (1985). Sex differences in nonverbal pain judgment. Manuscript in preparation.

I.A.S.P. (International Association for the Study of Pain: Subcommittee on Taxonomy). (1979). Pain terms: A list of definitions and notes on usage. Pain, 6, 249-252.

Izard, C. E. (1969). The emotions and emotion constructs in personality and culture research. In R. B. Cattell, & R. M. Dreger (Eds.), Handbook of modern personality theory (pp. 496-510). Chicago: Aldine.

Izard, C. E. (1971). The face of emotion. New York: Appleton-Crofts.

Izard, C. E. (1979). The maximally discriminative facial movement coding system (MAX). Newark, Del.: University of Delaware, Instructional Resources Center.

Jacox, A. K. (1980). The assessment of pain. In W. L. Smith, H. Mersky, & S. C. Cross (Eds.), Pain: Meaning and management (pp. 75-88). New York: Spectrum.

Johnson, M. (1977). Assessment of clinical pain. In A. K. Jacox (Ed.), Pain: A source book for nurses and other health professionals (pp. 139-166). Boston: Little Brown & Co.

Kahneman, D., Slovic, P., & Tversky, A. (1982). Judgment under uncertainty: Heuristics and biases. Cambridge University Press.

Kazdin, A. E. (1980). Research design in clinical psychology. New York: Harper and Row.

Keefe, F. J., & Block, A. R. (1982). Development of an observation method for assessing pain behavior in chronic low back pain patients. Behavior Therapy, 13, 363-375.

Keefe, F. J., Wilkins, R. H., & Cook, W. A. (1984). Direct observation of pain behavior in low back pain patients during physical examination. Pain, 20, 59-68.

Kendon, A. (1982). The organization of behavior in face-to-face interaction: Observations on the development of a methodology. In K. R. Scherer, & P. Ekman (Eds.), Handbook of methods in nonverbal behavior research (pp.440-505). Cambridge University Press.

Kleck, R. E., Vaughan, R. C., Cartwright-Smith, J., Vaughan, K. B., Colby, C. Z., & Lanzetta, J. T. (1976). Effects of being observed on expressive subjective, and physiological responses to painful stimuli. Journal of Personality and Social Psychology, 13, 172-184.

Langfeld, H. S. (1918). The judgment of emotions from facial expressions. Journal of Abnormal and Social Psychology, 13, 172-184.

Lanzetta, J. T., Cartwright-Smith, J., & Kleck, R. E. (1976). Effects of nonverbal dissimulation on emotional experience and autonomic arousal. Journal of Personality and Social Psychology, 33, 354-370.

Lanzetta, J. T., & Kleck, R. E. (1970). Encoding and decoding of nonverbal affect in humans. Journal of Personality and Social Psychology, 16, 12-19.

LaRusso, L. (1978). Sensitivity of paranoid patients to nonverbal cues. Journal of Abnormal Psychology, 87, 463-471.

LeResche, L. (1982). Facial expression of pain: A study of candid photographs. Journal of Nonverbal Behavior, 7, 46-56.

LeResche, L., & Dworkin, S. F. (1984). Facial expression accompanying pain. Social Science and Medicine, 19, 1325-1330.

Leventhal, H., & Everhart, D. (1979). Emotion, pain, and physical illness. In C. E. Izard (Ed.), Emotions in personality and psychotherapy (pp. 263-299). New York: Plenum Press.

Leventhal, H., & Sharp, E. (1965). Facial expressions as indicators of distress. In S. S. Tomkins, & C. E. Izard (Eds.), Affect, cognition and personality: Empirical Studies (pp. 296-318). New York: Springer.

Loeser, J. D. (1980). Low back pain. In J. J. Bonica (Ed.), Research publications: Association for research in nervous and mental disease: Vol. 58. Pain (pp.363-377). New York: Raven Press.

Main, C. L., & Waddell, G. (1982). Chronic pain, distress and illness behavior. In C. J. Main (Ed.), Clinical psychology and general medicine: A behavioral perspecitve (pp. 1-62). New York: Plenum.

Melzack, R. (1975). The McGill Pain Questionnnaire: Major properties and scoring methods. Pain, 1, 277-299.

Melzack, R. (1983). Concepts of pain measurement. In R. Melzack (Ed.), Pain measurement and assessment (pp. 1-5). New York: Raven Press.

Melzack, R., & Wall, P. D. (1965). Pain Mechanisms: A new theory. Science, 150, 971-979.

Melzack, R., & Wall, R. D. (1982). The challenge of pain. New York: Penguin Books.

Mohamed, S. N., Weisz, G. M., & Waring, G. M. (1978). The relationship of chronic pain to depression, marital adjustment and family dynamics. Pain, 5, 282-292.

