INFORMATION TO USERS

The most advanced technology has been used to photograph and reproduce this manuscript from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

U·M·I

University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700 800/521-0600

Order Number 9105927

Determinants of diagnostic prototypicality judgments of the personality disorders

Herbert, Diana Lee, Ph.D.

The University of North Carolina at Greensboro, 1990

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

DETERMINANTS OF DIAGNOSTIC PROTOTYPICALITY JUDGMENTS OF THE PERSONALITY DISORDERS

by

Diana Lee Herbert

A Dissertation Submitted to
the Faculty of the Graduate School at
The University of North Carolina at Greensboro
in Partial Fulfillment
of the requirements for the Degree
Doctor of Philosophy

Greensboro
1990

Approved by

Rosemery O. Nelson-Gray

Rosemery O. Nelson-Gray, Ph.D.

APPROVAL PAGE

This dissertation has been approved by the following committee of the Faculty of the Graduate School at the University of North Carolina at Greensboro.

Dissertation Advisor Rosemary O. Nelson-Gray

Committee Members Jacquelyn W. White

April 27, 1990
Date of Acceptance by Committee

April 25, 1990
Date of Final Oral Examination

©, 1990, by Diana Lee Herbert, Ph.D.

HERBERT, DIANA LEE, Ph.D. Determinants of Diagnostic Prototypicality Judgments of the Personality Disorders (1990). Directed by Dr. Rosemery O. Nelson-Gray. 176 pp.

The recent advent of the third edition of the Diagnostic and Statistical Manual of Mental Disorders represented a major shift from a monothetic to a polythetic system for the categorization of psychopathology. Accompanying this shift has been an increased interest in the prototype model of categorization, particularly with respect to the personality disorders. Despite this heightened interest in alternative models of psychiatric nosology, little is known about the precise factors that determine the diagnostic process. The present study examined three factors that have been suggested to be important determinants of personality disorder diagnoses within the prototype model of categorization. These factors included the number of features representative of a personality disorder category, the extent to which those features are typical of the category, and the "dominance" or proportion of category features to the total number of features. A series of personality profiles was constructed in which the above factors were varied factorially. Thirty-two practicing doctoral-level clinical psychologists read 12 profiles of hypothetical clients, and provided ratings of how prototypical each client was of each of the 11 DSM-III personality disorders. Subjects also selected one diagnosis that best categorized the profile. The results revealed

strong main effects for the factors of typicality and dominance, and limited effects for category feature number. Profiles containing features that were more typical of a given personality disorder resulted in higher prototypicality ratings of that category and an increased likelihood of a "correct" diagnosis. Similarly, profiles containing a high proportion of category features relative to the total number of features resulted in higher prototypicality ratings and increased diagnostic accuracy relative to low-dominance profiles. Although in the predicted direction, the effects of category feature number were weaker than predicted. The results support the utility of conceptualizing personality disorders within a prototype framework, and hold important implications for future revisions of nosologies of psychopathology. In particular, it is argued that greater attention should be given to differences in typicality of criterion features of the personality disorders, as well as to the issue of dominance in the assessment process.

ACKNOWLEDGMENTS

This dissertation is in many ways a product of the intellectually stimulating atmosphere of the psychology department at the University of North Carolina at Greensboro. I am deeply indebted to the unwaivering support and invaluable guidance of Rosemery Nelson-Gray throughout all phases of my graduate training. I would also like to thank the faculty of the UNCG psychology department, especially Richard Shull, for his masterful teaching and insightful feedback, and Jacquelyn White, for her guidance and expertise in the areas of personality and statistics. Thanks are also extended to Susan P. Keane and Rita O'Sullivan for their helpful comments on this dissertation. I am also indebted to Steven C. Hayes and to the late Aaron J. Brownstein for their guidance throughout my graduate training.

Appreciation is expressed to John Livesley of the University of British Columbia for furnishing data derived from his own research for incorporation into the personality profiles employed in the present project. Thanks are also extended to Ronald Boykin for his helpful comments, as well as to Lara Branscomb and Pamela Carroll for their help in subject solicitation and data preparation. Graditude is expressed to the clinicians who volunteered to participate as subjects in this study.

I would like to give special thanks to Ashiwell Undie for his statistical expertise, and to Debra A. Hope, Edward N. Sherin, Eileen Egan, and Hannah Jacobson for their assistance with data analysis. Thanks are also extended to Susan McLeer of the Medical College of Pennsylvania for her assistance with this project. Appreciation is also expressed to the UNCG Graduate School for a summer research stipend, and to the American Psychological Association and the Sigma Xi Scientific Society for dissertation grants.

I am also deeply indebted to Coby and Gracye Hobbs, to Gene and Ruby Huddleston, and to George and Maebell Evans for their unwaivering support and guidance throughout my graduate training. Finally, I would like to thank my husband and colleague, James Herbert, for his invaluable support and his insightful feedback on this project, and my son, Aaron, for his inspiration and patience.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF TABLES (Continued)

LIST OF TABLES (Continued)

CHAPTER I

INTRODUCTION

The ability to discriminate phenomena and to "re-present" them in categories is fundamental to effective behavior with respect to those phenomena. "Whenever we perform any kind of action, say something as mundane as writing with a pencil, hammering with a hammer or ironing clothes, we are using categories. The particular action we perform on that occasion is a kind of motor activity (e.g., writing, hammering, ironing), that is, it is in a particular category of motor actions. They are never done in exactly the same way, yet despite the differences in particular movements, they are all movements of a kind..." (Lakoff, 1987, p. 6).

From a behavioral perspective, such categories of actions are response classes that are consistently differentially reinforced when performed in the presence of particular discriminative stimuli. The specific responses may never have exactly the same topography, yet, functionally, they are all members of the same response class. Behaviorists make a further distinction between responding differentially to discriminative stimuli and verbally categorizing different stimuli. For example, pigeons can learn to respond differentially to different colored response keys, but one would not say that they have developed a verbal categorization system for colors. Verbal specifications of stimuli are not absolutely necessary for effective behavior with respect to those stimuli. Yet verbal specifications of stimuli can be useful in sharpening stimulus control by increasing the probability of differential responding to those stimuli.

The Functions of Category Labels

Verbal specifications of stimulus categories are necessary for the development of accurate descriptions of natural phenomena. Radical Behaviorism holds that verbal behavior helps sharpen stimulus control by allowing for the specification and construction of discriminative stimuli and contingencies of reinforcement (Skinner, 1969). The act of verbally specifying stimulus categories, such as the color "red," objects that are "hammers," or "histrionic" behavior, is called a "tacting" (Skinner, 1969, p. 81). The verbal response of tacting the color "red," for example, is evoked by the stimulus property of a particular range of wavelength of light (Skinner, 1969). A parent's differential reinforcement for the child's tacting of the color "red," the object "hammer," or other stimulus categories increases the probability that the child will readily respond differentially to those stimulus categories (Herrnstein, 1982).

Thus, the accurate specification or labeling of stimulus categories increases the probability that those listening will respond effectively with respect to those categories. "Similarly, when a speaker intraverbally reconstructs directions, rules of conduct, and 'laws of thought,' he increases the likelihood of successful practical, ethical, and intellectual behavior, respectively, and his success in doing so depends upon the 'purity' of the controlling relations" (Skinner, 1957, p. 418). Furthermore, explicit classification systems sharpen the discriminative control of verbal responses to related groups of stimuli (Skinner, 1957). For example, the revised third edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-III-R, American Psychiatric Association, 1987) serves as a classification system that helps psychodiagnosticians classify related groups of clinical behaviors into diagnostic categories, thereby facilitating effective behavior with respect to those phenomena.

From a cognitive perspective, category labels facilitate learning and recall. Labels of categories allow a large amount of related information to be organized hierarchically into a limited number of memory chunks (Bower & Hilgard, 1981; Tulving & Donaldson, 1972). Attribution theory stresses the importance of category labels in helping people predict and control their social environment (Yarkin, Harvey & Bloxum, 1981). A label about a person may indicate to others which behaviors would probably be most effective in interacting with that person, based upon their interactions with similar persons in the past. Each perspective on the utility of category labels emphasizes the increased effectiveness of behavior as the result of effective categorization. Of particular interest to the present study are psychodiagnostic classification systems, which may be viewed as verbal specifications of certain categories of behavioral phenomena that are intended to enhance the effectiveness of assessment and treatment of these phenomena.

The Classical View of Categorization

Much attention within the fields of cognitive and clinical psychology has recently been devoted to contrasting "prototype" categorization with "classical" categorization. The roots of the "classical" model of categorization are somewhat difficult to trace because the currently popular literature on the prototype model cites few proponents of the classical model.

The 17th century physician, Thomas Sydenham (1624-1689) has been credited as the founder of modern medical nosology (Dewhurst, 1966; Temkin, 1965). The roots of the classical model of categorization can be found in Sydenham's writings

(Dewhurst, 1966). Rather than recommend the virtually useless theoretical medical books of the 1600's, he suggested that physicians read Don Quixote as an text on the skills and merits of empirical observation. He asserted that the species and genera of diseases should be distinguished upon the basis of purely empirical observation of a constant set of necessary and sufficient symptoms existing in each member of a given class, rather than upon etiological theories (Dewhurst, 1965; Temkin, 1965).

Directly following from Sydenham's emphasis upon a set of necessary and sufficient criteria for category membership, Baissier de Sauvage (1790) developed a classification system of diseases consisting of ten classes, each with successive orders and genera. Note that Sydenham's requirement that classification systems be based upon a set of necessary and sufficient criteria is now simply referred to in the current prototype literature as the classical or "monothetic" model of categorization. The proponents of the prototype model may be quite surprised to learn, however, that even as the founder of the monothetic model of medical nosology, Sydenham strongly emphasized the importance of considering heterogeneity within classes of disease. Comparing himself to a botanist, he stated, "that botanist would have but little conscience who contented himself with a general description of a thistle and overlooked the special and peculiar characteristics in each species" (Dewhurst, 1966).

Application of the classical model of categorization was refined in the hierarchical classification system of the Swedish naturalist, Carolus Linnaeus (1701-1778). This taxonomic system specified a list of necessary and sufficient criteria for category membership. For example, the class of Mammalia is defined by the following necessary and sufficient criteria: "young nourished by milk glands, skin with hair or fur, body cavity divided by diaphragm, red corpuscles without nuclei, and high body temperature" (Curtis, 1979, p. 346).

The classical model was used as the basis of Charles Fourier's classification system of personality types, developed in the early 1800's. Beginning with three basic "passions," he then divided each of them into 12 orders, 12 genera, 134 species, and 404 varieties, making for a total of 810 character types (Allport, 1937).

The classical model has also been called the "monothetic" approach to categorization (Hempel, 1965; Schwartz & Wiggins, 1987). According to the monothetic model, category membership is determined by the presence or absence of a set of necessary and sufficient criteria. Thus, an entire set of criteria is necessary for category membership in that no criteria can be missing. The set of criteria is also sufficient to acquire category membership in that no additional criteria need be fulfilled. Although at face value the monothetic model appears to be very useful, it is based upon several assumptions that are quite problematic. If all members of a classical category possess the same set of necessary and sufficient criteria, it follows that all category members are equally good, representative instances of the category. Moreover, all category members are equally poor, unrepresentative examples of other categories. Despite Sydenham's emphasis upon the importance of considering the heterogeneity within members of the same class, a weakness of the monothetic model is that it does not account for different degrees of heterogeneity that exist between members of the same category. For example, within the monothetic class of Mammalia, a greater degree of heterogeneity exists between felines and whales than does between felines and canines.

Given its requirement of a constant set of necessary and sufficient criteria for category membership, the monothetic model of classification appears to be based on the assumption that the determination of category membership is very clear-cut, with homogeneity within categories, clearly demarcated boundaries of categories, and heterogeneity between categories. It would follow from this model that natural phenomena exist in a pure, homogeneous form, as if they had been parceled into discrete "abstract containers" (Lakoff, 1987, p. 6).

Moreover, the impact of motivational influences upon the observer's determination of category membership is not addressed. According to Lakoff (1987), the classical monothetic model is based upon the assumption that human reason is most successful when it operates autonomously of motivational influences. This assumption holds that categories of phenomena are clearly demarcated and self-evident to those who have succeeded in isolating their reasoning powers from motivational influences. Lakoff (1987, p. 9) suggests that the classical monothetic concept of a category is based upon the rather egocentric assumption that "there is a correct, God's eye view of the world -- a single correct way of understanding what is and what is not true." He argues that such ideas "need to be replaced by ideas that are not only more accurate, but more humane" (1987, p. 9).

The Prototype Model of Categorization

Whereas monothetic categories specify a list of necessary and sufficient criteria, polythetic categories specify a list of criteria, but only require a particular number of those criteria to be met for diagnostic inclusion. The evolution from the DSM-II to the DSM-III as the standard for psychiatric nomenclature represents a move from the monothetic approach to the polythetic approach. It has been suggested that a further move to a prototype model would be a further improvement (Millon, 1986).

The revolution from the classical, monothetic model to the prototype model is not really as dramatic a change as is described in the literature. According to this literature, the classical view dominated psychiatric classification until the creation of DSM-III. Calling this approach the "classical" model of categorization is misleading because it makes the prototype view appear new and innovative relative to the monothetic view, when actually the roots of the prototype view go just as far back in history as the roots of the "classical" view.

Although the prototype model is commonly presented as a major innovation, its roots can be traced back to the ancient Greek philosophers, making the term "classical" just as applicable to the prototype model as it is to the monothetic model. Within the literature on personality disorders, this alternative model of categorization specifies a prototype or an "ideal type" of a given personality style. "The ideal type itself may rarely be seen, but it represents the synthesis of many similar cases and serves as the exemplar against which future cases can be compared" (Francis & Widiger, 1986). This model of categorizing personality types can be traced to the "character writings" of

Theophrastus, a student and successor of Aristotle (Allport, 1937; Roback; 1927).

> "The 'portraits' by Theophrastus are similar both in form and substance to the DSM-III diagnoses of character pathology, but they surpass DSM-II and anticipate DSM-III by being atheoretical, and anticipate DSM-IV by including not only a narrative description of the ideal type, but also a list of behavioral acts that typify each style (for example, the 'penurious' character forbids anyone to pick a fig from his garden, daily checks his boundary markers on his property, and will move furniture to find a lost copper)" (Frances & Widiger, 1986, p. 242).

In fact this line of thought can be traced even further back to Aristotle's mentor, Plato. According to Plato's teachings, each type or class of phenomenon is represented by its perfect, ideal form in the world of ideas. The word prototype (from the Greek "prototupon," meaning first form), commonly denotes an original model of a phenomenon that is subsequently reproduced, in ways that represent significant modifications from the original form. It is argued that Plato's concept of the perfect, ideal form does not necessarily reflect the "classical view" that natural phenomena exist in pure, homogeneous forms. It is possible to use a prototype or ideal type as a perfect example of a particular category without believing that the actual members of the category must be perfect as well. "The relationship of the 'many' objects, belonging to a certain class of things in the sense world, to the 'One,' i.e. the single Idea which is their archetype, is another great source of difficulty to Plato" (Bourke, 1962). One of the solutions discussed in Plato's dialogues is "that the many participate imperfectly in the perfect nature of their Idea" (Bourke, 1962). This classical Greek concept of a perfect form being used to categorize real, imperfect phenomena appears to have been rejuvenated in the current prototype model, as can be seen in the following quotations. "Prototypic categories are organized around prototypical examples (the best examples of the concept) with less prototypical examples forming a continuum away from these central cases" (Livesley, 1985). "The clinician is then able to recognize the many imperfect cases by their resemblance or approximation to the whole" (Schwartz & Wiggins, 1987).

In more recent times, Wittgenstein (1953) has been credited with specifying the philosophical assumptions upon which the prototype model is based. Using the category of "games" as an example, he stated that board games, card games, and "ring-around-the-rosy" type games shared a particular "family resemblance." Yet, each instance of a game does not share all the category features of a game. Thus, category members need not all share the same set of necessary and sufficient criteria for membership. Rather, category members share variable, probabilistic family resemblances along any number of dimensions.

Whereas Wittgenstein's insights related more to family resemblances among objects having ordinary, prescientific category names, the German social scientist, Max Weber, has been credited with laying the foundations for a prototype model of scientific categorization (Jasper, 1953; Schwartz & Wiggins, 1987). Applying this model to historic sociology, Weber introduced the well-known "ideal types" of "the protestant ethic" and the "spirit of capitalism" in his analysis of individual Americans.

Weber's approach could also be applied to the assessment and diagnosis of maladaptive behavior patterns. When confronted with an individual patient, the clinician has an infinite amount of varying information, which can be better understood when it is compared with an "ideal type." In Weber's words,

> "... as soon as we attempt to reflect about the way in which life confronts us in immediate concrete situations, (we realize that) it presents an infinite multiplicity of successively and coexistently emerging and disappearing events, both "inside" and "outside" of ourselves. The absolute infinity of the multiplicity is seen to remain undiminished even when our attention is focused on a single "object," ... as soon as we seriously attempt an exhaustive description of **all** the individual components of this "individual phenomenon," to say nothing of trying to explain it causally" (p. 72).

According to Weber, the only way to make such information scientifically intelligible is to compare it against an "accentuation" of different perspectives that people have of themselves, of others, or of other social or natural phenomena. This "accentuation" must be "one-sided" in its emphasis of particular cultural values, or particular scientific or practical goals, such as the promotion of mental health in the field of clinical psychology.

> "An ideal type is formed by the one-sided accentuation of one or more points of view and by the synthesis of a great many defuse (sic), discrete, more or less present, and occasionally absent concrete individual phenomena, which are arranged according to those one-sidely emphasized viewpoints into a unified thought-construct" (p. 90).