Munn, N. L. (1940). The effect of the knowledge of the situation upon judgment of emotion from facial expressions. Journal of Abnormal and Social Psychology, 35, 324-338.

O'Sullivan, M. (1982). Measuring the ability to recognize facial expressions of emotion. In P. Ekman (Ed.), Emotion in the human face (2nd ed. pp. 281-317). Cambridge University Press.

Patrick, C. J., Craig, K. D., & Prkachin, K. M. (in press). Observer judgments of acute pain: Facial action determinants. Journal of Personality and Social Psychology.

Paulley, J. W., & Haskell, D. J. (1975). Treatment of migraine without drugs. Journal of Psychosomatic Research, 19, 367-374.

Polunin, I. (1977). The body as an indicator of health and disease (Monograph). In J. Blacking (Ed.), The anthropology of the body (pp. 85-98). New York: Academic Press.

Prieto, E. J., & Geisinger, K. F. (1983). Factor analytic studies of the McGill Pain Questionnaire. In R. Melzack (Ed.), Pain measurement and assessment (pp. 63-70). New York: Raven Press.

Prkachin, K, M., & Craig, K. D. (1985). Influencing nonverbal expression of pain: Sensory decision theory analysis. Pain, 21, 399-409.

Prkachin, K. M., Currie, N. A., & Craig, K. D. (1983). Judging nonverbal expressions of pain. Canadian Journal of Behavioral Science, 15, 408-420.

Redican, W. K. (1982). An evolutionary perspective on human facial displays. In P. Ekman (Ed.), Emotion in the human face (2nd ed. pp. 212-280). Cambridge University Press.

Rinn, W. R. (1984). The neuropsychology of facial expressions: A review of the neurological and psychological mechanisms for producing facial expressions. Psychological Bulletin, 95, 52-77.

Rosenthal, R. (1973). Estimating effective reliabilities in studies that employ judges' ratings. Journal of Clinical Psychology, 29, 342-345.

Rosenthal, R. (1982). Conducting judgment studies. In K. R, Scherer, & P. Ekman (Eds.), Handbook of methods in nonverbal behavior research (pp. 287-361). Cambridge University Press.

Rosenthal, R., & DePaulo, B. M. (1979). Sex differences in accomodation in nonverbal communication. In R. Rosenthal (Ed.), Skill in nonverbal communication: Individual differences (pp. 68-103). Cambridge Mass: Oelgeschager, Gunn & Hain.

Scherer, K. R., & Ekman, P. (1982). Methodological issues in studying nonverbal behavior. In K. R. Scherer, & P. Ekman (Eds.), Handbook of methods in nonverbal behavior research (pp. 1-44). Cambridge University Press.

Schwartz, G. E., Brown, S. L., & Ahern, G. L. (1980). Facial muscle patterning and subjective experience during affective imagery: Sex differences. Psychophysiology, 17, 75-82.

Schwartz, G. E., Fair, P. L., Salt, P., Mandel, M. R., & Klerman, G. L. (1976). Facial muscle patterning to affective imagery in depressed and nondepressed subjects. Science, 192, 489-491.

Stark-Ademec, C., & Kimball, M. (1984). Science free of sexism: A psychologists guide to the conduct of nonsexist research. Canadian Psychologist, 25, 23-34.

Stewart, M. L. (1977). Measurement of clinical pain. In A. K. Jacox (Ed.), Pain: A source book for nurses and other health professionals (pp. 107-137). Boston: Little Brown & Co.

Taqiuri, R. (1969). Person perception. In G. Lindsey, & E. Aronson (Eds.), Handbook of social psychology (Vol. 3, pp. 395-449). Reading, Mass.: Addison-Wesley.

Teske, K., Daut, R. L., & Cleeland, C. S. (1983). Relationship between nurses' observations and patients' self-reports of pain. Pain, 16, 289-296.

Tomkins, S. S. (1982). Affect theory. In P. Ekman (Ed.), Emotion in the human face (2nd ed. pp. 353-395). Cambridge University Press.

Turk, C. C., Meichenbaum, D., & Genest, M. (1983). Pain and behavioral medicine: A cognitive behavioral perspective. New

York: Guilford Press.

Tursky, B., Jamner, L. D., & Friedman, R. (1982). The Pain Perception Profile: A psychophysical approach to the assessment of pain report. Behavior Therapy, 13, 376-394.

Vaughan, K. B., & Lanzetta, J. T. (1980). Vicarious instigation and conditioning of facial expressive and autonomic responses to a model's expressive display of pain. Journal of Personality and Social Psychology, 38, 909-923.

Von Baeyer, C. L., Johnson, M. E., & McMillan, M. J. (1984). Consequences of nonverbal expression of pain: Patient distress and observer concern. Social Science and Medicine, 19, 1319-1324.