Based upon Weber's insights, the eminent German philosopher and psychiatrist, Karl Jaspers, applied the concept of "ideal types" in his book General Psychopathology (1963), the first edition of which was published in 1913. In his book, Jaspers draws a sharp contrast between two basic kinds of categories. One is the monothetic category, which he terms the "generic group." Note that the term generic commonly describes a attribute or group of attributes that is related to an entire class. In the field of biology, a generic attribute is an attribute that is common to all members of a Genus in the Linnaeun system of monothetic classification. Jaspers contrasts the "generic group" with the category based upon the "ideal type":

> "A case either belongs or does not belong to a generic group (e.g., paralysis) whereas a case only corresponds more or less to a type (e.g., hysterical personality). A generic group is the concept which

> represents an actually existing and definable variant. A type is a fictitious construct which in reality has fluid boundaries; it serves to assess a particular case ... Generic groups either exist or they do not. Types reveal themselves as either fruitful or not for the comprehension of individual cases ... Through the use of generic groups, real boundaries are established; through the use of types we only give structure to a transient manifold" (p. 560).

Upon the basis of Jasper's work, his student and colleague, Kurt Schneider, developed a categorization system using ideal types in his book on the clinical psychopathology of personalities (1958).

The current literature on the prototype model of categorization credits Eleanor Rosch (1973a, 1973b) for having revolutionized the study of categorization in experimental cognitive psychology with her development of "prototype theory." She, in turn, credits much of her work to the insights made by Wittgenstein. Rosch explicitly challenged two assumptions of the monothetic approach to categorization. The first assumption was that if category membership is truly determined only by a set of inherent properties of each member, then categories should exist independently of the peculiarities of the organisms doing the categorizing. Secondly, if category membership is truly determined by a set of necessary and sufficient properties inherent to each member, then all category members should be equally representative of the category. Rosch and others have demonstrated substantial evidence (discussed below) against both of these implications of the monothetic theory of categorization.

Rosch (1978) maintains that there are two major principles of prototype categorization. Both of these principles are based upon the notion that the categorization of phenomena is highly influenced by the idiosyncrasies of the organisms doing the categorizing. The first principle is that categorization functions to maintain a "cognitive economy" by providing "maximum information with the least cognitive effort" (p. 28). Rosch elaborates that this principle of categorization requires that the categorizer strike a balance between making as many fine discriminations between stimuli as possible, and generalizing enough across those stimuli to make sense and use of them.

> "On the one hand, it would appear to the organism's advantage to have as many properties as possible predictable from knowing any one property, a principle that would lead to formation of large numbers of categories with as fine discriminations between categories as possible. On the other hand, one purpose of categorization is to reduce the infinite differences among stimuli to behaviorally and cognitively usable proportions. It is to the organism's advantage not to differentiate one stimulus from others when that differentiation is irrelevant to the purposes at hand" (Rosch, 1978, pp. 28-29).

Rosch states that cognitive economy could be maintained by monothetic dichotomous categorization, but that an alternative way "to achieve separateness and clarity of actually continuous categories is by conceiving of each category in terms of its clear cases rather than its boundaries" (1978, pp. 35-36). Rosch's second major principle, "perceived world structure," is that the world is not perceived as "an unstructured total set of equiprobable co-occurring attributes. Rather, the material objects of the world are perceived to possess...high correlational structure...what attributes will be perceived given the ability to perceive them is undoubtedly determined by many factors having to do with the functional needs of the knower interacting with the physical and social environment" (1978, p. 29).

Although the principle of "cognitive economy" has often been interpreted as inherently cognitive in the prototype literature, Rosch's emphasis upon a delicate balance between stimulus discrimination and generalization is clearly compatible with a behavioral perspective on categorization. In behavioral terms, the features of stimulus categories are not all equiprobable representatives of those categories. The presence of some features is more highly correlated with a given category than is the presence of other features. Rather than categorizing phenomena according to a predetermined set of necessary and sufficient criteria, we categorize phenomena along a flexible gradient of family resemblances. Moreover, the exact "correlational structure" or generalization gradient of stimulus categories depends upon the function that the stimulus categories serve in increasing the effectiveness of the organism's behavior. Thus, whereas the monothetic approach holds that discrete categories exist independently of the categorizer, prototype theory holds that the process of categorization necessarily involves an interaction between the categorizer and the environment. Categorization systems are inevitably influenced by motivational forces acting upon the categorizer's behavior.

Rosch (1978) admonishes that prototype theory has been subjected to many simplistic interpretations and misunderstandings, the most common of which involves the reification of the concept of the prototype. "To speak of a prototype at all is simply a convenient grammatical fiction; what is really referred to are judgments of degree of prototypicality...For natural-language categories, to speak of a single entity that is the prototype is either a gross misunderstanding of the empirical data or a covert theory of mental representation" (Rosch, 1978, p. 40). Ironically, however, this reification of the prototype may be exactly what many thinkers have done to Plato's original concept of an abstract "ideal form" for each category that participates in the imperfect nature of the concrete category members. As Rosch warns, the tendency to reify prototypes and to ask whether a category feature is or is not part of a prototype harks back to the dichotomous question of whether a feature is or is not within the boundaries of a monothetic category. For example, the question of whether overly dramatic behavior is or is not a member of the Histrionic Personality Disorder prototype is only one step removed from the underlying question of whether overly dramatic behavior is or is not one of the necessary and sufficient criteria for membership within the monothetic category of Histrionic Personality Disorder. Rosch argues that such "thinking precisely violates the Wittgensteinian insight that we can judge how clear a case something is and deal with categories on the basis of clear cases in the total absence of information about boundaries" (1978, p. 36).

In addition to the reification of the prototype, prototype theory has also been widely misunderstood to presuppose various theories about how categories are learned and processed. As insightfully stated by Rosch (1978) and reiterated by Lakoff (1987), prototype theory does not constitute or presuppose any specific theory regarding the processing, representation, or learning of categories. "A representation of categories in terms of conjoined necessary and sufficient attributes alone would probably be incapable of handling all of the presently known facts, but there are many representations other than necessary and sufficient attributes that are possible...prototypes only constrain but do not specify representation and process models" (Rosch, 1978, pp. 40-41).

Prototype Research on Common Categories

In addition to her theoretical insights, Rosch has led a series of programmatic empirical investigations, providing overwhelming support for prototype theory. Her research in this area began with the study of how arbitrary names of primary or "focal" colors are learned more easily than names of nonprimary colors by both English-speaking children and adults and by Dani-speaking New Guinea natives (Heider, 1972; Rosch, 1973). The New Guinea natives spoke Dani, a language that has only two color categories: mili ("dark-cool" colors, such as blue, green, and black) and mola ("light-warm" colors, such as yellow, red, and white). Although the Dani-speakers categorized colors within these apparently dichotomous categories, they discriminated new colors along generalization gradients. They also learned arbitrary names more easily for colors that were presented within a range that had a regular primary color, such as red, at the center than for colors that were presented within a range that had a nonprimary color, such as orange, at the center (Rosch, 1973). Despite differences between English and Dani speakers in the formal linguistic categories into which colors are divided, speakers of both languages apparently discriminate colors according to similar generalization gradients. Similarly, Heider (1972) found that when four year-old English-speakers were asked to match a colored chip with another chip chosen from an array of color chips, they matched the colors best when the initial chip was a primary color. In this sense, primary colors appeared to be more prototypical of particular color ranges than nonprimary colors in the same general color range.

Rosch (1973) extended her research to the examination of prototype categorization of natural objects and animals, such as chairs and birds. Rosch and her colleagues found that some members of a given category were consistently rated as being more prototypical than other members of a category. For example, in the case of birds, robins were consistently rated as being more prototypical than penguins, chickens, and ostriches. Further studies have demonstrated that even in cases when subjects do not agree on the boundaries of categories, they overwhelmingly agree in their judgments of how representative or prototypical an example is of a given category (Rosch, 1974, 1975a, 1975b). When a quantifiable property of category members, such as size, is examined, prototypical members actually represent the means of that property (Reed, 1972; Rosch, Simpson & Miller, 1976). In related studies (Posner & Keele, 1968; Reed, 1972), subjects abstracted a prototypical visual pattern from a series of dot patterns or line-drawn faces. These prototypical visual patterns represented the central tendency of the series and were used by subjects as a standard against which to compare the prototypicality of novel visual patterns.

Subjects' prototypicality ratings of category members are measured by numerous dependent variables. These variables include reaction time in category naming (Rosch, 1976b), speed of learning artificial categories, order and probability of producing examples of category members, order of learning category members, and speed of recognizing members of a category (Mervis & Rosch, 1981; Smith & Medin, 1981).

Among the factors that have been proposed to influence prototypicality ratings, Rosch and Mervis (1975) have shown that the number of features a category member shares with other members of that category is highly positively correlated with that category members' prototypicality ratings. Conversely, the number of features a category member shares with contrasting categories is highly negatively correlated with the prototypicality ratings of that category member. Thus, prototypicality is highest for category members that have the most features in common with members of that category, and the least features in common with members of other categories. Given the importance of considering how the presence of noncategory features affects prototypicality, Rosch and her colleagues (Rosch & Mervis, 1975; Rosch, Mervis, Gray, Johnson & Boyes-Braem, 1976) maintain that prototypicality is a function of the "cue validity" of category members.

> "Cue validity is a probabilistic concept; the validity of a given cue x as a predictor of a given category y (the conditional probability of y/x) increases as the frequency with which cue x is associated with category y increases and decreases as the frequency with which cue x is associated with categories other than y increases. The cue validity of an entire category may be defined as the summation of the cue validities for that category of each of the attributes of the category." (Rosch, 1978, pp. 30-31).

The concept of cue validity, borrowed from previous investigators (Beach, 1964a, 1964b; Reed, 1972), is comparable to Tversky's (1977; Tversky & Gati, 1978) quantitative concept of "category resemblance," but cue validity places more weight upon the category features associated with contrasting categories. Rosch (1978, p. 37) holds that for natural language categories, "the extent to which items have attributes common to the category was highly negatively correlated with the extent to which they have attributes belonging to members of contrast categories...it is a fact that both representativeness within a category and distinctiveness from contrast categories are correlated with prototypicality in real categories."

In addition to examining the determinants of prototypicality of category members, researchers have also examined the determinants of prototypicality for specific features or feature combinations of category members. Based upon her research, Rosch (1978) reasons that the most salient stimulus features within a category are most likely to become associated with the category name. Such features, whose observation is most correlated with a category name, are judged to be the most representative or prototypical features of that category. For example, in Rosch's color experiments, primary colors were judged as being more prototypical than other colors within a given color range because their stimulus features were more likely to be associated with the learned category name. Thus, the relative salience of particular feature combinations appears to influence which feature combinations will tend to co-occur with the category name, thereby affecting which feature combination will tend to be judged as more prototypical than other combinations.

In their quest for the determinants of feature prototypicality, Malt and Smith (1984) have conducted experiments suggesting that prototypicality ratings of category feature combinations are influenced by the degree to which those features are correlated with one another. Prototypicality ratings are also strongly influenced by the extent to which the presented features represent "particularly salient or functional combinations" (1984, p. 250).

Tversky and Hemenway (1984) have arrived at a similar conclusion in examining various properties or "parts" of category members (such as presence of core, roundness, and juiciness for members of the apple category). These authors note that when subjects list the parts or properties of category members, they do not list every possible part, such as cells or molecules in the case of apples, as if each part were equally representative of apples. Rather, they list only "the attributes important for distinguishing the appearance or function of the object, so that the smell of flowers and the taste of fruit are mentioned, but not the smell or taste of clothing" (p. 178). The authors conclude that properties of natural categories are correlated with one another on the basis of how particular property configurations signal the category member's structure or function, and facilitate communication.

On the basis of these findings, the prototypicality of feature combinations appears to be influenced by the features' inter-correlation, saliency, and ability to signal a phenomenon's structure or function, or to facilitate communication about that phenomenon. From a behavioral perspective, it is argued that all of these factors are important determinants of feature prototypicality for a single underlying reason: they influence the degree to which feature combinations are discriminative for behaviors that have been effective in interacting with the phenomenon in question, as well as for behaviors that have been effective in the broader environment in which that phenomenon is presented.

A Behavioral Analysis of Prototype Categorization

Following from Rosch's (1978) point that prototype categorization does not presuppose any particular theory of processing, representation, or learning, it is argued that prototype categorization is very compatible with a behavioral perspective. In fact, a non-monothetic view of categorization is anything but revolutionary to the field of experimental behavioral analysis. Behavior analysts have long held that a concept, such as "chair," for example, is based upon our generalization among all compound stimuli we call "chair," and our discrimination between those stimuli and the stimuli we do not call "chair" (Keller & Shoenfeld, 1950). Behavior analysts have empirically demonstrated gradients of category membership based upon the principles of generalization and discrimination (e.g., Herrnstein & Loveland, 1964; Jenkins & Harrison, 1960; Lea & Harrison, 1978; Morgan, Fitch, Holman, & Lea, 1976). When a response has been reinforced in the presence of a specific stimulus property, and that stimulus property is varied along one or more stimulus dimensions, then responding to future stimuli is a function of the similarity of these stimuli to the original stimulus. Responses to some stimuli are more likely to be consequated by consistent differential reinforcement than are responses to other stimuli. Responding occurs along a generalization gradient, such that the effects of past reinforcement in the presence of the original stimulus are transferred more to more similar stimuli, and less to less similar stimuli (Catania, 1984). Rosch's first principle of prototypical categorization, "cognitive economy" is compatible with the behavioral emphasis upon maintaining a balance between stimulus discrimination and generalization in facilitating effective behavior. Rosch's second principle, "perceived world structure," is compatible with the behavioral position that the categorization of phenomena occurs along a generalization gradient.

Implications of Monothetic and Prototype Categorization for Social, Behavioral, and Personality Categories

As compared to natural phenomena such as colors and objects, social, behavioral, and personality phenomena are often composed of more complex compound stimuli that vary along many stimulus dimensions. As a result of this complexity, the contrast between monothetic and prototype views of social, behavioral, and personality phenomena is particularly striking. For example, an etiquette expert might construct a monothetic classification system of behaviors such that all behaviors would be neatly classified as either always absolutely proper or always absolutely improper according to the particular social situation in which they were exhibited. A common monothetic model of categorization that associates specific behaviors with personality types rather than with situations is the concept of a fixed set of personality traits that remains consistent across environmental contexts. This set of necessary and sufficient traits is assumed to be equally inherent to the various people who are members of a particular personality category. This approach to personality categorization implies that there is "only one correct God's eye view" of a person according to the particular personality category of which that person is a member (Lakoff, 1987, p. 9). Personality styles are particularly cumbersome for monothetic theory, given the heterogeneity of individuals within any personality category, the difficulty of specifying necessary and sufficient criteria for membership within a personality category, and the influence of personality theories upon different systems of personality categories.

From a prototype perspective, social situations, behaviors, and personality styles are not categorized according to a set of necessary and sufficient criteria. Rather, these phenomena are categorized so that certain phenomena are more prototypical of a given category than are other phenomena. For example, specific behaviors or behavioral patterns are seen as more prototypical than others of a particular personality style. Moreover, given two people with a particular personality style, one person may be considered more prototypical of that personality style than the other person. Even if two individuals are viewed as equally prototypical of a given personality style, it is possible that they may differ along many dimensions, sharing few specific features. According to the prototype view, the suitability of members to particular personality categories is probabilistic, rather than entirely suitable, or entirely unsuitable. As compared to the monothetic view of personality categorization, the prototype view recognizes a more representative proportion of within-category heterogeneity and between-category homogeneity.

Prototype Research on Nonclinical Personality Categories

Studies of the prototype categorization of personality categories and social situations have revealed a striking similarity between the factors influencing the categorization of these phenomena and more basic phenomena. For example, people categorize social situations such that some situations are judged to be more prototypical of a particular type of situation than are others (Cantor, Mischel, & Schwartz, 1982; Schutte, Kenrick, & Sandalla, 1985). Sparked by their interest in the interaction between person and situation, Cantor and Mischel (1977) have also focused on the prototype categorization of personality types. They hypothesized that personality traits actually function as conceptual prototypes that help us make sense of and predict other people's behavior. Using a series of trait adjectives, associated with the prototypes of extroversion and introversion, they formulated brief written descriptions of four characters. These included an extravert, an introvert, a nonextravert control character, and a nonintrovert control character. In one condition, the extravert and introvert characters were explicitly identified as such to the subjects. In another condition, however, they were not identified as such. Results from a memory recognition task showed that subjects selectively recognized conceptually related, but unpresented items for the extravert and introvert characters, whether or not they had been explicitly labeled as such. Thus, whether subjects received a personality label or not, they still formed an overall impression of the introvert and extrovert, which in turn resulted in their selective recognition of actually unpresented category-consistent information. Cantor and Mischel concluded from these results that even when people are not exposed to an explicit label of others' personality styles, they tend to organize and recall information about others as a function of others' similarity to pre-existing, trait-based prototypes.

Cantor (1978) continued this line of research in her dissertation, suggesting that at least three factors influence the prototypical categorization of personality, given extensive exposure to another person. These factors are termed breadth, dominance, and differentiation. First, breadth is the number and variety of different category-consistent attributes displayed (Cantor, 1978; Cantor & Mischel, 1979b). It was hypothesized that person who exhibits a greater number of category-consistent attributes will be judged as more prototypical of that particular category than a person who exhibits a smaller number of category-consistent attributes. Second, dominance is the number of category-consistent attributes relative to the total number of attributes displayed. It was hypothesized that "perceived prototypicality increases with increases in the ratio of the category-consistent attributes displayed relative to the total set of attributes displayed by the individual. Prototypicality will increase as the category-consistent attributes assume a position of 'figure' against the 'background' of the total configuration" (Cantor & Mischel, 1979b, p. 32). A third possible factor affecting the categorization of personality is differentiation: the degree to which the displayed attributes can be differentiated from contrasting or incompatible categories. The authors hypothesize that people "negatively weight such attributes in judgments of prototypicality according to the degree of incompatibility with the type" (Cantor & Mischel, 1979b, p. 33).

In her dissertation, Cantor (1978) examined the effects of breadth, dominance, and differentiation upon subjects' ratings of characters' prototypicality. She first had subjects generate personality descriptions of friends they had who were good, moderate, or poor examples of an extravert. Then, a second group of six judges rated the prototypicality of the descriptions. The ratings among the judges were reliable at 0.87. A "breadth-dominance-differentiation" index for each character was determined by statistically combining the three factors, and was then used to predict subjects' prototypicality ratings of each character. The breadth-dominance-differentiation scores for each character were highly correlated with the prototypicality ratings.