Weisenberg, M. (1977). Pain and pain control. Psychological Bulletin, 84, 1008-1044.

Winer, B. J. (1971). Statistical principles in experimental design (2nd ed.). New York: McGraw-Hill.

Wolff, B. B. (1971). Factor analysis of human pain responses: Pain endurance as a specific pain factor. Journal of Abnormal Psychology, 78, 292-298.

APPENDIX A

Rating Form Name:______________________________

No Discomfort	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Distracting	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Annoying	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Uncomfortable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Unpleasant	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Irritating	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Upsetting	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Distressing	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Miserable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Frightful	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Dreadful	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Horrible	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Agonizing	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Intolerable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Unbearable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Excruciating	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

No Discomfort	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
Distracting	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
Annoying	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
Uncomfortable	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
Unpleasant	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
Irritating	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
Upsetting	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
Distressing	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
Miserable	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
Frightful	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
Dreadful	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
Horrible	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
Agonizing	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
Intolerable	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
Unbearable	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
Excruciating	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37

No Discomfort	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55
Distracting	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55
Annoying	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55
Uncomfortable	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55
Unpleasant	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55
Irritating	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55
Upsetting	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55
Distressing	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55
Miserable	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55
Frightful	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55
Dreadful	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55
Horrible	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55
Agonizing	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55
Intolerable	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55
Unbearable	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55
Excruciating	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55

No Discomfort	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73
Distracting	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73
Annoying	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73
Uncomfortable	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73
Unpleasant	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73
Irritating	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73
Upsetting	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73
Distressing	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73
Miserable	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73
Frightful	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73
Dreadful	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73
Horrible	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73
Agonizing	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73
Intolerable	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73
Unbearable	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73
Excruciating	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73

No Discomfort	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91
Distracting	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91
Annoying	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91
Uncomfortable	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91
Unpleasant	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91
Irritating	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91
Upsetting	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91
Distressing	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91
Miserable	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91
Frightful	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91
Dreadful	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91
Horrible	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91
Agonizing	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91
Intolerable	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91
Unbearable	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91
Excruciating	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91

No Discomfort	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108
Distracting	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108
Annoying	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108
Uncomfortable	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108
Unpleasant	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108
Irritating	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108
Upsetting	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108
Distressing	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108
Miserable	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108
Frightful	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108
Dreadful	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108
Horrible	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108
Agonizing	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108
Intolerable	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108
Unbearable	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108
Excruciating	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108

No Discomfort	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124
Distracting	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124
Annoying	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124
Uncomfortable	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124
Unpleasant	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124
Irritating	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124
Upsetting	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124
Distressing	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124
Miserable	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124
Frightful	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124
Dreadful	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124
Horrible	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124
Agonizing	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124
Intolerable	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124
Unbearable	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124
Excruciating	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124

No Discomfort	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140
Distracting	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140
Annoying	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140
Uncomfortable	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140
Unpleasant	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140
Irritating	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140
Upsetting	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140
Distressing	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140
Miserable	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140
Frightful	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140
Dreadful	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140
Horrible	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140
Agonizing	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140
Intolerable	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140
Unbearable	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140
Excruciating	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140

No Discomfort	141	142	143	144	145	146	147	148
Distracting	141	142	143	144	145	146	147	148
Annoying	141	142	143	144	145	146	147	148
Uncomfortable	141	142	143	144	145	146	147	148
Unpleasant	141	142	143	144	145	146	147	148
Irritating	141	142	143	144	145	146	147	148
Upsetting	141	142	143	144	145	146	147	148
Distressing	141	142	143	144	145	146	147	148
Miserable	141	142	143	144	145	146	147	148
Frightful	141	142	143	144	145	146	147	148
Dreadful	141	142	143	144	145	146	147	148
Horrible	141	142	143	144	145	146	147	148
Agonizing	141	142	143	144	145	146	147	148
Intolerable	141	142	143	144	145	146	147	148
Unbearable	141	142	143	144	145	146	147	148
Excruciating	141	142	143	144	145	146	147	148

APPENDIX B

Instructions for Study 1

In our lab we have been studying peoples' facial expressions of discomfort from receiving painful stimuli. We are presently interested in how accurately people can rate another person's discomfort when viewing that person's facial expressions.

I would like to give you a little background to the experiences of the people you are about to watch. These people were volunteers in a study looking at the effects of different kinds of acute noxious stimulation on peoples' facial expressions. They went through a series of tests in which they received electric shock, radiant heat, and had strong pressure applied to their index finger above the bone. These stimuli varied in duration form 1/2 to 2 seconds. In all cases, the stimuli were varied from noticeable levels of sensation, which were not uncomfortable, to levels where the subjects felt the stimulation to be as much as they could endure and asked that it be terminated. Under these conditions, the people reported their reactions, which varied throughout the range of descriptors on your rating form. However, no attempt has been made to present equal numbers of facial expressions for all reported levels of unpleasantness.

The video segments you are about to see show peoples' faces during the brief presentation of the various stimuli. As the sole focus of this study is on facial expressions all segments have been specially selected and carefully editted so that you will see only the subjects' facial expressions. All portions of the original tapes which revealed body or head movements have been eliminated either by selection or editting. To control for possible cues due to eye movements the people were instructed to fixate on one of several spots on the wall in front of them. Thus you will see little or no eye movements in these video segments. We have randomly scrambled all the videotape segments so that the order in which you see the segments won't disclose any information about what rating the subjects' gave the stimulation.

The segments you are to rate are separated by a 6 second blank segment during which time you are to make your rating. The beginning of each segment will be announced by a voice indicating the number of the upcoming segment. The numbers announced on the tape correspond with the segment numbers listed on your rating forms. This will help guide you in placing your ratings in the appropriate places.

As you can see on your rating forms there is a list of descriptors shown on the left side of each form which indicate increasing levels of unpleasantness. This is the same list of descriptors used by the people who received the stimuli. You are to use these descriptors in making your rating of the severity of the subject's distress. For example, if you thought the person was in no discomfort you would use the descriptor "no discomfort". If you thought that the person was experiencing a moderate level of pain then you might chose a descriptor such as "distressing".

To make a rating choose the best descriptor and circle the appropriate segment number that is on the same row across from that descriptor. For example, if you decided that in segment 8 the person felt the stimulus was "unpleasant" you would circle number "8" across from the word "unpleasant".

Do you have any questions?

There are 148 segments to rate. The video is 26 minutes long. There are 5 initial practice segments.

APPENDIX C

POST-EXPERIMENTAL QUESTIONNAIRE

Name

1. How accurate do you think you are at judging people from viewing their facial expressions? (circle one)

Not at all accurate						Very accurate
1	2	3	4	5	6	7

2. Have you had any of the following interactions with people experiencing pain?

   a. interacting with many people who have been demonstrating acute or short term pain.

   b. interacting with anyone experiencing chronic (more than 6 months) pain.

   c. none of the above.

3. Is it easier for you to judge the feelings of either females or males from observing their facial expressions? (check one)

   a. females

   b. males

   c. no difference

4. Remembering that the sole focus of this study was on facial expressions, how well did you feel the facial expressions represented the range of distress you rated the subjects as experiencing.

Very representative						Not at all representative
1	2	3	4	5	6	7

5. If you rated the representativeness of the facial expressions as 5 or more would you please indicate what lead you to see them as unrepresentative.

APPENDIX D

Debriefing for Study 1

As stated in the instructions this study is interested in how people rate different types of facial expressions related to pain.

In three prior studies we were able to identify a small constellation of facial movements that were consistently related to self-reported pain. However, it is unclear how well these facial movements communicate the person's pain to judges, independent of other visual cues. The video segments you viewed were portrayals of all possible combinations of the significant facial movements found in the prior studies. Thus the video segments you saw were posed expressions which attempted to duplicate the significant facial movements we found in our previous research in which we used eletric shock, a cold pressor test, and giving back pain patients a range of motion test. Although all the facial movements you saw actually happened in the previous studies it was necessary to show you posed expressions in order to control the visual context in which you viewed the expressions. This control enables me to evaluate the independent role of facial expressions on your ratings.

There are two hypotheses I am investigating. The first relates to what effect does making the facial expressions more complex have on your ratings. It is expected that the higher the complexity of the expression the higher will be your rating. The second hypothesis focuses on the effect of seeing the facial movements portrayed sequentionally versus as a single expression. It is predicted that the segments with a single unified expression will be rated higher than segments in which the same movements are seen sequentially.

Besides attempting to investigate the role of facial expressions in observer judgments this research is also a preliminary step towards the development of a possible method for training health care workers to more accurately assess their patient's pain from observing the patient's facial expressions.

Are there any questions?

APPENDIX E

Gracely Scale Affective Descriptors of Pain

and Scaled Scores

Descriptors	Scaled Scores
No Discomfort[a]	0.0
Distracting[b]	3.1
Annoying	3.5
Uncomfortable	3.8
Unpleasant	4.0
Irritating	4.7
Upsetting	6.1
Distressing	7.7
Miserable	11.7
Frightful	12.4
Dreadful	15.0
Horrible	17.8
Agonizing	25.1
Intolerable	26.7
Unbearable	26.8
Excruciating	38.7

[a] This descriptor was added to the original scale to provide a non-pain base point.