Given restricted exposure, rather than extensive exposure, Cantor predicted that "prototypicality would be increased when the target individual exhibited the most central (highly associated) category attribute(s) consistently and intensely across many situations and particularly in situations where such behavior is non-normative" (Cantor, 1978, p. 4645-B). Cantor constructed three "behavioral episodes" or paragraph-length short stories describing extravert or bright-intelligent characters. The characters were rated as more prototypical when they were consistently outgoing or intelligent in three non-normative situations, as opposed to normative situations. Given these results, Cantor (1978) concluded that prototypicality is "a joint function of stimulus factors associated with the target person's behavior and with the observation environment" (Cantor, 1978, p. 4645).

Continuing to pursue the question of how consistency in behavior influences prototypicality judgments, Cantor and Mischel (1979b) constructed a series of characters who were to some degree introverts, extravert, or good samaritans. The characters were pure, mixed, or inconsistent characters. The pure types were purely introverts, extravert, or good samaritans; the mixed (but not inconsistent) types were either part introvert or extravert and always part good samaritan; and the inconsistent types were part introvert and part extravert. The authors found that information about consistent characters was remembered more accurately than information about inconsistent characters. In a free recall test, subjects were the most accurate and wrote the most information about the pure consistent characters, as opposed to the mixed or inconsistent characters. Subjects were more accurate and wrote more information about mixed characters than inconsistent characters. Cantor and Mischel (1979b) conclude that these results support the hypothesis that information is structured and remembered according to how it matches with pre-existing personality prototypes.

Prototype Categorization and Psychodiagnosis

The empirical research in the realm of common object and personality categories can be extended to diagnostic categories. Many theorists of psychopathology view diagnostic categories as a "necessary evil." On the one hand, diagnostic categories are essential in communicating information about clients to other clinicians, as well as to oneself at a future date. Diagnoses may also provide clues to clinicians regarding additional features to inquire about during assessment, as well as clues about etiology and treatment. On the other hand, many clinicians feel that diagnostic categories are too narrow to reflect the diversity of clinical phenomena in the real world. A common criticism that clinicians have of diagnostic systems is that they invariably pigeon-hole individual clients into homogeneous categories that are not particularly descriptive of any one category member. Diagnostic systems have also received the opposite criticism of failing to reduce the heterogeneity that exists among clients with the same diagnosis (e.g., Blum, 1978; King, 1954; Rotter, 1954). Possibly as a result of this within-category heterogeneity, as well as between-category homogeneity, the reliability of diagnostic categories is reduced (e.g., Eysenck, 1952; Scott, 1958). As Cantor, Smith, French & Mezzich (1980) have noted, however, most "critiques of the psychiatric diagnostic system seem to presuppose a classical view of categorization" (p. 181).

An advantage of prototype categorization over monothetic categorization is that the prototype approach explicitly recognizes heterogeneity within diagnostic categories, and homogeneity between diagnostic categories. Thus, compared to the monothetic model of categorization, the prototype model represents a more realistic map of the complexity of various clinical phenomena. Moreover, the prototype approach holds different implications for the appropriateness of particular diagnoses for particular clients. According to the assumptions underlying monothetic theory, atypical and borderline cases reflect aberrations in an otherwise accurate classification system. Such cases suggest that not all of the necessary and sufficient criteria for category membership have been identified. Given this assumption that atypicalities reflect a deficiency in an otherwise refined classification system, the monothetic approach fails to recognize that atypicalities are inevitable real world phenomena. In contrast, the prototype approach to categorization uses the existence of atypicalities to its advantage, basing categorization upon the principle that some category members are more typical of a given category than are other category members.

The relatively low diagnostic reliability of the personality disorders within the current psychiatric nosological system, for example, can be understood as context specific. In borderline cases reliability is expected to be low, whereas in prototypical cases reliability is expected to be high (Blashfield, Sprock, Haymaker, & Hodgin, 1989). Reliability in diagnosing an instance of a particular disorder could be predicted, at least in part, by its degree of similarity to the prototype for that diagnostic category. Methods of determining diagnostic reliability could be adapted to this model. Traditional "indices of reliability based upon the agreement between independent clinicians when assigning patients to discrete diagnostic categories, are inappropriate for fuzzy categories with probabilistic membership. The more significant question is whether clinicians agree when rating prototypicality" (Livesley, 1985b, pp. 356-357). Diagnostic reliability based upon prototypicality ratings may be much higher than diagnostic reliability based upon the "all or nothing" designation of a particular diagnosis.

As noted earlier, the introduction of the DSM-III and the DSM-III-R represents a move from monothetic categorization to polythetic categorization (Cantor et al. 1980; Clarkin, Widiger, Frances, Hurt, & Gilmore, 1983). The DSM-III-R categories do not require that a set of necessary and sufficient diagnostic criteria be met equally by each individual client given a particular diagnosis. Rather, each client need only exhibit a portion of a given set of correlated features in a variety of combinations. For example, the DSM-III-R requires that a client meet any five of nine possible criteria to be diagnosed with Narcissistic

Personality Disorder. The DSM-III-R also represents a move toward prototype categorization in that some diagnoses require the presence of one or two critical features, and then only a portion of several less critical features. For example, a DSM-III-R diagnosis of Major Depression requires that either dysphoric mood or loss of interest and enjoyment in activities be present, along with a portion of less diagnostically efficient criteria. This requirement suggests that some features are considered more prototypical than others of a given diagnostic category. Clarkin et al. (1983) note that the DSM-III is more prototypic than DSM-II in that it "(a) requires multiaxial diagnoses, (b) encourages multiple diagnoses within each of its axes, and (c) does not rely on the assumption that mental disorders are discrete entities" (p. 263). It should be noted, however, that despite this progress toward prototype categorization, the current psychiatric diagnostic system is still based largely upon a monothetic view of categorization.

Prototype Research on Psychodiagnostic Categories

Several major lines of research on prototype categorization have been conducted in the area of psychodiagnosis. Cantor and her colleagues were among the first to bridge basic prototype research on natural categories to clinical research on psychodiagnosis. Cantor et al. (1980)

constructed prototypes of nine DSM-II (American Psychiatric Association, 1968) diagnoses based upon the characteristic features of those categories according to the consensus of 13 experienced clinicians. The diagnostic categories included Functional Psychosis, Schizophrenia, Affective Disorder, Paranoid Schizophrenia, Schizoaffective Disorder, Chronic Undifferentiated Schizophrenia, Manic-Depressive--Manic, Manic-Depressive--Depressed, and Involutional Melancholia. These consensual prototypes served as standards against which to compare actual cases. The authors then selected 12 actual patient cases from a psychiatric hospital. The cases were selected so as to represent four different disorders with three instances of each disorder. The four disorders were Manic-Depressive--Manic, Manic-Depressive-Depressed, Paranoid Schizophrenia, and Chronic Undifferentiated Schizophrenia. For each of the four disorders, there were three cases with different levels of typicality, relative to the standard prototypes obtained previously. The high typical patients' case histories contained 8 to 13 features of the standard prototype, whereas the medium typical patients' case histories contained 5 to 8 prototypical features, and the low typical patients' case histories contained 4 prototypical features. Note that the category features had already been determined to be of high typicality, and the level of typicality was determined by the number of highly typical features included in the case histories, rather than by variations in the typicality of the features themselves.

The authors asked nine experienced clinicians to diagnose the 12 patients on the basis of unedited medical histories. As predicted, diagnostic reliability was significantly higher for cases with a medium or high number of highly typical category features than for cases with a low number of highly typical category features. Cantor et al. (1980) concluded that the number of prototypical features presented in a case has a strong bearing on clinicians' subsequent diagnostic reliability for that case.

In a later study, Cantor and Genero (1986) used the same typical and atypical case histories to examine the effects of a summary versus exemplar prototype teaching paradigm upon the diagnostic reliability of novice versus expert diagnosticians. In the summary prototype condition, a list of prototypical features and their associated typicality weights for four disorders were presented to the subjects. The four disorders were the same as those presented to subjects in Cantor et al.'s (1980) study. Whereas information about the diagnostic categories in the summary prototype condition was given in the form of a list of weighted features, information about the diagnostic categories in the exemplar prototype condition was given in the form of two paragraph-length case vignettes. For each of the four diagnostic categories, two heterogeneous case exemplars were given with different category features reported in each case vignette. The vignettes contained various demographic features to personalize each vignette, and to prevent them from being confused with other vignettes. Cantor and Genero (1987) predicted that both novice undergraduate diagnosticians would diagnose later cases more reliably and confidently when they had received exemplar prototype training, rather than summary prototype training. They predicted that in contrast to the summary prototype training, exemplar prototype training would emphasize the heterogeneity among members of a given diagnostic category. Those results were borne out for the expert clinicians only. The authors found that exemplar prototype training, as opposed to summary prototype training, enhanced the expert clinicians' diagnostic reliability and confidence ratings of the accuracy of their diagnoses. In contrast, summary prototype training, relative to exemplar prototype training, enhanced the novice diagnosticians' reliability and confidence ratings.

Horowitz, Wright, Lowenstein, and Parad (1981b) have extended the prototype categorization research on adult diagnostic categories to child diagnostic categories. According to expert staff members at a child inpatient psychiatric facility, the three most common types of children seen there were the "aggressive-impulsive" child, the "depressed-withdrawn" child, and the "borderline-disorganized" child. Horowitz et al. (1981b) had the clinicians generate a prototype for each category by listing typical behaviors, thoughts, and feelings for children in each category. In Part 2 of this study, the authors compared the expert clinicians' lists of prototypical features for the three child categories with lists of prototypical features provided by college students who had either minimal clinical experience at the facility or no clinical experience. In general, the lists of features provided by the students represented only a subset of the prototypical features listed by the experts. Moreover, some of the novices' listed features were not included in the experts' lists of prototypical features. The authors conclude that compared to novices' prototypes, experts' prototypes appear to be more sophisticated in including important prototypical features and excluding irrelevant features, again illustrating the importance of subject variables in categorization.

In another study, Horowitz, Post, French, Wallis, and Siegelman (1981a) examined the extent to which the number of prototypical features presented in a case history influenced clinicians' diagnostic reliability. Horowitz et al. (1981a) generated prototypes of depression by asking 35 undergraduates to describe a person who was a good example of someone who was depressed. The essays were then divided into 3 groups: the first group contained only 1 prototypical feature, the second group contained 4 to 9 prototypical features, and the third group contained 17 to 20 prototypical features. The word "depressed" was not contained in any of the essays. The essays were then rated by another group of undergraduates along several dimensions. The authors found that persons described in essays with a greater number of prototypical features of depression were rated as significantly more depressed than persons described in essays with fewer prototypical features. Horowitz et al. (1981a) also found that the number of prototypical depressive features influenced clinicians' diagnostic reliability. The authors had clinicians observe a series of videotapes of clients who were depressed to different degrees. They found that clinicians disagreed more about the degree of a client's depression given few prototypical features, suggesting that the low number of features may have "activated" the full prototype for some clinicians, but not for others.

Horowitz et al. (1981a) then turned to a different question. In the diagnostic process, how important is the absence of irrelevant or contradictory features relative to the presence of relevant prototypical features? The authors presented case summaries of 26 psychiatric patients to 20 experienced clinicians, who provided diagnoses. The authors included with each case description a checklist of commonly-given DSM-II diagnoses. The clinicians were asked to check off any diagnoses on the list that might be possible for each client. The authors then generated eight diagnostic sets of case descriptions, with three instances of each of eight DSM-II disorders. The three instances varied from low, medium, to high diagnostic reliability obtained by the clinicians. Horowitz et al. (1981a) then asked undergraduates to rate on a 5-point scale the presence or absence of prototypical features. The diagnostic criteria listed for the corresponding new DSM-III diagnoses were randomly presented to the undergraduates on a checklist. The authors found that relative to cases with a low number of "relevant" category features, cases with a high number of category features were associated with significantly higher diagnostic reliability. In contrast, the number of "irrelevant" noncategory features was not significantly different across cases with different levels of diagnostic reliability.

The question of whether or not competing diagnostic features undermine clinicians' diagnostic reliability is especially pertinent to the personality disorders. Approximately two-thirds of the clients who meet the criteria for one personality disorder meet the criteria for at least one more (Clarkin et al, 1983; Mellsop et al, 1982; Stangl et al, 1984). Moreover, many of the features of personality disorders are shared by other personality disorders, making for an even more fuzzy distinction between these categories.

Prototype Research on Personality Disorders

Results similar to those of Horowitz et al. (1981a) were found by Blashfield, Sprock, Haymaker, and Hodgin (1989) in a study of factors affecting diagnostic judgments of the personality disorders. These investigators found a significant positive correlation between diagnostic reliability and number of category features. The number of features from other personality disorders, however, did not significantly affect diagnostic reliability. These findings contradict Livesley's (1985a, 1985b) hypothesis of a negative correlation between measures of prototypicality and the number of competing diagnostic features.

In one study, Blashfield, Sprock, Pinkston, and Hodgin (1985) attempted to isolate prototypical case examples of the 11 DSM-III personality disorders on the basis of high diagnostic reliability and rapid diagnostic reaction time. Thirty cases were selected from sources such as the DSM-III Case Book (Spitzer, Skodol, & Gibbon, 1981), the DSM-III Training Guide (Webb, DiClemente, Johnston, Sanders, & Perley, 1981), and various professional papers. Among the 30 cases, prototype cases were isolated for only seven of the personality disorders. Unlike other studies, prototypicality in this study was not determined by prototypicality ratings of cases or features, but rather by diagnostic reliability and diagnostic reaction times.

Whereas the Blashfield et al. (1985) study examined the prototypicality of clinical feature combinations in the form of case summaries, other studies have attempted to quantify the prototypicality of specific clinical features. In a study of Borderline Personality Disorder (BPD), Clarkin et al. (1983) calculated conditional probabilities as a means of quantifying the "diagnostic efficiency" of both individual diagnostic criteria and combinations of diagnostic criteria. The conditional probability of a BPD diagnosis associated with the presence of all possible diagnostic features and feature combinations was calculated. The authors note that a feature that is highly prevalent in BPD clients may not necessarily be a discriminating diagnostic feature if that feature overlaps with other personality disorders (e.g., poor interpersonal relationships), or with other person categories (e.g., being in therapy). For example, the authors found that although impulsivity occurred in 100% of the BPD cases, its conditional probability associated with a BPD diagnosis was lowered by the fact that it also occurred in 25% of patients with another personality disorder. The results revealed that both individual and combined features did vary in diagnostic efficiency. The authors concluded that computing the conditional probability of particular feature combinations represents an empirical method of demonstrating how some features and feature combinations are more prototypical than others of the personality disorders.

Whereas Clarkin et al. (1983) examined the prototypicality of features present in BPD subjects, other researchers have obtained prototypicality ratings of features from clinician subjects. For example, Livesley (1986a, 1986b) collected trait and behavior descriptions for each of the 11 DSM-III personality disorders from various sources, including DSM-III, the ninth edition of the International Classification of Diseases (ICD-9; U.S. Department of Health and Human Services, 1980), and texts by Vaillant and Perry (1980), Millon (1981), and Lion (1981). Livesley then designed a series of 22 questionnaires, with a separate trait and behavior questionnaire for each of the 11 personality disorders. One of the 22 questionnaires was then mailed to each of 2,960 psychiatrists belonging to either the Canadian Psychiatric Association or to the American Psychiatric Association. In completing the questionnaires, the psychiatrists were asked to think of prototypical patients rather than average patients, and to rate each item on a 7-point scale on the basis of its prototypicality. The results of the study revealed that the prototypicality ratings for both traits and behaviors were highly reliable across the 938 respondents. Livesley also found that some personality disorders were associated with more unique, distinctive features than were other personality disorders. Moreover, certain features of each disorder were rated as being more prototypical than other features. The study eventuated in a list of the mean prototypicality ratings for various traits and behaviors associated with each of the personality disorders.

Expanding upon Livesley's work on the prototypicality and distinctiveness of traits associated with the personality disorders, Boykin (1987) examined how these variables actually affect the psychodiagnostic process. Neither feature prototypicality nor feature distinctiveness had been investigated previously within the context of the psychodiagnostic process. In his dissertation, Boykin (1987) designed a set of personality profiles based upon Livesley's extensive data on the prototypicality and distinctiveness of the traits associated with the "erratic cluster" of the personality disorders. This cluster includes the Antisocial, Borderline, Histrionic, and Narcissistic Personality Disorders. Each profile consisted of a constant number of six features associated with a given personality disorder. The features were varied according to whether they were distinctive or shared with other personality disorders, and were of either high or low prototypicality or "centrality."

A total of 20 licensed clinical psychologists were asked provide diagnoses and prototypicality ratings for a series of

32 personality profiles. A composite diagnostic accuracy-typicality score was calculated for each subject. Diagnostic accuracy was determined according to which disorder the investigator had intended to portray in each profile, based upon the feature prototypicality ratings provided by Livesley. Boykin hypothesized that clinicians would assign higher accuracy-typicality scores to cases with features that were more prototypical of the intended disorder than to cases with features that were less prototypical. It was also predicted that cases with fewer features that are shared with other personality disorders would be diagnosed as more prototypical than cases with more shared features. Additionally, Boykin predicted that cases with both highly prototypical and distinctive features would be assigned higher prototypicality ratings than cases with both low prototypical and shared features.

As predicted, significant main effects for feature prototypicality and distinctiveness, as well as a significant interaction between the two were revealed. Relative to profiles with low central features, those with high prototypical features were associated with much higher accuracy-typicality scores. Moreover, relative to profiles with features shared by other disorders, profiles with distinctive features were associated with higher accuracy-typicality scores. As predicted an interaction emerged, in which profiles with high prototypical and distinctive features yielded the highest accuracy-typicality scores, whereas profiles with low prototypical and shared features yielded the lowest accuracy-typicality scores. Additionally, the remaining conditions consisting of either low prototypical, distinctive features, or high prototypical, shared features were associated with moderate diagnostic reliability and prototypicality ratings.