[b] From Gracely, R. H., McGrath, P., & Dubner, R. (1978). Ratio scales of sensory and affective verbal pain descriptors. Pain, 5, 5-18.

APPENDIX F

Hypothesis 1 - 2 X 4 ANOVA sex of judge x level of complexity

Source	SS	df	MS	F	p
Sex	194.03998	1	194.03998	8.08	0.0066
Error	1153.34708	48	24.02806		
Level	741.02791	3	247.00930	75.89	0.0001
Sex X Lev.	58.65509	3	19.55170	6.01	0.0007
Error	468.68156	144	3.25473		

APPENDIX G

Hypothesis 1 - Simple Effects for Sex of Judge and Level of Complexity

Source	SS	df	MS	F	p
Sex at L1	2.89042	1	2.89042	0.34	n.s.
Sex at L2	24.77658	1	24.77658	2.93	n.s.
Sex at L3	100.13106	1	100.13106	11.85	0.001
Sex at L4	124.89902	1	124.89902	14.78	0.001
Error	1622.02864	192	8.44807		
Lev-Male	195.58139	3	65.19380	20.03	0.001
Lev-Female	604.10161	3	201.36720	61.87	0.001
Error	468.68156	144	3.25473		

APPENDIX H

Hypothesis 1 - Newman-Keuls Post Hoc Comparisons for Level of Complexity for Each Sex of Judge

Males

	$\bar{X}_1$	$\bar{X}_2$	$\bar{X}_4$	$\bar{X}_3$
	2.59	4.65	5.93	6.08
$\bar{X}_1$		2.06***	3.34***	3.49***
$\bar{X}_2$			1.28*	1.43**
$\bar{X}_4$				0.15

Females

	$\bar{X}_1$	$\bar{X}_2$	$\bar{X}_3$	$\bar{X}_4$
	3.07	6.06	8.91	9.09
$\bar{X}_1$		2.99***	5.84***	6.02***
$\bar{X}_2$			2.85***	3.04***
$\bar{X}_3$				0.18

*** = .001

** = .025

* = .05

APPENDIX I

Hypothesis 2 - 2 X 2 X 3 ANOVA Sex of Judge X Type of Configuration X Level of Complexity

Source	SS	df	MS	F	p
Sex	170.29088	1	170.29088	7.65	0.008
Error	1067.81038	48	22.24605		
Level	151.10846	2	75.554423	36.85	0.0001
Sex X Lev	16.15179	2	8.07590	3.94	0.0227
Error	196.84912	96	2.05051		
Config	174.98137	1	174.98137	33.29	0.0001
Sex X Conf	30.67999	1	30.67999	5.84	0.0195
Error	252.29390	48	5.25612		
Conf X Lev	37.74372	2	18.87186	14.51	0.0001
S X C X L	6.82375	2	3.41188	2.62	n.s.
Error	124.82809	96	1.30029		

APPENDIX J

Hypothesis 2 - Simple Effects for Type of Configuration and Level of Complexity

Source	SS	df	MS	F	p
Conf at L2	9.01020	1	9.01020	3.44	n.s.
Conf at L3	68.80702	1	68.80702	26.27	0.001
Conf at L4	134.90822	1	134.90822	51.51	0.001
Error	377.12199	144	2.61890		
Lev at Overlap	169.73417	2	84.86709	50.65	0.001
Lev at Non-over	19.11802	2	9.55901	5.71	0.025
Error	321.67721	192	1.67540		

APPENDIX K

Hypothesis 2 - Newman-Keuls Post hoc Comparisons for Level of Complexity by Type of Configuration

Non-Overlapping Configurations

	$\bar{X}_2$	$\bar{X}_3$	$\bar{X}_4$
	4.23	4.65	5.10
$\bar{X}_2$		0.42	0.87 *
$\bar{X}_3$			0.45

Overlapping Configurations

	$\bar{X}_2$	$\bar{X}_3$	$\bar{X}_4$
	4.83	6.31	7.42
$\bar{X}_2$		1.48 **	2.54 **
$\bar{X}_3$			1.11 **

* = .025

** = .001

APPENDIX L

Hypothesis 2 - Simple effects for Sex of Judge and Type of Configuration

Source	SS	df	MS	F	p
Sex at Overlap	172.76642	1	172.76642	12.56	Ø.ØØ1
Sex at Non-Overlap	28.2Ø46Ø	1	28.2Ø46Ø	2.Ø5	n.s.
Error	132Ø.1Ø428	96	13.751Ø9		
Config for Males	29.55685	1	29.55685	5.62	Ø.Ø25
Config for Females	176.1ØØ72	1	176.1ØØ72	33.5Ø	Ø.ØØ1
Error	252.2939Ø	48	5.25612		