Statement of Purpose

The purpose of the present study was to examine how three critical variables interact to impact clinicians' diagnosing behavior. The results of previous research suggest that the number of category features and the prototypicality of those features are among the most important factors affecting judgments of category membership. The factor that has received the most attention in the literature is number. Studies of the categorization of basic objects (e.g., Rosch & Mervis, 1975), of personality types (e.g., Cantor, 1978), and of psychodiagnosis (Blashfield et al., 1989; Cantor et al., 1980; Horowitz et al., 1981a) reveal that the higher the number of category features comprising a stimulus, the more likely that stimulus will be judged as a prototypical member of that category. Note, however, that a review of the studies on the importance of category feature number (Blashfield et al., 1989; Horowitz et al., 1981a; Rosch & Mervis, 1975), reveals that the category features included in the number that was varied were derived from previously constructed prototypes of diagnostic categories. Therefore, rather that simply varying the number of category features, these studies varied the number of highly prototypical category features. Although this process precluded an analysis of the separate effects of feature number and feature typicality, the highly significant effect for number of highly prototypical features was attributed, in each of the studies, to category feature number, rather than to feature typicality. The present study is the first to examine the independent effects of category feature number and category feature typicality.

The effects of the prototypicality of individual category features have been examined in several studies (Cantor, 1978; Cantor & Mischel, 1979b; Clarkin et al., 1983; Schutte et al. 1985), with the general finding that stimuli composed of more prototypical features are more likely to be judged as members of the corresponding category, and are rated as more prototypical members of the category, relative to stimuli composed of less prototypical features. Again, however, aside from the present study, only one study (Boykin, 1987) has examined feature prototypicality as a determinant of psychodiagnostic judgments.

Finally, the proportion of category features to the total number of features, termed "feature dominance", has been examined in only one study (Cantor, 1978). In this study, however, dominance and number were not varied factorially, but rather, were combined in a single breadth-dominance-differentiation index. Therefore, these factors could not be analyzed for their independent contribution to prototypicality judgments. Moreover, Cantor's (1978) study did not examine dominance as a determinant of the actual psychodiagnostic process. Three other studies examined the impact of mixed or irrelevant features upon prototypicality judgments. Cantor and Mischel (1979) found that features associated with consistent characters were recalled more accurately than were features associated with mixed personality characters, than were features associated with contradictory personality characters. Although the authors did not examine dominance per se, they found that the presence of extraneous and contradictory noncategory features significantly impacted the accuracy with which information about the characters was recalled. In another study by Horowitz et al. (1981a), variations in the number of "irrelevant" noncategory features in clinical case histories were not associated with different levels of diagnostic reliability. Rosch and Mervis (1975) found that the number of contrasting noncategory features of a category member was highly negatively correlated with prototypicality ratings of that category member. The present study was the first to examine dominance as the proportion of category features over all presented features, and to vary dominance and number factorially prior to the psychodiagnostic process. The present study varied dominance and number factorially to compare the relative importance of the number of category features present with the proportion of category features.

As noted above, the present study examined the role of feature number, feature typicality, and dominance as determinants of the psychodiagnostic process. By examining the possibility of main effects for each of these factors across the other two factors, the present study assessed the impact of each of the variables across various contexts. This study examined how feature number, feature typicality, and dominance impact diagnostic judgments about the DSM-III personality disorders. The personality disorders readily lend themselves to an investigation of prototype categorization for several reasons. First, the criteria for the various personality disorders often overlap, thereby highlighting the degree of similarity between the different personality disorders. Second, clients with one personality disorder often meet some or all of the criteria for other personality disorders, thereby highlighting the heterogeneity of clients who meet the criteria for a given personality disorder. Third, although personality disorder diagnoses are widely used by clinicians in describing clients, these diagnoses typically are associated with poorer diagnostic reliability than other diagnostic categories. This lower diagnostic reliability may reflect the inappropriateness of a monothetic categorization system for clinical diagnoses in general, and for the more nebulous personality disorders in particular.

On the basis of past research, main effects were predicted for feature typicality and feature number. Specifically, it was hypothesized that personality descriptions containing highly typical category features would be associated with higher measures of diagnostic reliability, diagnostic "accuracy," and prototypicality compared with personality descriptions containing relatively low typical category features. Similarly, personality descriptions containing a high number of category features were hypothesized to be associated with higher measures of diagnostic reliability, diagnostic "accuracy," and prototypicality relative to personality descriptions containing a low number of category features.

Based upon past research, it was predicted that the variables of feature typicality and feature number would be so highly significant, and would consume so much of the variance in any one analysis, that an otherwise significant main effect for dominance would not emerge. Therefore, a main effect for dominance was predicted only upon the condition that enough variance would be spared by what were predicted to be the highly significant main effects for feature typicality and feature number. Specifically, it was hypothesized that personality descriptions containing a high proportion of category features to noncategory features would be associated with higher measures of prototypicality, diagnostic "accuracy," and diagnostic reliability compared with personality descriptions containing a low proportion of category features. In addition, interactions between the variables of dominance, feature typicality, and feature number were predicted. In particular, dominance was predicted to have a greater influence upon diagnostic judgments under certain conditions in which either feature typicality or feature number, or both feature typicality and feature number, were relatively low.

CHAPTER II

METHOD

Subjects

A total of 32 licensed doctoral-level clinical and counseling psychologists practicing within a 100-mile radius of Greensboro, North Carolina, served as participants in the study. Potential subjects were randomly selected from the telephone directories of Greensboro, High Point, Chapel Hill, Durham, and Raleigh, North Carolina, and from a mailing list of licensed clinical psychologists provided by the North Carolina State Board of Examiners of Practicing Psychologists. Initially, 125 potential subjects were solicited, given a projected positive response rate of 30%. Subjects were mailed a solicitation letter explaining the nature of the study, and a response form on the back of a pre-addressed stamped postcard (see Appendix A for all materials mailed to subjects). The solicitation letter stated that potential participants must fulfill the following inclusion criteria: (a) permanent licensure as a psychologist for at least three years; (b) clinical work primarily with adults; and (c) expertise in diagnosing personality disorders. Potential participants were asked to return the postcard indicating whether or not they would be interested in participating in the study. Initially, 35 clinicians agreed to participate. Experimental materials were mailed to the first 32 of these clinicians who agreed to participate. Only 27 of the 32 subjects who were sent the experimental materials actually completed the study. The other five subjects either reported that they did not find time to complete the study, or mailed back incomplete materials. Therefore, experimental materials were mailed to the remaining three clinicians who had initially agreed to participate, as well as to two newly solicited subjects, making for a final total of 32 subjects.

The subjects also completed a brief demographics questionnaire, the results of which are presented in Table 1 (This table and all subsequent tables are located in Appendix B). The number of subjects that were included in the calculation of each demographic measure reported in Table 1 varied because some subjects did not respond to all of the demographic questions. As depicted in the table, an equal number of male and female clinicians participated as subjects. A total of 91% of the subjects worked primarily in private practice, and 9% worked primarily at a hospital setting. The clinicians' mean age was 41.5 years, and their mean number of years of clinical experience was 13.6 years. The mean number of personality disorder cases (with or without coexisting Axis I diagnoses) assessed in the last six months was 14.4 cases.

Experimental Design

Three experimental designs, depicted in Appendix C, were created so that specific combinations of the designs could later be analyzed. Design 1 is a two (medium typicality vs. high typicality) by two (low category feature number vs. high category feature number) factorial design with each of the factors being within-subjects. Design 2 is a two (medium typicality vs. high typicality) by two (low category feature number vs. high category feature number) factorial design with each of the factors being within-subjects. Design 3 is a two (medium typicality vs. high typicality) by two (low category feature number vs. high category feature number) factorial design with each of the factors being within-subjects. As illustrated in Appendix C, Designs 1, 2, and 3 differ from one another in the proportion of category features over total features, as well as in the number of total features included in each design. For Design 1 and Design 2, one-third of the total number of features within each cell are category features. For Design 3, however, two-thirds of the total number of features within each cell are category features. Therefore, the dominance of the category features over the total number of features within the cells of Design 1 and Design 2 is low relative to the dominance of the category features within the cells of Design 3. For Design 1 and Design 3, the number of total features within each cell is either 3 or 9. For Design 2, however, the number of total features within each cell is either 6 or 18. Therefore, overall, the total number of features within each cell of Design 1 and Design 3 is lower relative to the total number of features within the corresponding cells of Design 2. By combining all three designs, there were 12 different experimental conditions.

Each subject received twelve personality profiles, corresponding to each of the twelve experimental conditions. Of the twelve profiles, three profiles corresponded to each of the four personality disorders that comprise the "erratic cluster" according to DSM-III-R. These are the antisocial, borderline, histrionic, and narcissistic personality disorders. The disorders portrayed in the profiles were counterbalanced across the experimental conditions.

The twelve experimental conditions were determined by the following four independent variables or factors: feature typicality, feature dominance, feature number, and total feature number. The first factor is the level of typicality of the category features included in the personality profiles. The second factor, feature dominance, is the relative proportion of category features to the total number of features presented in the personality profiles. The third factor is the number of category features presented in the profiles. The fourth factor is the total number of both category and noncategory features depicted in the profiles. These independent variables, as well as the dependent variables, are operationalized below.

Stimulus materials

The entire set of materials sent to each subject upon their agreement to participate in the study is presented in Appendix A, and includes the following: an introductory letter, experimental instructions, the twelve profile sheets, a demographics questionnaire, and a stamped envelope, pre-addressed to the principal investigator. At the bottom of each of the profile sheets was a list of all 11 personality disorders described in the DSM-III-R. Subjects were asked to indicate on this list the diagnosis that best fit the person described in the profile. Next to each listed disorder is a 1-7 Likert-scale; subjects were to rate how typical the person described in the profile was of each of the disorders, with one being the least typical case, and seven being the most typical case.

The personality profiles were based upon data kindly provided by Dr. John Livesley of the University of British Columbia. As described previously, Livesley (1986a, 1986b) generated a set of lists of the clinical features associated with each of the 11 personality disorders. The features were given typicality ratings by nearly 1000 psychiatrists. Based upon these data, Livesley has listed the clinical features associated with each disorder in descending order of typicality, from those features that were rated as being most typical of the disorder, to those that were rated as being least typical of the disorder. Livesley then divided the feature lists for each disorder into four quartiles, with the first quartile composed of the most typical features, and the fourth quartile composed of the least typical features.

Livesley provided the experimenter with the first, second, third, and fourth quartiles of the feature lists for the "erratic cluster" of the personality disorders, which includes the Antisocial, Borderline, Histrionic, and Narcissistic personality disorders. He also provided the fourth quartile of the feature lists for the other seven personality disorders that are not included in the erratic cluster.

Based upon these data, a total of 48 personality profiles were constructed by the present author, with a profile representing each of the four erratic cluster personality disorders in each of the twelve experimental conditions. Although each subject received twelve profiles, the disorders portrayed by the profiles were counterbalanced across the experimental conditions so that each subject received only three instances of each of the four disorders. The twelve profiles sent to each subject were also randomly ordered to control for sequence effects.

Additionally, 4 versions of each of the 48 profiles were created in which the order of the features' presentation was varied randomly. Thus, for any 1 of the 48 profiles, there were 8 clinicians who received that profile, and among those 8 clinicians, only 2 received the same order of feature presentation. The low number of features in some of the profiles prevented the creation of 8 versions of feature order that would have corresponded to the 8 clinicians who received any 1 of the 48 profiles.

The experimental designs, depicted in Appendix C, required that the profiles be constructed in such a way that typicality, dominance, category feature number, and total feature number were varied factorially. For Design 1, the numbers of category features chosen for the low and high number conditions were one and three, respectively. These numbers appear as the numerators of the fractions shown within the cells of experimental designs. The dominance or proportion of category features over the total number of features presented is 1/3 for the low dominance condition and 2/3 for the high dominance condition. Therefore, for Design 1, the number of total features for low and high total number were three and nine, respectively. For both Designs 2 and 3, the number of category features varied from two to six within each low or high dominance condition. The proportion of category features over the total number of features, however, remained constant within each dominance condition.

The typicality of the category features varied from medium to high, rather than from low to high. The medium versus high typicality conditions were selected for two reasons. First, the noncategory features were of low typicality. Thus, it was necessary to present category features of at least medium typicality in order to obtain a sufficient contrast with the low typicality noncategory features. Secondly, the high versus medium typicality conditions were selected to reduce the contrast between the typicality conditions. This was expected to reduce the likelihood that a highly robust main effect for typicality would override any interactions among the independent variables. The medium typicality features were selected from the second and third quartiles of Livesley's data for the erratic cluster, whereas the high typicality features were selected from the first quartile.

The noncategory features were not features that were completely irrelevant to the diagnostic process; they were contrasting features that were somewhat associated with the other personality disorders. The noncategory features for each profile constituted a combination of features, selected from the fourth quartiles of the typicality lists for the seven personality disorders that were not included in the erratic cluster. Noncategory features that were not distinctive to any one personality disorder were selected. Thus, all noncategory features were both low distinctive and low typical.

Dependent variables

The subjects were asked to provide a 1-7 rating of how typical each profile was of each of the 11 personality disorders, with one being the least typical, and seven being the most typical. Subjects also provided a diagnosis that they felt best fit the person described in the profile.

As discussed previously, Boykin (1987) used a dependent variable that combined accuracy and typicality by assigning a positive value to the 1-7 typicality ratings of accurate diagnoses, and a negative value to the 1-7 typicality ratings of inaccurate diagnoses. In the present study, however, an alternative dependent measure was judged to be more compatible with the prototype model. (For further discussion of the dependent measure used by Boykin (1987), as well as experimental results from the present study using this measure, see Appendix D).

The first dependent variable is each clinician's prototypicality rating for the disorder that the category features in each profile had been intended to describe, divided by the mean typicality rating given by that clinician for the other 10 personality disorders that were intended to be "noncategory" diagnoses. This dependent measure for each individual subject profile is represented in the following proportion:

$$\frac{\text{typicality rating for the intended category dx}}{\text{mean of the typicality ratings for the noncategory dx's}}$$

Note that, for each profile, each clinician assigned a rating of how typical the profile was of all 11 personality disorders. This being the case, the clinicians' typicality ratings for the intended category diagnosis could be assessed whether or not they actually assigned that particular diagnosis to the profile.

Examining the proportion of prototypicality for the intended category diagnosis over that for the noncategory diagnoses parallels Cantor's emphasis that the diagnostic process involves determining not only what the category is, but also what the category is not. For example, suppose that two clinicians assign a typicality rating of seven to the intended category diagnosis of a particular personality profile. The first clinician may have assigned a mean typicality rating of two to the other, noncategory diagnoses, whereas the second clinician may have assigned a mean typicality rating of five to the noncategory diagnoses. These cases illustrate that even if clinicians assign the same ratings to the category diagnosis, the significance of those ratings depends upon the context of noncategory prototypicality in which they were assigned. A further benefit of using this proportion as a dependent variable also controls for variability in the data due to any possible tendencies of subjects to give either low or high typicality ratings in general.

Since prototypicality was rated on a 1-7 scale, the highest possible proportion of category to noncategory prototypicality was 7/1 or seven. The lowest possible proportion was 1/7. If the category and noncategory diagnoses were rated with equal typicality (e.g., 5/5 or 7/7), then the proportion would equal one.

The second dependent variable is similar to the first except that it combines typicality with diagnostic reliability, rather than with diagnostic accuracy. The second dependent variable was each subject's prototypicality rating of the diagnosis most commonly assigned by the clinicians in a given experimental condition, divided by the subject's mean typicality rating for the remaining nonmodal diagnoses. This dependent measure for each individual subject profile is represented in the following proportion:

$$\frac{\text{typicality rating for the modal dx}}{\text{mean of the typicality ratings for the nonmodal dx's}}$$

This measure allows for an examination of the relative typicality ratings for the most reliably assigned diagnoses of the personality profiles. To the extent that the subjects' modal diagnosis in a given condition corresponded to the diagnoses intended by the experimenter according to Livesley's (1986a, 1986b) data, then this second dependent measure would be identical to the first measure described above. The inclusion of this measure was important given Boykin's (1987) finding that for cases in which clinicians' diagnoses disagreed with the intended diagnosis of the profiles, the clinicians often tended to agree with each other upon an alternative diagnosis.

The third dependent variable for each clinician is the typicality rating of the diagnosis chosen by each clinician as being the most prototypical of the hypothetical client, divided by the mean of the typicality ratings for the other diagnoses that were not chosen by this clinician. This dependent measure for each individual subject profile is represented in the following proportion:

$$\frac{\text{typicality rating for the chosen dx}}{\text{mean of the typicality ratings for the nonchosen dx's}}$$

The fourth dependent variable is the number of chosen diagnoses that correspond to the intended diagnosis in a given condition. Finally, the fifth dependent variable is the number of chosen diagnoses that correspond to the modal diagnosis in a given condition.

Procedure

After consenting to participate in the study, the subjects received the complete packet of materials described above, including a set of twelve personality profiles. It was expected that more clinicians would be willing to participate if the study were conducted through the mail, rather than in person during their regular office hours. In this way they were able to complete the experiment at a time that was most convenient for them. The subjects were asked to complete the experiment in one sitting within one week of receiving their packet of materials. Prompts were sent to those clinicians whose materials were not returned within three weeks.

In a procedure similar to that employed by Boykin (1987), subjects were instructed to complete each of the profiles one at a time, in the order presented. Subjects were reminded that the experiment was not a test of their clinical abilities, but rather a survey of their own impressions of the personality disorders, based upon their experience with actual clients. The subjects were therefore asked not to consult any outside sources such as diagnostic manuals. They were also asked to spend a maximum of three - four minutes per case. The entire procedure was designed to take approximately one hour. The subjects were instructed to read each personality profile, and to imagine the person described in the profile. They were then asked to provide a 1-7 rating of how typical that person was of each of the 11 personality disorders, with one being the least typical, and seven being the most typical. After completing the typicality ratings, the subjects were asked to put a check mark next to one of the 11 listed personality disorders that they felt best fit the person described in the profile. After completing these steps for each of the twelve profiles, the subjects completed the brief demographics questionnaire described above. Upon receipt of all the participants' materials, the investigator immediately sent the clinicians a written debriefing statement fully explaining the nature of the study, and thanking them for their participation (Appendix A).

CHAPTER III

RESULTS

Overview

This chapter is divided into five sections, corresponding to each of the five dependent variables. For the reader's convenience, a summary list of these dependent variables is presented in Table 2 (see Appendix B). In addition, a one-page summary chart of the experimental results for the five dependent variables is presented in Table 3. Within each of the five sections of this chapter, the effects of each of the independent variables are discussed. Also addressed is the consistency or inconsistency of each of these effects across the three different combinations of Designs 1, 2, and 3 (see Appendix C). These combinations are as follows:

Design Combination 1 + 2;

Design Combination 1 + 3;

Design Combination 2 + 3.

The effects of the three independent variables upon each of the five dependent variables were analyzed within the context of these three design combinations.

The three combinations of Designs 1, 2, and 3 were each analyzed separately in order to examine the effects of the experimental variables while controlling for the effects of total feature number. If there had been no control for total feature number, then it would have covaried with dominance. Specifically, if the number of category features were held constant across the low and high dominance conditions, then by definition, the high dominance conditions would have been associated with a lower number of total features relative to the low dominance conditions. The addition of total feature number as an experimental control ensured that any main effect for dominance could not be attributable to the total number of features presented. In order to implement this control, the three combinations of the experimental designs described above were necessary.

As illustrated in Appendix C, three of the four independent variables were analyzed in each of the design combinations. The specific set of independent variables varied according to the design combination, as described below. First, Design Combination 1 + 3 was analyzed as a 2 (medium typicality vs. high typicality) by 2 (low dominance vs. high dominance) by 2 (low category feature number vs. high category feature number) factorial design, with all factors being within-subjects. Second, Design Combination 1 + 2 was analyzed as a 2 (medium typicality vs. high typicality) by 2 (low total feature number vs. high total feature number) by 2 (low category feature number vs. high category feature number) factorial design, with all factors being within-subjects. Third, Design Combination 2 + 3 was analyzed as a 2 (medium typicality vs. high typicality) by 2 (low dominance vs. high dominance) by 2 (low category feature number vs. high category feature number) factorial design, with all factors being within-subjects.

In general, the experimental results provided support for the above hypotheses, but with a few exceptions. As predicted, a highly significant main effect across all three designs, in the predicted direction, was revealed for category feature typicality. Contrary to predictions, a main effect for category feature number was revealed for only one of the three combinations of the experimental design. A highly significant effect across all three designs, in the predicted direction, was revealed for dominance. As predicted, no main effect for the experimental control of total feature number was revealed. In addition, no significant interactions among any of the factors were revealed.

<u>Dependent Variable 1: The Prototypicality Rating for the Intended Diagnosis Divided by the Mean of the Ratings for the Nonintended Diagnoses</u>

Three separate 2 by 2 by 2 repeated measures analyses of variance (ANOVA's) were performed for each of the three Design Combinations (1 + 3, 1 + 2, and 2 + 3), the levels of which are described above. Across all three design combinations,

a main effect for typicality was revealed. Consistent with the predicted results, the clinicians were significantly more likely to rate the hypothetical client as being more prototypical of the intended diagnosis, relative to the other diagnoses, when the features in the profile were of high typicality, compared to when the features were of medium typicality. The F and probability values for the typicality main effect for each combination of designs are as follows:
Design Com. 1 + 3, $F(1, 222) = 31.516$, $p < .0001$ (Table 4);
Design Com. 1 + 2, $F(1, 218) = 26.203$, $p < .0001$ (Table 5);
Design Com. 2 + 3, $F(1, 216) = 39.715$, $p < .0001$ (Table 6).

Across both of the two design combinations that included dominance as a factor (1 + 3 and 2 + 3), a main effect for dominance was revealed. The clinicians were significantly more likely to rate the hypothetical client as being more prototypical of the intended diagnosis when the category features represented a highly dominant proportion of all the features presented relative to when the category features represented a less dominant proportion of all the features presented. The F and p values for the dominance main effects are as follows:
Design Com. 1 + 3, $F(1, 222) = 13.137$, $p < .0001$ (Table 4);
Design Com. 2 + 3, $F(1, 216) = 21.946$, $p < .0001$ (Table 6).

Although the number of category features was included as an independent variable in all three design combinations, a main effect for this variable was revealed only for Design Combination 2 + 3. Subjects were significantly more likely to rate the hypothetical client as being more prototypical of the intended diagnosis relative to the other diagnoses when the profile contained a high number of category features, compared to when the profile contained a low number of category features, $F(1, 216) = 5.616$, $p = .019$ (Table 6). As opposed to the other two design combinations, Design Combination 2 + 3 was unique in that it kept the exact number of category features within each level constant across the two designs (see Appendix C). That is, there were two category features in the low category number level of both Designs 2 and 3, and there were six category features in the high category number level of both designs. This fact may explain why a significant effect for the number of category features was obtained only for Design Combination 2 + 3.

Consistent with predictions, no main effects nor interactions were revealed for total feature number, which was included as an experimental control variable in Design Combination 2 + 3. As predicted, this finding was consistent across all five dependent variables. Moreover, there were no significant interaction effects with any of the experimental variables for this, nor any other, dependent variable.

Dependent Variable 2: The Prototypicality Rating for the Modal Diagnosis Divided by the Mean of the Ratings for the Nonmodal Diagnoses

The second dependent variable was each subject's prototypicality rating of the diagnosis most commonly assigned by the clinicians in a given experimental condition, divided by the subject's mean typicality rating for the remaining nonmodal diagnoses. The results of the three ANOVA's performed on the data for the second dependent variable were consistent with the results performed on the data for the first dependent variable, with the exception that no significant main effect was revealed for the category feature number. Across all three Design Combinations (1 + 3, 1 + 2, and 2 + 3), the ANOVA's revealed a main effect for feature typicality. As predicted, the clinicians were significantly more likely to rate the hypothetical client as being more prototypical of the modal diagnosis, relative to the other diagnoses, when the features in the profile were of high typicality, compared to when the features were of medium typicality. The F and p values for the typicality main effects are as follows:

Design Com. 1 + 3, $F(1, 134) = 9.990$, $p < .002$ (Table 7);

Design Com. 1 + 2, $F(1, 136) = 3.917$, $p < .050$ (Table 8);

Design Com. 2 + 3, $F(1, 154) = 9.849$, $p < .002$ (Table 9).

Across both design combinations that included dominance as a factor (1 + 3 and 2 + 3), a main effect for dominance was revealed. The clinicians were significantly more likely to rate the hypothetical client as being more prototypical of the modal diagnosis when the category features represented a highly dominant proportion of all the features presented relative to when the category features represented a less dominant proportion of all the features presented. The F and p values for the dominance main effects are as follows:
Design Com. 1 + 3, $F(1, 134) = 5.038$, $p = .026$ (Table 7);
Design Com. 2 + 3, $F(1, 154) = 12.342$, $p = .001$ (Table 9).

As noted above, the three ANOVA's performed on the data for this dependent variable revealed no significant main effect for number of category features, nor for number of total features. Similarly, none of the interaction effects approached statistical significance.

It was predicted that the degree of correspondence between the clinicians' modal diagnoses and the intended diagnoses would determine whether or not the effects revealed for the modal diagnoses would correspond with the effects revealed for the intended diagnoses. The extent to which the clinicians' modal diagnoses corresponded with the intended diagnoses is discussed in the section of this chapter entitled "Dependent variable 5 - Number of Chosen Diagnoses Corresponding to the Modal Diagnoses."

Dependent Variable 3 - Prototypicality Rating for the Chosen Diagnosis Divided by the Mean of the Ratings for the Nonchosen Diagnoses

No significant main effects nor interactions were revealed by the three ANOVA's performed upon the data for this dependent variable. The subjects' ratings of how prototypical the hypothetical client was of the chosen diagnoses were not differentially affected by the levels of typicality, dominance, category feature number, and total feature number (Tables 10, 11, and 12). Although an effect for feature typicality and dominance was revealed for both the intended and the modal diagnosis, the typicality and dominance of the features associated with the intended diagnosis did not affect the clinicians' rating of how prototypical the patient was of the chosen diagnosis, relative to the nonchosen diagnoses. It was predicted that the degree of correspondence between the clinicians' individual chosen diagnoses and the intended diagnoses would determine whether or not the effects revealed for the chosen diagnoses would correspond with the effects revealed for the intended diagnoses. The extent to which the clinicians' chosen diagnoses corresponded with the intended diagnoses is discussed in the following section.

Dependent Variable 4 - Number of Chosen Diagnoses Corresponding to the Intended Diagnosis

Overall, there were a total of 384 cases in which a chosen diagnosis could have been provided by the clinicians. These 384 cases corresponded to the twelve experimental conditions, the four intended diagnoses within each condition, and the eight clinicians asked to diagnose the hypothetical clients representing each of the four intended diagnoses. Due to seven missing cases in which no chosen diagnosis was given by the clinicians, 377 of the 384 cases were actually assigned a chosen diagnosis. Among these 377 cases in which a chosen diagnosis was provided, there were 140 or 37% of the cases in which the chosen diagnosis was the same as the intended diagnosis. Table 13 presents the number of chosen diagnoses that matched each of the four intended diagnoses in each of the twelve experimental conditions. There was a maximum of eight possible "correct" chosen diagnoses in each subcondition, in which the chosen and intended diagnoses could match. Due to missing data, however, six of the 48 subconditions had only seven valid cases, and one subconditon had only six valid cases.

The number of cases in which the chosen diagnosis matched the intended diagnosis for each hypothetical client was calculated across the four possible intended diagnoses for each experimental condition in Designs 1, 2, and 3, and in

Design Combinations 1 + 3, 1 + 2, and 2 + 3 (Tables 14, 15, 16, and 17, respectively). The chi-square test was judged as inappropriate for these data because the assumption of independence of observations was not met. That is, each of the 32 subjects diagnosed 12 profiles corresponding to the each of the 12 experimental conditions. Since the same subjects responded to more than one profile, observations within the data set were not completely independent of one another. An alternative nonparametric test was therefore employed. Like the chi-square test, the McNemar test is appropriate for nominal data with two samples. Unlike the chi-square test, however, the two samples compared in the McNemar test can be related. The McNemar test yields a statistic that is distributed as chi-square with one degree of freedom (Segal, 1956). A discussion of the procedures used in the McNemar test is presented in Appendix E. A total of nine McNemar tests were performed on the data for this dependent variable, and corresponded to each of the independent variables that were examined in the three design combinations (see Tables 18, 19, 20, and 21).

A significant main effect was revealed for typicality in all three design combinations. As predicted, the clinicians were significantly more likely to give diagnoses that corresponded to the intended diagnosis when the features of the hypothetical client were of high typicality, relative to when they were of medium typicality. The corresponding chi-square and probability values for the typicality main effect for each combination of designs are as follows:

Design Combination 1 + 3, $X = 22.74$, $p < .001$ (Table 18);

Design Combination 1 + 2, $X = 14.38$, $p < .001$ (Table 18);

Design Combination 2 + 3, $X = 31.13$, $p < .001$ (Table 18).

A significant main effect for dominance was also revealed for both design combinations that included dominance as a factor. The clinicians were significantly more likely to give diagnoses that corresponded to the intended diagnosis when there was a high proportion of category features over noncategory features, relative to when there was a low proportion of category features over noncategory features. The corresponding chi-square and probability values for each combination of designs are as follows:

Design Combination 1 + 3, $X = 15.85$, $p < .001$ (Table 19);

Design Combination 2 + 3, $X = 16.98$, $p < .001$ (Table 19).

A significant effect for number of category features was also revealed, but only for Design Combination 2 + 3. As noted above, this was the only combination of designs in which the exact number of category features for the two levels of this factor was kept constant across the two designs. For this design combination, the clinicians were significantly more likely to give diagnoses that corresponded to the intended diagnosis when there was a high number of category features, relative to when there was a low number of category features (Design Combination 2 + 3, $X = 5.92$, $p < .02$; Table 20). As predicted, no significant effects were revealed for the total number of both category and noncategory features associated with each hypothetical client (Table 21).

Dependent variable 5 - Number of Chosen Diagnoses Corresponding to the Modal Diagnosis

There were 48 subconditions, corresponding to the four intended diagnoses examined within each of the twelve experimental conditions. Within each subcondition, there were eight clinicians who were asked to diagnose the profile for that subcondition. The diagnoses provided by the clinicians within each subcondition were examined to determine whether or not a modal diagnosis occurred within that subcondition. A modal diagnosis was defined simply as a diagnosis that was more frequently assigned than any other diagnosis in that subcondition. The number of times that a modal diagnosis could occur in a given subcondition ranged from two times to eight times. For example, as illustrated in Table 13, the subcondition for the Antisocial Personality Disorder intended diagnosis in the first cell of Design 1 was associated with the modal diagnosis of Obsessive-Compulsive Personality Disorder. This diagnosis was assigned by four of the eight clinicians within that subcondition. Two other clinicians diagnosed the profile as Schizotypal Personality Disorder, and the remaining two diagnosed it as Schizoid Personality Disorder and as Mixed Personality Disorder, respectively. In subconditions in which more than one diagnosis occurred with the same frequency and no single diagnosis occurred at a higher frequency, no modal diagnosis was determined.

Table 13 presents the number of chosen diagnoses within each of the 48 subconditions that matched the four intended diagnoses, as well as the number of chosen diagnoses that matched the modal diagnosis for in each subcondition. Among the 48 subconditions, a modal diagnosis occurred in only 36 subconditions. Among the 36 subconditions in which a modal diagnosis did occur, there were 17 subconditions in which the modal diagnosis matched the intended diagnosis.

The number of cases in which the chosen diagnosis matched the modal diagnosis was calculated across the four possible intended diagnoses for each experimental condition in Designs 1, 2, and 3, and in Design Combinations 1 + 3, 1 + 2, and 2 + 3 (Tables 22, 23, 24, and 25, respectively). Consistent with the procedure described above, a total of nine McNemar tests were performed on the number of chosen diagnoses that corresponded to the modal diagnosis (see Tables 26, 27, 28, and 29). These analyses revealed a pattern of results similar to the results found when the intended diagnosis was used as the standard for comparison. The pattern of results differed, however, in that only three of the nine analyses (as opposed to six of the nine analyses above) yielded significant effects. Moreover, the magnitude of the effects was smaller relative to when the intended diagnosis was used as the standard for comparison. A significant effect for typicality was revealed across two of the three combinations of designs (1 + 3 and 2 + 3). For these two design combinations, the clinicians were significantly more likely to give diagnoses that corresponded to the modal diagnosis when the features of the hypothetical client were of high typicality, relative to when they were of medium typicality. The corresponding chi-square and probability values for each combination of designs are as follows:

Design Combination 1 + 3, $X = 4.05$, $p < .05$ (Table 26);

Design Combination 2 + 3, $X = 5.76$, $p < .02$ (Table 26).

A significant main effect for dominance was also revealed, but only for the Design Combination 2 + 3. For this combination of designs, the clinicians were significantly more likely to give diagnoses that corresponded to the modal diagnosis when there was a high proportion of category features over noncategory features, relative to when there was a low proportion of category features over noncategory features (Design Combination 2 + 3, $X = 4.97$, $p < .05$; Table 27). No significant effects were revealed for number of category features, nor for number of total features for any of the three design combinations (Tables 28 and 29, respectively).

As an additional measure, the data were also visually inspected using the more stringent criteria that the chosen diagnoses (a) match the intended diagnosis, and (b) be agreed upon by at least half of the clinicians in a given subcondition. There were 13 subconditions in which the diagnosis matched the intended diagnosis and was also assigned by at least 50% of the clinicians within that subcondition. These subconditions are underlined in Table 13. Note that the exact number of clinicians that represented 50% of the clinicians in a subcondition varied across the subconditions due to missing data for seven of the 48 subconditions. Within these 13 subconditions, 61% of the chosen diagnoses matched the intended diagnosis and were agreed upon by at least half of the clinicians in that subcondition. Table 30 presents the data on a more molar scale, for each of the twelve experimental conditions, collapsed across intended diagnosis. Even more striking than the data presented in Table 22 for the number of chosen diagnoses simply matching the modal diagnoses, a visual inspection of the data in Table 30 reveals that both category feature typicality and dominance strongly influenced the number of chosen diagnoses that matched the intended diagnosis and were agreed upon by at least half of the clinicians in that subcondition.

CHAPTER IV
DISCUSSION

Verbal specifications of categories of phenomena can enhance the effectiveness of one's behavior with respect to those phenomena. Psychodiagnostic classification systems are verbal specifications of categories of maladaptive behavioral phenomena. These classification systems are designed with the goal of enhancing the clinician's ability to assess and treat these phenomena effectively. To the extent that the categories specified in diagnostic nosologies correspond to functionally meaningful subtypes of psychopathology, the reliability and validity with which clinicians apply the DSM diagnoses is enhanced. There is wide agreement in the field of clinical psychopathology that the utility of diagnostic categories was enhanced with the evolution from the second edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-II; American Psychiatric Association, 1968) to the third edition of this manual (DSM-III; American Psychiatric Association, 1980) as the standard for psychiatric nomenclature (Klerman, 1986; Millon, 1986).

Whereas the DSM-II was based on a monothetic model of categorization, the DSM-III was based on a polythetic model of categorization. Monothetic categories specify a list of necessary and sufficient criteria for category membership. Polythetic categories also specify a list of criteria, but require that only a certain number of those criteria be met for category membership. It has been suggested that the existing polythetic categories of DSM-III-R be further refined according to a prototype model in future psychiatric nosologies, at least for the personality disorder categories (Millon, 1986). The prototype model is essentially an alternative method of distinguishing important differences between actually continuous categories by emphasizing clear cases, rather than boundaries between categories. The prototype model is particularly applicable to the categorization of the personality disorders, given the heterogeneity of individual members of any one category, as well as the overlap between individual members diagnosed with other personality disorders. The prototype model does not specify a set of necessary and sufficient category features, but rather specifies the extent to which category features are prototypical of a particular personality disorder.