APPENDIX M

Hypothesis 2 - Simple Effects for Sex of Judge

and Level of Complexity

Source	SS	df	MS	F	p
Sex at L2	24.37891	1	24.37891	2.78	n.s.
Sex at L3	50.33902	1	50.33902	5.73	0.025
Sex at L4	111.72490	1	111.72490	12.72	0.001
Error	1264.65984	144	8.78236		
Level for Males	35.12025	2	17.56013	8.56	0.005
Level for Females	132.13797	2	66.06899	32.22	0.001
Error	196.84912	96	2.05051		

APPENDIX N

Hypothesis 2 - Newman-Keuls Post hoc Comparisons for

Level of Complexity by Sex of Judge

Male Judges

	$\bar{X}_2$	$\bar{X}_3$	$\bar{X}_4$
	4.03	4.77	5.21
$\bar{X}_2$		0.74	1.18 *
$\bar{X}_3$			0.44

Female Judges

	$\bar{X}_2$	$\bar{X}_3$	$\bar{X}_4$
	5.02	6.19	7.32
$\bar{X}_2$		1.17 *	2.30 ***
$\bar{X}_3$			1.13 **

* = 0.025

** = 0.01

*** = 0.001

APPENDIX O

Instructions for Study 2

In our lab we have been studying peoples' facial expressions of discomfort from receiving painful stimuli. We are presently interested in how accurately people can rate another person's discomfort when viewing that person's facial expressions.

I would like to give you a little background to the experiences of the people you are about to watch. These people were volunteers in a study looking at the effects of different kinds of acute noxious stimulation on peoples' facial expressions. They went through a series of tests in which they received electric shock, radiant heat, and had strong pressure applied to their index finger above the bone. These stimuli varied in duration from 1/2 to 2 seconds. The experiment was divided into two parts. In the first part the people received the stimulation and rated its level of discomfort. In the second part of the study the people received the same stimulation but were instructed to try and control their facial expressions so no one would know how much discomfort they felt. In all cases, the stimuli were varied from noticeable levels of sensation, which were not uncomfortable, to levels where the subjects felt the stimulation to be as much as they could endure and asked that it be terminated. Under both conditions, the people reported their reactions, which varied throughout the range of descriptors on your rating form. However, for the purposes of this study only selected levels of discomfort will be focused on.

The video segments you are about to see show peoples' faces during the brief presentation of the various stimuli. As the sole focus of this study is on facial expressions all segments have been specially selected and carefully editted so that you will see only the subjects' facial expressions. All portions of the original tapes which revealed body or head movements have been eliminated either by selection or editting. To control for possible cues due to eye movements the people were instructed to fixate on one of several spots on the wall in front of them. Thus you will see little or no eye movements in these video segments.

The videotape is comprised of a random selection of segments from the non-instruction condition and the condition where the subjects were instructed to control or hide their facial expressions. In addition, we have randomly scrambled the order of all our videotape segments so that the order in which you see the segments won't disclose any information about what rating the subjects gave the stimulation.

The segments you are to rate are separated by a 6 second blank space during which time you are to make your rating. The beginning of each segment will be announced by a voice indicating the number of the upcoming segment. The numbers announced on the tape correspond with the segment numbers listed on your rating forms. This will help guide you in placing your rating in the appropriate place.

You have two tasks. The first is to decide if the person was trying to hide his or her facial expression or not. The second is to rate as accurately as possible how much discomfort the person was experiencing.

As you can see on your rating forms there is a list of descriptors shown on the left side of each form which indicates increasing levels of unpleasantness. This is the same list of descriptors used by the people who received the stimluli. You are to use these descriptors in making your rating of the subject's discomfort.

Two examples of how to use the rating form are as follows: if you thought the person in the eighth segment felt annoyed by the stimulus and was trying to hide their facial expression, you would place an "X" in the box under the word "hidden" and circle the number "8" across from the word "annoyed". If you thought the person in segment 12 felt the stimulus was unpleasant and they were not trying to hide their facial expression, you would place an "X" in the box under the words "not hidden" and circle the number "12" across from the word "unpleasant".

There are five practice segments and then 60 segments for the actual study. The tape is approximely 9 minutes in duration.

Are there any questions?