The incorporation of the prototype model into a diagnostic classification system might enhance the relatively low diagnostic reliability that has generally been associated with the personality disorders compared to Axis I disorders. Highly prototypical cases may be diagnosed with greater reliability relative to less prototypical cases. Under the current system of discrete present-or-absent judgments, however, the effects of prototypicality on diagnostic reliability are not acknowledged. A classification scheme based on a prototype model would explicitly predict a relationship between reliability and prototypicality. Under such a system, low diagnostic reliability would not necessarily be taken to indicate clinician error, since atypical cases would be expected to yield lower agreement than more typical cases. Moreover, two clinicians might agree that a given case is characteristic of two personality disorder categories, but they might differ in their categorical judgments of which of the two is the best single diagnosis for the case. Under the present system of all-or-nothing labels, the clinicians would appear to disagree completely. If prototypicality ratings of the two categories were used, however, considerable agreement would be revealed.

In order to move toward a useful psychodiagnostic system based on a prototype model, it is critical to understand the factors that impact clinicians' diagnostic judgments. The purpose of the present study was to examine how the features associated with particular personality disorders, as well as how the contexts in which those features are presented, impact clinicians' diagnostic judgments. Subjects were presented with brief descriptions of 12 hypothetical clients. Information as to the sex of the client, which might have affected the clinicians' use of particular diagnostic categories, was not provided. The subjects were instructed to quantify their judgments of how prototypical each hypothetical client was of each of the 11 personality disorders specified in the DSM-III. The subjects were also asked to diagnose the hypothetical client with one of the personality disorders. A measure of how prototypical the hypothetical client was of the intended, chosen, or modal diagnosis for that condition was determined in proportion to how prototypical the client was of the other diagnoses. This dependent variable was judged to be particularly compatible with the prototype model.

The descriptions of each of the hypothetical clients were developed from previous experimental data as to the typicality and distinctiveness of features associated with each of the personality disorders. These data were kindly provided to the investigator by Dr. John Livesley (1986a, 1986b) of the University of British Columbia for incorporation into the client descriptions that were used in the present study. These data were used to generate 48 client profiles. Each of the profiles contained features that were associated with the particular personality disorder category which that particular profile was intended by the investigator to portray. These features are referred to as "category features." Each of the profiles also contained features that were associated with other personality disorder categories which that particular profile was not intended to portray. These features are referred to as "noncategory features." The 12 client profiles were designed such that the impact of three factors upon the diagnostic process used by the clinicians could be examined. Each of the three factors represented a dimension along which either the features themselves, or the contexts in which they were presented were varied.

The first experimental factor was feature typicality. The typicality of each of the individual category features within each profile was varied at either a high or a medium level. The levels of high and medium, as opposed to high and low, were selected for two reasons. One was that the noncategory features were of low typicality. Thus, it was necessary to present category features of at least medium typicality in order to obtain a sufficient contrast with the low typicality noncategory features. Additionally, the high versus medium typicality conditions were selected to reduce the contrast between the typicality conditions. This was expected to reduce the likelihood that a highly robust main effect for typicality would override other main effects or any interactions among the independent variables. The second experimental factor was the number of features associated with a given personality disorder in proportion to the total number of features presented, and was varied at either a high or a low level. This proportion is referred to as the <u>dominance</u> of the category features over the total number of features, in which the total number of features consists of both category and noncategory features. Third, the <u>number of the category features</u> included in each profile for the intended personality disorder was varied as either high or low. In addition, as an experimental control, the number of total features in each profile was adjusted such that any potential effects of dominance upon the diagnostic process could be discerned from the influence of total feature number. Moreover, as an additional experimental control, the proportion of category features that were unique to the particular personality disorder category portrayed in each profile over the category features that were shared by other personality disorder categories was kept constant across all conditions. The necessity of this control was highlighted by a previous study demonstrating that feature distinctiveness impacts clinicians' prototypicality judgments (Boykin, 1987). Boykin (1987) found that the factors of category feature typicality and category feature distinctiveness both had strong effects upon clinicians' prototypicality ratings of the personality disorders.

<u>Overview of Hypotheses and Results</u>

On the basis of past research, it was hypothesized that both feature typicality and feature number would have a strong impact upon the diagnostic process. Specifically, personality descriptions containing highly typical category features were hypothesized to be associated with higher measures of diagnostic reliability, diagnostic accuracy, and prototypicality compared with personality descriptions containing relatively low typical category features. Similarly, personality descriptions containing a high number of category features were hypothesized to be associated with higher measures of diagnostic reliability, diagnostic accuracy, and prototypicality relative to personality descriptions containing a low number of category features.

Based upon past research, it was predicted that the variables of feature typicality and feature number would be so powerful that they might mask any effects of dominance. Therefore, a main effect for dominance was predicted only upon the condition that enough variance would be spared by what were hypothesized to be the highly significant main effects for feature typicality and feature number. In this case it was predicted that personality descriptions containing a high proportion of category features to noncategory features would be associated with higher measures of prototypicality, diagnostic accuracy, and diagnostic reliability compared with personality descriptions containing a low proportion of category features.

In general, the experimental results provided support for the above hypotheses, but with a few exceptions. As predicted, a highly significant main effect across all three designs, in the predicted direction, was revealed for category feature typicality. Contrary to predictions, a main effect for category feature number was revealed for only one of the three combinations of the experimental design. A highly significant effect across all three designs, in the predicted direction, was revealed for dominance. As predicted, no main effect for the experimental control of total feature number was revealed. In addition, no significant interactions among any of the factors were revealed.

Category Feature Typicality

The results of the present study for typicality are consistent with those of Boykin (1987) in that the typicality of the features associated with the intended diagnosis strongly affected the clinicians' rating of how prototypical the patient was of the intended diagnosis, as well as the "accuracy" of their diagnostic judgments. These results are also consistent with those of Cantor and Mischel (1980), who found a strong main effect for feature typicality. Note, however, that unlike the present study and that of Boykin

(1987), Cantor and Mischel (1980) did not actually vary the degree of typicality of each individual feature. Rather, they created three levels of typicality by varying the number of features associated with prototypes of disorders agreed upon by 13 clinicians, resulting in an effect due to both typicality and number. In the present study, however, the effects of feature typicality and feature number upon the diagnostic process were assessed independently.

The present study was the first to examine the effect of the typicality of the features associated with the intended diagnosis upon clinicians' judgments of how prototypical the client was of the modal diagnosis for that condition, relative to the nonmodal diagnoses (dep. variable #2). Although a strong main effect was revealed, it was not as powerful as the effect for the intended diagnosis. This was because the modal diagnosis for a given condition did not always correspond to the intended diagnosis. Recall that there were 48 subconditions, corresponding to the four intended diagnoses within each of the 12 experimental conditions. In 12 of the 48 subconditions, no single modal diagnosis was revealed. That is, there was not sufficient consensus in these subconditions for a single modal diagnosis to emerge. For example, of the eight subjects in a subcondition, three might have chosen diagnosis A, three might have chosen diagnosis B, and the remaining two might have chosen diagnoses C and D. Since no single diagnosis received more "votes" than the other diagnoses, no modal diagnosis was obtained for the present purposes. Among the 36 subconditions in which a modal diagnosis did occur, the modal diagnosis was equivalent to the intended diagnosis in 17 conditions.

Although a main effect of category feature typicality was revealed for both the intended and the modal diagnosis, the typicality of the features associated with the intended diagnosis did not affect the clinicians' rating of how prototypical the patient was of the chosen diagnosis, relative to the nonchosen diagnoses (dep. variable #3). This pattern reflects the fact that overall only 37% of the chosen diagnoses were the same as the intended diagnoses. Interestingly, even when the clinicians "misdiagnosed" a profile relative to the intended diagnosis, their prototypicality ratings of the intended diagnosis were still influenced by the typicality of the features associated with the intended diagnosis.

Behaviorally, the typicality of a stimulus may be viewed as the strength of that stimulus to function discriminatively in relation to the response of making a particular diagnosis. Features associated with a given personality disorder category have, in the past, been differentially associated with reinforcing consequences when that diagnosis was made. In a clinician's past, colleagues have praised diagnostic judgments in the presence of certain features, and have withheld approval or even punished the same judgments in the presence of other features. Similarly, the response of making a certain diagnostic judgment in the presence of certain features may have led to more effective intervention relative to when that judgment was made in the presence of other features. In either case, the former features have acquired a discriminative function. Of course, this process is not an either-or phenomenon, as stimuli may vary widely in their degree of association with the behavior-reinforcer contingency. The present results suggest similarities in the learning histories of many practicing clinicians in that certain features have been more consistently associated with diagnostic judgments that were reinforced relative to other features.

The strong effects for typicality in the present study as well as in prior research underscores the importance of incorporating features of prototype models in psychodiagnostic nosologies. Even the existing polythetic personality disorder categories of DSM-III-R, although vastly improved relative to earlier monothetic systems, are based on the assumption of equal importance of the category features. For example, in the case of Borderline Personality Disorder, the DSM-III-R requires that any five of a list of eight symptoms be present. The present results suggest that some symptoms are much more critical to the diagnostic process than others, and this variability must be taken into account if the utility of the diagnostic system is to be maximized.

Dominance of Category Features Over Total Features

Consistent with the results for category feature typicality, dominance had a powerful effect upon clinician's prototypicality judgments of how relatively prototypical the patient was of the intended diagnosis (dep. variable #1), a lesser effect upon judgments of the modal diagnosis (dep. variable #2), and no effect upon judgments of the chosen diagnosis (dep. variable #3). Although this was the first study to examine the independent effect of dominance in a factorial design, the present findings are consistent with those of Cantor (1978). She found that a breadth-dominance-differentiation index of each character profile was highly positively correlated with subjects' prototypicality ratings of that character. Although Cantor's "index" included dominance, it was confounded with other factors (breadth and differentiation). It is therefore impossible to conclude if dominance per se was responsible for the observed effects. The present results are also consistent with those of Rosch and Mervis (1975), who found that the number of noncategory features included in the presentation of category features was highly negatively correlated with clinicians' prototypicality ratings of the overall feature combination.

The highly significant main effect for dominance in the present study supports Livesley's (1985a, 1985b) hypothesis that a negative correlation exists between the number of competing features and measures of prototypicality. The current findings do contradict the findings of some related studies, however (Blashfield, Sprock, Haymaker, & Hodgin, 1989; Horowitz, Post, French, Wallis, & Siegalman, 1981a). Horowitz et al. (1981a) found that a high number of prototypical category features was associated with high diagnostic reliability, but that a high versus low number of "irrelevant" features was not differentially associated with diagnostic reliability. These features, however, were not competing features, but rather, were "irrelevant" features that were not descriptive of other diagnostic categories. Blashfield et al. (1989) found a positive correlation between diagnostic reliability and number of category features, but no correlation between diagnostic reliability and the number of competing features from other personality disorders. This finding, however, may be due in part to the fact that the authors did not control for the typicality and distinctiveness of the individual noncategory features, as well as the dominance of the category features in proportion to the total number of category and noncategory features.

In the present study, the dominance of category features in proportion to the number of total features had a much more critical role in the diagnostic process than did the absolute number of category features or the number of total features. Even when the clinicians "misdiagnosed" a profile relative to the intended diagnosis, their prototypicality ratings of the intended diagnosis were still influenced by the dominance of the intended category features. Such a powerful effect for dominance was unexpected, and raises an interesting theoretical point. The high versus low dominance conditions contained the same amount of information about the category features themselves; both conditions included the same number of category features, and the same level of typicality of the category features. If only the presence of certain category features were critical to the diagnostic process, then no dominance effect would have occurred. The strong dominance effect in the present study suggests that the same information as to the typicality, number, and distinctiveness of the category features can have very different effects upon the diagnostic process as a function of the total context in which the information is portrayed. Behaviorally, category features are viewed as discriminative stimuli in the presence of which certain diagnostic responses have been reinforced in the past. Relative to the high dominance conditions, the low dominance conditions were characterized by additional stimuli that interfered with the controlling relation between key discriminative stimuli and the corresponding diagnostic behavior.

The strong dominance effect in the present study contradicts the generally accepted clinical lore that the more assessment information clinicians gather, the more in touch they are with the most important pathognomonic features of their clients. A clinician's assessment behavior may or may not be in touch with contingencies that have in the past resulted in effective and expedient diagnosis, assessment, and treatment of clients' maladaptive behavior patterns. Of course, detailed assessment for the purpose of monitoring the effectiveness of treatment is likely to be quite useful. Some clinicians, however, may engage in assessment for assessment's sake, and develop intricate historical accounts of their clients' maladaptive behavior patterns. Some of these more subtle behavior patterns are likely to be associated with contrasting diagnostic categories. Although such historical accounts may be quite interesting in capturing the subtleties of a client's behavior, they may encourage both the clinician and client to lose sight of the features of a client's behavior that are the most indicative of procedures leading to effective assessment and treatment. Not only can assessment for assessment's sake reach a point of diminishing returns, it can also mask the importance of highly typical, distinctive, and numerous features that would otherwise be critical to effective diagnosis, assessment, and treatment.

The strong effect for dominance in the present study is consistent with Rosch's probabilistic concept of cue validity, which holds that "the validity of a given cue x as a predictor of a given category y (the conditional probability of y/x) increases as the frequency with which cue x is associated with category y increases and decreases as the frequency with which cue x is associated with categories other than y increases" (Rosch, 1978, pp. 30-31). The dominance effect in the present study attests to the powerful impact that features from competing personality disorders can have upon the diagnostic process. The strong effect for dominance in the present study is also consistent with Cantor and Genero's (1986) suggestion that the diagnostic process involves simultaneously determining what a clinical phenomenon is and what it is not. Thus, the process of diagnosing a personality disorder, for example, must determine not only "what it is," along some gradient of stimulus dimensions, but must also involve a discrimination between "what it is " and "what it is not." In fact, the Greek roots of "diagnosis" (dia - meaning both "through" and "apart," and gno - meaning "to perceive" or "to know") reveal the word as a metaphor for the simultaneous processes of "perceiving through" a phenomenon to the essential core of "what it is" and of "perceiving" the phenomenon "apart" from other phenomena.

Number of Category Features

The fact that the number of category features significantly affected the clinicians' prototypicality judgments for the intended diagnosis (dep. variable 1) in only one of the three combinations of the experimental designs may be due to the fact that this combination was unique in that it kept the exact number of category features within each level constant across the two designs (see Appendix C, Design Combination 2 + 3). That is, there were two category features in the low category number level of both Designs 2 and 3, and there were six category features in the high category number level of both designs. In contrast, for the other two Design Combinations, the exact number of features in the high vs. low levels differed across the two designs that comprised the combination. This difference may have contributed extraneous variance to the category number effect for these two Design Combinations, thereby reducing the likelihood of a significant effect. In any case, the effects for number of category features were clearly weaker than predicted.

The present results and those of Boykin (1987) both contradict the general conclusion in the prototype literature that category feature number is the most critical determinant of the diagnostic process. The present finding of only limited effects of the number of category features is consistent with Boykin's finding that even if category feature number is experimentally controlled, other qualities of the category features, such as typicality and distinctiveness, significantly affect clinicians' prototypicality judgments. As Boykin (1987, p. 80) concluded, "which attributes are in the body of information, not just how many of them there are, is also important in the diagnostic process." Boykin had controlled for category feature number in his study upon the basis of previous research demonstrating the importance of category feature number in determining the outcome of the diagnostic process (specifically, Cantor and Mischel, 1980). A review of this and the other studies on the importance of category feature number (Blashfield et al., 1989; Horowitz et al., 1981a; Rosch & Mervis, 1975), however, reveals that the category features that were used in these studies had already been judged to be of very high typicality. Therefore, rather that simply varying the number of category features, these studies varied the number of highly prototypical category features. This process precluded an analysis of the separate effects of feature number and feature typicality. Yet, the highly significant effect for number of highly prototypical features was attributed, in each of the studies, to category feature number, rather than to feature typicality. An advantage of the present study, like that of Boykin (1987), is that it utilized previous data regarding feature typicality and distinctiveness (Livesley, 1986a, 1986b), thereby enabling an examination of the effects of other qualities of category features independent of those related to category feature number.

The limited impact of category feature number in the present study may be due in part to the strong effects of feature typicality. Each of the personality disorder categories are characterized by one, two, or in some cases three very highly prototypical features. It may be that when these features are present the diagnosis will generally made, and the number of additional features becomes largely irrelevant. This interpretation must be made with caution, however, given the lack of interaction effects in the present study. Relative to the effects of category feature number, the effects of category feature typicality, distinctiveness, and dominance has not been given sufficient attention in the majority of prior investigations. On the basis of the current findings it is concluded that although category feature number can play a significant role in determining the outcome of the diagnostic process, other qualities of the category features themselves, as well as the context of all the features with which they are presented, are at least as important.

Total Number of Features

The inclusion of low versus high total feature number in this study is comparable to some degree to Cantor and Mischel's (1980) contrast between situations in which diagnosticians have a "restricted view" versus a "full view" of a client. Cantor and Mischel (1980) hypothesized that in the "restricted view," when the diagnostician has only a limited amount of information, the typicality of the observed features would be of particular importance. The authors hypothesized that in the "full view," however, when the diagnostician has more information, the number of observed features would be of particular importance. In the present study, the total number of features of the hypothetical client was varied at an either high or low level, corresponding roughly to Cantor and Mischel's notion of restricted vs. full views, respectively. The fact that the total number of features did not significantly affect the clinicians' diagnostic process does not lend support to the hypothesis that feature typicality and feature number are differentially important to the diagnostic process under restricted versus full view conditions.

This factor was included primarily in order to ensure that any effects attributed to dominance would not be due to total feature number. The null effects of total feature number support the conclusion that the effects of dominance described above are in fact due to that variable.

Strengths and Limitations

Many of the strengths of the present study have been discussed above. These include an experimental design permitting the disambiguation of the effects of category feature number, typicality, and dominance. Other strengths include the examination of correspondence of chosen diagnoses with both intended and modal diagnoses, the independent assessment of both diagnostic accuracy and prototypicality ratings, and the examination of prototypicality ratings for each disorder within the context of the prototypicality ratings for other disorders. In addition to examining the context in which the category features were presented, the proportion dependent variable is compatible with the prototype model in that, unlike the previous measure that combines accuracy and prototypicality (Boykin, 1987), it is not based upon the assumption that a feature's prototypicality of one category is on a continuum with its prototypicality of other categories. Moreover, by combining the prototypicality ratings for a given category with those for other categories, the proportional dependent measure takes a greater amount of the variability in the data into account. Additionally, the proportional dependent variable resulted in data that were normally, rather than bimodally distributed, thereby allowing for the use of more powerful parametric statistical analysis as opposed to nonparametric analyses.

The strengths of the study that specifically enhanced its external validity were the use of licensed, doctoral-level, practicing clinicians as subjects. Additionally, clinicians were given an open-choice format for diagnosis within the specified realm of the personality disorders. In comparison to the more restricted format of only four possible diagnoses used in Boykin's (1987) study, however, the present study's more open-ended format increased the variability in the chosen diagnoses and decreased the probability of diagnostic accuracy and reliability.

The weaknesses of the study are primarily related to limitations in external validity. These include the use of paragraph-length descriptions of hypothetical clients, as opposed to videotaped vignettes of hypothetical clients, or preferably edited videotapes of actual clients. Along these lines, the features within the client profiles were not descriptions of actual behaviors, but rather were descriptions of personality traits. Personality traits are themselves functional categories of behavioral patterns. Had they been available, data on the typicality of behaviors (Livesley, 1986a, 1986b) would have been preferable to use in place of the data on the typicality of traits. Additionally, the fact that the study was conducted through the mail, rather than in person, may have decreased to some extent the clinicians' ability to comply with the experimental instructions, such as completing all of the profiles one after another in an uninterrupted time period. This weakness, however, may be outweighed by the fact that the sample of clinicians willing to participate in a study of this nature was possibly more representative of the population of practicing clinicians than would have been a sample of clinicians who had self-selected to participate in a study that required an appointment with the investigator during regular work hours. Note, however, that only 25% of the clinicians who were solicited agreed to participate in the study. It is possible that the data provided by these self-selected subjects may have been somewhat different from the data that would have been provided by a more representative sample that included both clinicians who were interested in participating in such studies, as well as clinicians who were not interested in such studies.

An additional weakness of the study is that in an attempt to include noncategory features from the 10 noncategory personality disorders, and to control for the typicality, number, and distinctiveness of those features, some of the profiles included features that were somewhat contradictory to one another. Relative to the hypothetical clients with more consistent features, the few hypothetical clients with contradictory features may have seemed less representative of actual clients.

Directions for Future Research

The findings of the present study, as well as those of Boykin (1987), attest to the utility of the present methodology for studying factors affecting the diagnostic process. Despite its analogue nature, this methodology allows for the disambiguation of the effects of factors that are otherwise difficult to tease apart. One avenue for future research would be to increase the external validity of the present methodology by using videotapes of actual or dramatized clients, rather than lists of traits. For example, one could utilize previous data on the typicality and distinctiveness of actual behaviors (Livesley, 1986a, 1986b) for each of the personality disorders to develop scripts for videotapes of dramatized clients. It may also be possible to develop videotapes of edited segments with actual clients, although this would be more difficult. The exact content of the videotapes would, of course, have to be carefully checked by raters. In a procedure similar to that employed by Herbert, Nelson, and Herbert (1988), after viewing each videotape, clinicians could also note on a behavior checklist behaviors that they either did or did not observe, with corresponding ratings of how prototypical each behavior was of the chosen diagnosis. The correspondence between the accuracy with which clinicians observed behaviors associated with DSM criteria could also be compared with the clinicians' diagnostic accuracy. This study would also allow for a comparison between Livesley's (1986a, 1986b) data on feature typicality and the clinicians' prototypicality ratings of each of the videotaped behaviors for the chosen diagnosis.

Another interesting avenue for future research within the personality disorders is the issue of cross-situational consistency and temporal stability. Along with the factors in the present study, one could examine the effect of cross-situational consistency upon clinicians' prototypicality ratings. Clinicians could observe either consistent or inconsistent behavior of hypothetical clients in role-play situations. Additionally, as a variation of a study by Cantor (1978), clinicians could observe hypothetical clients engaging in behaviors that were highly typical of particular personality disorders in situations in which such behavior was normative versus situations in which the behavior was not normative. Perhaps of even greater interest, one could cross the factor of normative versus non-normative behavior with a factor of effective versus ineffective behavior. Clinicians could observe videotaped segments of a hypothetical client engaging in behaviors that were highly typical of a given personality disorder in situations in which such behaviors were effective, and also in situations in which the behaviors were not effective. The following factors have now been empirically demonstrated to be important to the diagnostic process: typicality, dominance, distinctiveness, and category feature number. Regardless of methodological details, future studies that investigate any of these factors should sufficiently control for the other factors.

Clinical Implications

The present study holds several important clinical implications, many of which relate to the development and evaluation of future versions of the DSM. First, given the strong effect for feature typicality and the relatively weak effect for feature number, future versions of the DSM need to place greater emphasis on the typicality of the criterion features that a client portrays as opposed to the current emphasis on the number of criterion features that a client portrays. For example, a system in which features were weighted for typicality could be empirically-derived. Research is needed to determine whether or not the increased specificity and reliability that might result from such a modification would outweigh the increased cumbersomeness that the system would entail.

The fact that clinicians' prototypicality ratings were significantly affected by how typical the category features were of the intended diagnosis, even in cases when diagnostic accuracy was low, suggests that future versions of the DSM would do well to encourage clinicians to attend more to the prototypicality ratings of the personality disorders, rather than the current categorical yes-no diagnostic format. Moreover, the main effect for dominance suggests that clinicians need to be careful not to carry out their assessments in so much detail that they end up attending to too many low typical noncategory, or irrelevant features, thereby decreasing the salience of the key category features.

Research utilizing videotaped profiles could lead to an empirically-derived training procedure in which individual clinicians could compare their prototypicality ratings and diagnoses for videotaped client dramatizations with those of clinicians known in the field for their diagnostic expertise with particular personality disorders. Computerized training programs could be devised to accompany videotapes of dramatized clients who had previously been judged to be more or less prototypical of specific personality disorders. The training programs could allow the clinician to pause the videotape in order to rate the prototypicality of key behaviors for the various personality disorders, and to compare these ratings to those of recognized experts in the field. Such programs and videotapes could be included as part of a training kit to accompany future versions of the DSM.

Clearly, much work remains to be done in this relatively new area. In addition to increasing the utility of our psychodiagnostic classification systems, such work will hopefully enhance our understanding of the process of verbal categorization in general.

BIBLIOGRAPHY

Allport, G. (1937). Personality: A Psychological Interpretation. New York: Henry Holt and Company.

American Psychiatric Association (1968). Diagnostic and statistical manual of mental disorders (2nd ed.). Washington, DC: Author.

American Psychiatric Association (1980). Diagnostic and statistical manual of mental disorders (3rd ed.). Washington, DC: Author.

American Psychiatric Association (1987). Diagnostic and statistical manual of mental disorders (3rd ed., revised). Washington, DC: Author.

Baissier de Sauvage, F. (1790) Nosologia methodica. In C. F. Daniel, (Trans.) Prolegomena, article 74, Leipzig: Schwicker.

Beach, L. R. (1964a). Cue probabilism and inference behavior. Psychological Monographs, 78, No. 582.

Beach, L. R. (1964b). Recognition, assimilation, and identification of objects. Psychological Monographs, 78, No. 583.

Blashfield, R., Sprock, J., Pinkson, K., & Hodgin, J. (1985). Exemplar prototypes of personality disorder diagnoses. Comprehensive Psychiatry, 26, 11-21.

Blum, J. D. (1978). On changes in psychiatric diagnosis over time. American Psychologist, 33, 1017-1031.

Bower, G. H., & Hilgard, E. R. (1981). Theories of learning (5th ed.). Englewood Cliffs, NJ: Prentice Hall.

Boykin, R. A. (1987). Prototypes and Diagnoses: Effects of Attribute Centrality and Attribute Distinctiveness. Unpublished doctoral dissertation, University of North Carolina at Greensboro, Greensboro, NC.

Cantor, N. (1978). Prototypicality and personality judgments. Unpublished doctoral dissertation. Dissertation Abstracts International, 4645-B. (University Microfilms No. 7905826, 1-267).

Cantor, N. (1986). Psychiatric diagnosis and natural categorization. In T. Millon, & G. L. Klerman (Eds.), Contemporary directions in psychopathology: Toward the DSM-IV. New York: Guilford Press.

Cantor, N., & Genero, N. (1986). Psychiatric diagnosis and natural categorization: A close analogy. In T. Millon, & G. L. Klerman (Eds.), Contemporary directions in psychopathology: Toward the DSM-IV (pp. 233-256). New York: Guilford Press.

Cantor, N., & Mischel, W. (1977). Traits as prototypes: Effects on recognition memory. Journal of Personality and Social Psychology, 35, 38-48.

Cantor, N., & Mischel, W. (1979a). Prototypes in person perception. In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol. 12). New York: Academic Press.

Cantor, N., & Mischel, W. (1979b). Prototypicality and personality: Effects on free recall and personality impressions. Journal of Research in Personality, 13, 187-205.

Cantor, N., Mischel, W., & Schwartz, J. C. (1982). A prototype analysis of psychological situations. Cognitive Psychology, 14, 45-77.

Cantor, N., Smith, E. E., French, R., & Mezzich, J. (1980). Psychiatric diagnosis as prototype categorization. Journal of Abnormal Psychology, 89, 181-193.

Catania, A. C. (1984). Learning (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Clarkin, J. F., Widiger, T. A., Frances, A., Hurt, S. W., & Gilmore, M. (1983). Prototypic typology and the borderline personality disorder. Journal of Abnormal Psychology, 92, 263-275.

Curtis, H. (1979). Biology. New York: Worth Publishers.

Dewhurst, K. (1966, Ed.). Dr. Thomas Sydenham (1624-1689); his life and original writings. Berkeley, CA: University of California Press.

Eysenck, H. (1952). The scientific study of personality. London: Routledge & Kegan Paul.

Genero, N., & Cantor, N. (1987). Exemplar prototypes and clinical diagnosis: Towards a cognitive economy. Journal of Social and Clinical Psychology, 5, 59-78.

Heider, E. R. (1972). Universals in color naming and memory. Journal of Experimental Psychology, 93, 10-20.

Herbert, D. L., Nelson, R. O., Herbert, J. D. (1988). Effects of psychodiagnostic labels, depression severity, and instructions on assessment. Professional Psychology: Research and Practice, 19, 496-502.

Herrnstein, R. J. (1982). Stimuli and the texture of experience. Neuroscience and Biobehavioral Reviews, 6, 105-117.

Herrnstein, R. J., & Loveland, D. H. (1964). Complex visual concept in the pigeon. Science, 146, 549-551.

Horowitz, L. M., Post, D. L., French, R., Wallis, K. D., & Siegelman, E. Y. (1981a). The prototype as a construct in abnormal psychology: 2. Clarifying disagreement in psychiatric judgments. Journal of Abnormal Psychology, 90, 575-585.

Horowitz, L. M., Wright, J. C., Lowenstein, E., & Parad, H. W. (1981b). The prototype as a construct in abnormal psychology: 1. A method for deriving prototypes. Journal of Abnormal Psychology, 90, 568-574.

Jenkins, H. M., & Harrison, R. H. (1960). Effect of discrimination training on auditory generalization. Journal of Experimental Psychology, 59, 246-253.

Keller, F. S., & Schoenfeld, W. N. (1950). Principles of psychology. New York: Appleton-Century-Crofts.

King, G. (1954). Research with neuropsychiatric samples. Journal of Psychology, 614, 462-472.

Lakoff, G. (1987). Women, fire, and dangerous things: What categories reveal about the mind. Chicago: University of Chicago Press.

Lea, S. E. G., & Harrison, R. H. (1978). Discrimination of polymorphous stimulus sets by pigeons. Quarterly Journal of Experimental Psychology, 30, 521-537.

Livesley, W. J. (1985a). The classification of personality disorder: 1. The choice of category concept. Canadian Journal of Psychiatry, 30, 353-358.

Livesley, W. J. (1985b). The classification of personality disorder: 2. The problem of diagnostic criteria. Canadian Journal of Psychiatry, 30, 359-362.

Livesley, W. J. (1986). Trait and behavioral prototypes of personality disorders. American Journal of Psychiatry, 143, 728-732.

Malt, B. C., & Smith, E. E. (1984). Correlated properties in natural categories. Journal of Verbal Learning and Verbal Behavior, 28, 250-269.

Mellsop, G., Varghese, F., Joshua, S. (1982). The reliability of Axis II of DSM-III. American Journal of Psychiatry, 139, 1360-1361.

Mervis, C., & Rosch, E. (1981). Categorization of natural objects. Annual Review of Psychology, 32, 89-115.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits in our capacity for processing information. Psychological Review, 63, 81-97.

Millon, T. (1986). Personality prototypes and their diagnostic criteria. In T. Millon, & G. L. Klerman (Eds.), Contemporary directions is psychopathology: Toward the DSM-IV (pp. 671-712). New York: Guilford Press.

Morgan, M. J., Fitch, M. D., Holman, J. G., & Lea, S. E. G. (1976). Pigeons learn the concept of an "A". Perception, 5, 57-66.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. Journal of Experimental Psychology, 83, 304-308.

Reed, S. K. (1972). Pattern recognition and categorization. Cognitive Psychology, 3, 382-407.

Roback, A. (1927). The Psychology of Character. New York: Harcourt, Brace, and Company.

Rosch, E. (1973a). Natural categories. Cognitive Psychology, 4, 328-350.

Rosch, E. (1973b). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.). Cognitive development and the acquisition of language. New York: Academic Press.

Rosch, E. (1974). Linguistic relativity. In A. Silverstein (Ed.). Human communication: Theoretical perspectives. New York: Halsted Press.

Rosch, E. (1975a). Cognitive representations of semantic categories. Journal of Experimental Psychology: General, 104, 192-233.

Rosch, E. (1975b). The nature of mental codes for color categories. Journal of Experimental Psychology: Human Perception and Performance, 1, 303-322.

Rosch, E. (1978). Principles of categorization. In E. Rosch, & B. Lloyd (Eds.). Cognition and categorization (pp.28-49). Hillsdale, NJ: Lawrence Erlbaum.

Rosch, E., & Mervis, C. (1975). Family resemblances: Studies in the internal structure of categories. Cognitive Psychology, 7, 573-605.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. Cognitive Psychology, 15, 346-378.

Rosch, E., Simpson, C., & Miller, R. S. (1976). Structural bases of typicality effects. Journal of Experimental Psychology: Human Perception and Performance, 2, 491-502.

Rotter, J. (1954). Social learning and clinical psychology. New York: Prentice-Hall.

Schutte, N. S., Kenrick, D. T., & Sadalla, E. K. (1985). The search for predictable settings: Situational prototypes, constraint, and behavioral variation. Journal of Personality and Social Psychology, 49, 121-128.

Schwartz, M. A., and Wiggins, O. P. (1987). Diagnosis and ideal types: A contribution to psychiatric classification. Comprehensive Psychiatry, 28, 277-291.

Scott, W. (1958). Research definitions of mental health and mental illness. Psychological Bulletin, 55, 1-45.

Skinner, B. F. (1957). Verbal behavior. New York: Appleton-Century-Crofts.

Skinner, B. F. (1969). Contingencies of reinforcement: A theoretical analysis. Englewood Cliffs, NJ: Prentice-Hall.

Smith, E. E., & Medin, D. L. (1981). Categories and concepts. Cambridge, MA: Harvard University Press.

Spitzer, R. L., Skodol, A. E., & Gibbon, M. (1981). DSM-III casebook. Washington, D.C.: American Psychiatric Association.

Spitzer, R. L. (1976). More on pseudoscience in science and the case for psychiatric diagnosis. Archives of General Psychiatry, 33, 459-470.

Temkin, O. (1965). The history of classification in the medical sciences. In M. M. Katz, J. O. Cole, and W. E. Barton (Eds.). The role and methodology of classification in psychiatry and psychopathology. Washington D.C.: American Psychiatric Association.

Tulving, E., & Donaldson, W. (Eds.). (1972). Organization of memory. New York: Academic Press.

Tversky, A. (1977). Features of similarity. Psychological Review, 84, 327-352.

Tversky, A., & Gati, I. (1978). Studies in similarity. In E. Rosch, & B. Lloyd (Eds.). Cognition and categorization (pp.81-98). Hillsdale, NJ: Lawrence Erlbaum.

Tversky, B., & Hemenway, K. (1984). Objects, parts, and categories. Journal of Experimental Psychology: General, 113, 169-193.

United States Department of Health and Human Services (1980). The international classification of diseases (9th ed.). Washington, D.C.: Author.

Webb, L. J., DiClemente, C. C., Johnston, E. E., Sanders, J. L., & Perley, R. A. (eds.). (1981). DSM-III training guide. New York: Brunner/Mazel.

Wittgenstein, L. (1953). Philosophical investigations. New York: MacMillian.

Yarkin, K. L., Harvey, J. H., & Bloxum, B. M. (1981). Cognitive sets, attribution, and social interaction. Journal of Personality and Social Psychology, 41, 243-252.

APPENDIX A

Materials Mailed to Subjects

Solicitation Letter

Dear

I am a doctoral candidate in the clinical psychology program at UNCG. I am conducting a dissertation research project that depends upon the participation of practicing doctoral level clinical and counseling psychologists. The study is not a test of clinicians' knowledge. Rather, it is a survey of clinicians' own impressions of the personality disorders, based upon their experience with actual clients. The survey is designed to take no longer than one hour. Moreover, because the materials can be mailed to each participating clinician, the study can be completed at home, rather than during regular office hours. This study has been fully reviewed and approved by both my departmental dissertation committee and the Human Subjects Review Committee at UNCG. It has been judged to satisfy the American Psychological Association's ethical guidelines, and there is no misinformation or discomfort involved. Each individual participant's responses will be kept strictly confidential.

I am contacting you to ask if you would be willing to participate in this study. We are seeking participating clinicians who have been permanently licensed for at least 3 years, who work primarily with adults, and whose current clinical activities include the diagnosis of clients with personality disorders. Enclosed is a self-addressed, stamped postcard for your reply. If you agree to participate, you will receive a packet of information containing a paragraph-length personality profile for each of 12 hypothetical clients. You will be asked to provide your impressions of each client's diagnostic picture, based upon your experience with actual clients.