APPENDIX P

Name: ______________________________

Rating Form

	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden
	____	____	____	____	____	____	____	____	____	____	____	____
No Discomfort	1	2	3	4	5	6	7	8	9	10	11	12
Distracting	1	2	3	4	5	6	7	8	9	10	11	12
Annoying	1	2	3	4	5	6	7	8	9	10	11	12
Uncomfortable	1	2	3	4	5	6	7	8	9	10	11	12
Unpleasant	1	2	3	4	5	6	7	8	9	10	11	12
Irritating	1	2	3	4	5	6	7	8	9	10	11	12
Upsetting	1	2	3	4	5	6	7	8	9	10	11	12
Distressing	1	2	3	4	5	6	7	8	9	10	11	12
Miserable	1	2	3	4	5	6	7	8	9	10	11	12
Frightful	1	2	3	4	5	6	7	8	9	10	11	12
Dreadful	1	2	3	4	5	6	7	8	9	10	11	12
Horrible	1	2	3	4	5	6	7	8	9	10	11	12
Agonizing	1	2	3	4	5	6	7	8	9	10	11	12
Intolerable	1	2	3	4	5	6	7	8	9	10	11	12
Unbearable	1	2	3	4	5	6	7	8	9	10	11	12
Excruciating	1	2	3	4	5	6	7	8	9	10	11	12

Name: ______________________________

Rating Form

	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden
	____	____	____	____	____	____	____	____	____	____	____	____
No Discomfort	13	14	15	16	17	18	19	20	21	22	23	24
Distracting	13	14	15	16	17	18	19	20	21	22	23	24
Annoying	13	14	15	16	17	18	19	20	21	22	23	24
Uncomfortable	13	14	15	16	17	18	19	20	21	22	23	24
Unpleasant	13	14	15	16	17	18	19	20	21	22	23	24
Irritating	13	14	15	16	17	18	19	20	21	22	23	24
Upsetting	13	14	15	16	17	18	19	20	21	22	23	24
Distressing	13	14	15	16	17	18	19	20	21	22	23	24
Miserable	13	14	15	16	17	18	19	20	21	22	23	24
Frightful	13	14	15	16	17	18	19	20	21	22	23	24
Dreadful	13	14	15	16	17	18	19	20	21	22	23	24
Horrible	13	14	15	16	17	18	19	20	21	22	23	24
Agonizing	13	14	15	16	17	18	19	20	21	22	23	24
Intolerable	13	14	15	16	17	18	19	20	21	22	23	24
Unbearable	13	14	15	16	17	18	19	20	21	22	23	24
Excruciating	13	14	15	16	17	18	19	20	21	22	23	24

Name: ______________________________

Rating Form

	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden
	____	____	____	____	____	____	____	____	____	____	____	____
No Discomfort	25	26	27	28	29	30	31	32	33	34	35	36
Distracting	25	26	27	28	29	30	31	32	33	34	35	36
Annoying	25	26	27	28	29	30	31	32	33	34	35	36
Uncomfortable	25	26	27	28	29	30	31	32	33	34	35	36
Unpleasant	25	26	27	28	29	30	31	32	33	34	35	36
Irritating	25	26	27	28	29	30	31	32	33	34	35	36
Upsetting	25	26	27	28	29	30	31	32	33	34	35	36
Distressing	25	26	27	28	29	30	31	32	33	34	35	36
Miserable	25	26	27	28	29	30	31	32	33	34	35	36
Frightful	25	26	27	28	29	30	31	32	33	34	35	36
Dreadful	25	26	27	28	29	30	31	32	33	34	35	36
Horrible	25	26	27	28	29	30	31	32	33	34	35	36
Agonizing	25	26	27	28	29	30	31	32	33	34	35	36
Intolerable	25	26	27	28	29	30	31	32	33	34	35	36
Unbearable	25	26	27	28	29	30	31	32	33	34	35	36
Excruciating	25	26	27	28	29	30	31	32	33	34	35	36

Name: ______________________________

Rating Form

	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden
	____	____	____	____	____	____	____	____	____	____	____	____
No Discomfort	37	38	39	40	41	42	43	44	45	46	47	48
Distracting	37	38	39	40	41	42	43	44	45	46	47	48
Annoying	37	38	39	40	41	42	43	44	45	46	47	48
Uncomfortable	37	38	39	40	41	42	43	44	45	46	47	48
Unpleasant	37	38	39	40	41	42	43	44	45	46	47	48
Irritating	37	38	39	40	41	42	43	44	45	46	47	48
Upsetting	37	38	39	40	41	42	43	44	45	46	47	48
Distressing	37	38	39	40	41	42	43	44	45	46	47	48
Miserable	37	38	39	40	41	42	43	44	45	46	47	48
Frightful	37	38	39	40	41	42	43	44	45	46	47	48
Dreadful	37	38	39	40	41	42	43	44	45	46	47	48
Horrible	37	38	39	40	41	42	43	44	45	46	47	48
Agonizing	37	38	39	40	41	42	43	44	45	46	47	48
Intolerable	37	38	39	40	41	42	43	44	45	46	47	48
Unbearable	37	38	39	40	41	42	43	44	45	46	47	48
Excruciating	37	38	39	40	41	42	43	44	45	46	47	48

Name: ______________________________

Rating Form

	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden	Hidden / Not Hidden
	______	______	______	______	______	______	______	______	______	______	______	______
No Discomfort	49	50	51	52	53	54	55	56	57	58	59	60
Distracting	49	50	51	52	53	54	55	56	57	58	59	60
Annoying	49	50	51	52	53	54	55	56	57	58	59	60
Uncomfortable	49	50	51	52	53	54	55	56	57	58	59	60
Unpleasant	49	50	51	52	53	54	55	56	57	58	59	60
Irritating	49	50	51	52	53	54	55	56	57	58	59	60
Upsetting	49	50	51	52	53	54	55	56	57	58	59	60
Distressing	49	50	51	52	53	54	55	56	57	58	59	60
Miserable	49	50	51	52	53	54	55	56	57	58	59	60
Frightful	49	50	51	52	53	54	55	56	57	58	59	60
Dreadful	49	50	51	52	53	54	55	56	57	58	59	60
Horrible	49	50	51	52	53	54	55	56	57	58	59	60
Agonizing	49	50	51	52	53	54	55	56	57	58	59	60
Intolerable	49	50	51	52	53	54	55	56	57	58	59	60
Unbearable	49	50	51	52	53	54	55	56	57	58	59	60
Excruciating	49	50	51	52	53	54	55	56	57	58	59	60

APPENDIX Q

Debriefing for Study 2

As stated in the instructions this study is interested in how people rate different types of facial expressions related to pain, particularly how people rate hidden versus non-hidden facial expressions.

From neurological, and clinical sources it has been suggested that people experience difficulty in controlling facial muscles around the eyes and forehead which makes these muscles most likely to remain active even if someone is trying to control their facial expression. The video segments you viewed were portrayals of facial expressions involving muscles solely from either the upper face or the lower face. These expressions were imitations of actual facial exprerssions found in our prior research with low back pain patients. It was necessary to show you imitations of the actual facial movements in order to control the visual context in which you viewed the expressions. This control enabales me to evaluate the independent role of facial expressions on your ratings.

The major hypothesis I am investigating is that judges will chose more expressions showing only upper face movements as hidden than lower face movements. Secondarily, I will be comparing your ratings of the peoples' discomfort to ratings gathered in a prior study, using the same models, to see what effect different instructions have on ratings of discomfort.

Besides attempting to investigate the role of facial expressions on observer judgments this research is also a preliminary step towards the development of a possible method for training health care workers to more accurately assess their patient's pain from observing the patient's facial expressions.

Are there any questions?

APPENDIX R

Hypothesis 3 - 2 X 2 X 2 ANOVA Sex of Judge X Area of Face X Level of Complexity

Source	SS	df	MS	F	p
Sex	0.22501	1	0.22501	5.87	0.0193
Error	1.84095	48	0.03835		
Face Area	0.78647	1	0.78647	10.07	0.0026
Sex X Area	0.05147	1	0.05147	0.66	n.s.
Error	3.75061	48	0.07814		
Level	1.43793	1	1.43793	58.05	0.0001
Sex X Lev	0.00834	1	0.00834	0.34	n.s.
Error	1.18915	48	0.02477		
Area X Lev	0.00147	1	0.00147	0.09	n.s.
Sex X Area X Level	0.00383	1	0.00383	0.23	n.s.
Error	0.79852	48	0.01664		

APPENDIX S

Hypothesis 4 - 2 X 2 X 2 ANOVA Judgment Condition X Sex of Judge X Level of Complexity

Source	SS	df	MS	F	p
Jud Cond	179.77621	1	179.77621	17.36	0.0001
Sex	6.00094	1	6.00094	0.58	n.s.
Jud X Sex	85.91140	1	85.91140	8.29	0.0049
Error	994.27808	96	10.35706		
Level	261.70392	1	261.70392	123.20	0.0001
Jud X Lev	0.00886	1	0.00886	0.00	n.s.
Sex X Lev	1.93676	1	1.93676	0.91	n.s.
Jud X Sex X Level	4.11307	1	4.11307	1.94	n.s.
Error	203.92025	96	2.12417		

APPENDIX T

Hypothesis 4 - Simple Effects for Judgment Condition and Sex of Judge

Source	SS	df	MS	F	p
Jud Cond for Male	257.12043	1	257.12043	12.41	0.001
Jud Cond for Female	8.56659	1	8.56659	0.41	n.s.
Sex at Jud Cond 1	23.25048	1	23.25048	2.24	n.s.
Sex at Jud Cond 2	68.65835	1	68.65835	6.63	0.025
Error	994.27808	96	10.35706		