If you have any questions about the study, please do not hesitate to contact me or my dissertation chairperson, Rosemery O. Nelson, Ph.D., at the UNCG Psychology Department, (919) 334-5013. We believe that by tapping the clinical experience of practicing psychologists, this study will have important implications for an improved understanding of the personality disorders, and the variables that lead to their diagnosis. We hope to publish the study's findings and implications, while maintaining the confidentiality of individual data.

Recalling your own years as a graduate student, I am sure you know how grateful I would be if you would agree to participate in this study. Thank you for your consideration.

Sincerely,

Diana L. Herbert, M.A.
Doctoral Candidate

Rosemery O. Nelson, Ph.D.
Dissertation Chairperson

enc.

Response Form

Please take a moment to indicate your reply:

NAME:____

( ) YES, I would like to participate in your study. I understand that I will receive the materials for the study by mail, and that the survey is designed to take about 45 minutes.

( ) I have been permanently licensed for at least 3 years.

( ) I work primarily with adult clients.

( ) My current clinical activities include the diagnosis of personality disorders.

( ) SORRY, but I will not be able to participate in your study.

Follow Up Letter to Participants

Dear

I am very grateful that you have agreed to participate in my dissertation research. This study is designed to take approximately 1 hour, and may be completed at home, rather than during your regular office hours. This study has been approved by my dissertation committee and by our department's Human Subjects Review Committee. It involves no misinformation, risk, or discomfort. If you have any further questions concerning the study, please do not hesitate to call me or my dissertation chairperson, Dr. Rosemery O. Nelson, at the UNCG Psychology Department, (919) 334-5013.

So that your individual responses can be kept strictly confidential, they will be identified only by a code number. This code number is already written on each of your data sheets. Only the primary investigator, Diana Herbert, has access to the list matching subject code numbers and names. This list will only be used to note whether or not each participant's data have been received. To ensure the confidentiality of your data, please do not write your name on the actual data sheets.

Before beginning the study, please read, sign, and date the enclosed consent form. Please try to complete the entire survey uninterrupted, and within one week of receiving this packet. Once all the participants have returned their responses, you will receive a debriefing statement, explaining the exact nature of this study. A summary of the general experimental results will also be mailed to you as soon as they are available. Thank you for your support in this project.

Sincerely,

Diana Herbert, M.A.
Doctoral Candidate

Rosemery O. Nelson, Ph.D.
Dissertation Chairperson

enc.

Consent Form

I agree to participate in the present investigation on psychological assessment with the understanding that I will be free to terminate my participation at any time. I understand that the information I provide in this study will be assigned an anonymous subject identification number and will be treated as confidential material. I have been informed as to the nature of the experimental procedures. I understand that I will be assigning diagnoses and prototypicality ratings to a series of case profiles. I understand that the present investigation is in no way meant to represent an evaluation of my diagnostic skills, but is instead a survey of practicing clinicians' clinical impressions. I understand that I will be fully debriefed as to the details of the study as soon as I mail the enclosed materials back to the principal investigator.

Signature:______________________________

Date:____________________________________

Instruction Sheet

(Please read this sheet before you begin the survey.)

General Instructions:

1. On the pages that follow, you will be asked to read a paragraph-length personality profile for each of 8 hypothetical clients. Note that all of the clients are adults. The features that are decribed in the profiles for each client are inflexible and maladaptive. These features interfere significantly with the client's social or occupational functioning and cause subjective distress. They are representative of the client's current and long-term functioning, and they are not limited to episodes of illness.

2. All too often the descriptions of disorders in diagnostic systems are not based upon empirical data from the valuable resource of practicing clinicians' experience. This study is designed to tap the clinical experience of "on-line" clinicians like yourself, independent of the information in existing diagnostic systems. Therefore, please remember that this study is specifically not a test of how your knowledge corresponds to diagnostic manuals or other sources. Rather, it is a survey of your own current impressions and preferences, based upon your clinical experience with actual clients. We are interested in your personal impressions of how typical or atypical the hypothetical clients are of the various personality disorders. Thus, we ask that you not consult any outside sources, such as diagnostic manuals, books, or colleagues in completing this survey, as this would defeat the survey's purpose.

3. Because we are interested in your spontaneous reactions to the cases, please try not to spend more than 3 to 4 minutes on any one case.

Specific Instructions for Cases:

1. On the next page you will see the personality profile of your first hypothetical client. Please read the entire profile, and then try your best to imagine this person and what they are like.
2. Following the first profile is a list of the 11 personality disorders. Next to each disorder is a 1-7 rating scale. Please circle only one number between 1 and 7 corresponding to how typical or atypical the client is of EACH

of the 11 personality disorders. The higher the number, the more typical the client is of the disorder; the lower the number, the less typical the client is of the disorder. (Note that in any category, such as the category of birds, for example, some category members, such as robins, sparrows, and cardinals, are very good "classic" examples of that category. Other category members, however, such as ostriches, penguins, and turkeys, are considered more "atypical" examples of the bird category).

3. After you have completed the first client's typicality ratings for ALL 11 personality disorders, please decide which single category best fits the client described in the profile. Indicate this by putting a check mark in the blank next to the category you have selected. Feel free to use the optional comments section to clarify any responses that you think might be misunderstood.

4. After you have completed these steps for the first case, continue on to the remaining 7 cases. Please complete each of the cases one at a time, in order, and please refrain from returning to a case after you have begun the next one. Thank you again for participating. Please begin the survey.

Sample Personality Profile

1. PLEASE READ THE FOLLOWING PROFILE AND THINK ABOUT THE INDIVIDUAL WHO IS DESCRIBED. THIS INDIVIDUAL CAN BE DESCRIBED IN THE FOLLOWING WAY:

- -- reacts intensely to separation from others
- -- unstable interpersonal relationships
- -- loss of appreciation for total context due to preoccupation with trivia
- -- intense, irrational, inappropriate anger
- -- frequently overwhelmed by intense affect, either hostility or depression
- -- unable to experience pleasure; anhedonic
- -- shows impaired reality testing under stress
- -- feelings of depersonalization and derealization; sees self as artificial
- -- conflicting emotions of love, anger, and guilt felt towards those upon whom he/she depends

2. For EACH of the categories listed below, please circle a number between 1 and 7 to indicate how typical the individual is of EACH CATEGORY.

3. After step 2, please decide which single category best describes the individual. Put a checkmark (✓) in the blank next to the category you have selected. Use the optional comments section to clarify any responses you think might be misunderstood.

CATEGORY	TYPICALITY RATING (how well the person described fits each category; 7 = best fit or "classic" example, 1 = poorest fit or most atypical example)
___ Antisocial P.D.	1 -- 2 -- 3 -- 4 -- 5 -- 6 -- 7
___ Avoidant P.D.	1 -- 2 -- 3 -- 4 -- 5 -- 6 -- 7
___ Borderline P.D.	1 -- 2 -- 3 -- 4 -- 5 -- 6 -- 7
___ Dependent P.D.	1 -- 2 -- 3 -- 4 -- 5 -- 6 -- 7
___ Histrionic P.D.	1 -- 2 -- 3 -- 4 -- 5 -- 6 -- 7
___ Narcissistic P.D.	1 -- 2 -- 3 -- 4 -- 5 -- 6 -- 7
___ Obsessive-Compulsive P.D.	1 -- 2 -- 3 -- 4 -- 5 -- 6 -- 7
___ Paranoid P.D.	1 -- 2 -- 3 -- 4 -- 5 -- 6 -- 7
___ Passive-Aggressive P.D.	1 -- 2 -- 3 -- 4 -- 5 -- 6 -- 7
___ Schizoid P.D.	1 -- 2 -- 3 -- 4 -- 5 -- 6 -- 7
___ Schizotypal P.D.	1 -- 2 -- 3 -- 4 -- 5 -- 6 -- 7
___ Other (specify:___________)	1 -- 2 -- 3 -- 4 -- 5 -- 6 -- 7

4. Comments (optional):

WHEN YOU HAVE COMPLETED YOUR CHOICES FOR THIS PROFILE, PLEASE TURN THE PAGE.
222121

This profile corresponds to the cell in Design 3, in which the category features were of high typicality, high dominance, and high number (see Table 13). The profile was intended to portray the diagnostic category of Borderline Personality Disorder. Seven of the eight clinicians who received this profile diagnosed it as Borderline Personality Disorder. For each profile corresponding to each of the 48 profiles, there were four versions in which the presentation of the features was randomly ordered to control for sequence effects. Each of the four versions was received by two of the eight clinicians in each subcondition.

Demographic Questionnaire

1. Age: ____

2. Gender: ____

3. Primary place of clinical work (please check one):

   ____ Private practice

   ____ Hospital

   ____ Mental health clinic

   ____ Correctional facility

   ____ Medical school

   ____ Other, please specify: ____________________

4. Number of years of clinical experience: ____

5. Approximate number of personality disorder cases you assessed in the last 6 months: ____

6. (Optional) - Any specific comments or feedback about this this study are welcome.

Thank you again for your participation. Please be sure to enclose this questionnaire and all 8 completed profiles when you return your data in the enclosed envelope.

Debriefing Letter

Dear

We would like to thank you for your recent participation in my dissertation project, in which you provided diagnoses and typicality ratings for several hypothetical clients. This letter is intended to explain the nature of the study in greater detail.

Clinical diagnoses of psychopathology have often been viewed as a necessary evil. On the one hand, diagnostic categories are essential in communicating information about clients to other clinicians, as well as to oneself at a future date. Diagnoses may also provide clues to clinicians regarding additional features to inquire about during assessment, as well as clues about etiology and treatment. On the other hand, many clinicians feel that diagnostic categories are too narrow to reflect the diversity of clinical phenomena in the real world. A common criticism that clinicians have of diagnostic systems is that they invariably pigeon-hole individual clients into categories that do not completely fit.

DSM-III-R represents a move toward a system that acknowledges the heterogeneity of clients diagnosed with a given disorder. This classification system, however, is still based upon the traditional classical theory of categorization. According to the classical theory, category membership is determined by the presence or absence of a set of necessary and sufficient criteria. Thus, an entire set of criteria is necessary for category membership in that no criteria can be missing. The set of criteria is also sufficient to acquire category membership in that no additional criteria need be fulfilled. If all members of a classical category possess the same set of necessary and sufficient criteria, it follows that all category members are equally good, representative instances of the category. Moreover, all category members are equally poor, unrepresentative examples of other categories.

Classical theory has recently been challenged by proponents of the alternative prototype theory of categorization. According to prototype theory, category membership is not determined by a set of necessary and sufficient criteria. Rather, it is determined by a category member's degree of "family resemblance" to the category. In the case of Histrionic Personality Disorder, for example, one client may be a "classic" example of the disorder, whereas another client may be an atypical example of the disorder. Thus, according to prototype theory, all clients diagnosed with a particular disorder are <u>not</u> equally representative of that disorder. Some are more prototypical than others.

We are interested in examining how three critical factors interact to determine the prototypicality of individual cases of the personality disorders. The more prototypical a client is of a given disorder, the more likely he or she will be diagnosed with that disorder. The first factor is the <u>absolute number</u> of features associated with a given diagnostic category described in the case. The second factor is the <u>proportion</u> of "category" features over the total number of features described in the case. The third factor is the <u>prototypicality</u> of the individual category features described in the case. We varied these factors in a two (medium typicality vs. high typicality) by two (low dominance vs. high dominance) by two (low number vs. high number) factorial design. We predicted that the interaction between feature number, feature typicality, and dominance would impact diagnostic judgments according to the following principle. When the salience of one variable (or pair of variables) is <u>low</u>, then the remaining variable has a <u>greater</u> impact relative to when the salience of the first variable (or pair of variables) is <u>high</u>. For example, when the <u>number</u> of category features is <u>low</u>, as opposed to high, then <u>dominance</u> is predicted to have a <u>greater</u> impact upon the clinicians' judgments.

The results of this study may have important implications for identifying the processes that clinicians use in making diagnostic judgments. The study may help us understand how clinicians diagnose clients with personality disorders according to factors that appear to "call up" the clinicians' prototypes of these disorders. These decision-making processes could then be incorporated into the development and evaluation of future psychodiagnostic systems.

Thank you again for contributing your clinical experience to our understanding of the personality disorders and how they are diagnosed. A summary of the results of the project will be mailed to you as soon as they are available.

Sincerely,

Diana L. Herbert, M.A.
Doctoral Candidate

Rosemery O. Nelson, Ph.D.
Dissertation Chairperson

APPENDIX B

Tables

Table 1

Subject Demographic Data

Gender	50% male (n = 16) 50% female (n = 16)
Age	$\bar{x}$ = 41.5 years S.D. = 8.0 years min = 31 years max = 69 years (n = 30)
Primary Place of Practice	91% private practice 9% hospital (n = 32)
Years of Clinical Experience	$\bar{x}$ = 13.6 years S.D. = 7.6 years min = 4 years max = 34 years (n = 30)
No. of Personality Disorder Cases Assessed in Past Six Months	$\bar{x}$ = 14.4 cases S.D. = 12.5 cases min = 2 cases max = 50 cases (n = 27)

Table 2

Summary List of the Five Major Dependent Variables

Dependent Variable 1:

$$\frac{\text{typicality rating for the intended category dx}}{\text{mean of the typicality ratings for the noncategory dx's}}$$

Dependent Variable 2:

$$\frac{\text{typicality rating for the modal dx}}{\text{mean of the typicality ratings for the nonmodal dx's}}$$

Dependent Variable 3:

$$\frac{\text{typicality rating for the chosen dx}}{\text{mean of the typicality ratings for the nonchosen dx's}}$$

Dependent Variable 4:

Number of chosen diagnoses corresponding to the intended diagnosis in a given condition.

Dependent Variable 5:

Number of chosen diagnoses corresponding to the modal diagnosis in a given condition.

KEY: dx = diagnosis

Table 3

Summary Chart of Experimental Results

Measure	Factor	Design Combination 1 + 3	1 + 2	2 + 3
DV1	Typ	****	****	****
	Dom	****	.	****
	Cat#	-	-	*
	Tot#	.	-	.
DV2	Typ	**	*	**
	Dom	*	.	***
	Cat#	-	-	-
	Tot#	.	-	.
DV3	Typ	-	-	-
	Dom	-	.	-
	Cat#	-	-	-
	Tot#	.	-	.
DV4	Typ	**	**	**
	Dom	**	.	**
	Cat#	-	-	*
	Tot#	.	-	.
DV5	Typ	*	-	*
	Dom	-	.	*
	Cat#	-	-	-
	Tot#	.	-	.

KEY: **** = p < or = .0001
*** = p < or = .001
** = p < or = .01
* = p < or = .05
- = not significant
. = not applicable

Typ = Typicality
Dom = Dominance
Cat# = Category number
Tot# = Total number
For Dv1 to DV5, refer to Table 2.

Table 4

Three-Way Analysis of Variance on Dependent Variable 1

$$\frac{\text{typicality rating for the intended category dx}}{\text{mean of the typicality ratings for the noncategory dx's}}$$

For Design Combination 1 + 3

Source	Sums of Squares	d.f.	F	p
Typicality	25.51	1	31.52	0.0001
Dominance	10.63	1	13.18	0.0001
Category Number	0.58	1	0.72	0.097
Typ * Dom	1.87	1	2.31	0.130
Typ * Cat#	0.12	1	0.15	0.701
Dom * Cat#	1.76	1	2.17	0.142
Typ * Dom * Cat#	0.80	1	0.99	0.321
Error	179.71	222		

Table 5

Three-Way Analysis of Variance on Dependent Variable 1

$$\frac{\text{typicality rating for the intended category dx}}{\text{mean of the typicality ratings for the noncategory dx's}}$$

For Design Combination 1 + 2

Source	Sums of Squares	d.f.	F	p
Typicality	16.92	1	26.20	0.0001
Category Number	0.52	1	0.80	0.37
Total Number	0.78	1	1.20	0.27
Typ * Cat#	2.31	1	3.58	0.060
Typ * Tot#	0.19	1	0.30	0.587
Cat# * Tot#	1.66	1	2.58	0.110
Typ * Cat# * Tot#	0.10	1	0.15	0.702
Error	140.74	218		

Table 6

Three-Way Analysis of Variance on Dependent Variable 1

$$\frac{\text{typicality rating for the intended category dx}}{\text{mean of the typicality ratings for the noncategory dx's}}$$

For Design Combination 2 + 3

Source	Sums of Squares	d.f.	F	p
Typicality	30.90	1	39.71	0.0001
Dominance	16.63	1	21.95	0.0001
Category Number	4.26	1	5.62	0.019
Typ * Dom	0.80	1	1.06	0.304
Typ * Cat#	0.38	1	0.50	0.479
Dom * Cat#	0.001	1	0.001	0.977
Typ * Dom * Cat#	1.42	1	1.87	0.173
Error	163.66	216		

Table 7

Three-Way Analysis of Variance on Dependent Variable 2

$$\frac{\text{typicality rating for the modal dx}}{\text{mean of the typicality ratings for the nonmodal dx's}}$$

For Design Combination 1 + 3

Source	Sums of Squares	d.f.	F	p
Typicality	9.13	1	9.99	0.002
Dominance	4.61	1	5.04	0.026
Category Number	1.09	1	1.19	0.276
Typ * Dom	1.55	1	1.69	0.196
Typ * Cat#	0.23	1	0.25	0.618
Dom * Cat#	0.53	1	0.58	0.448
Typ * Dom * Cat#	1.05	1	1.15	0.285
Error	122.53	134		

Table 8

Three-Way Analysis of Variance on Dependent Variable 2

$$\frac{\text{typicality rating for the modal dx}}{\text{mean of the typicality ratings for the nonmodal dx's}}$$

For Design Combination 1 + 2

Source	Sums of Squares	d.f.	F	p
Typicality	2.71	1	3.92	0.050
Category Number	0.395	1	0.571	0.451
Total Number	0.002	1	0.003	0.955
Typ * Cat#	0.69	1	1.00	0.319
Typ * Tot#	0.12	1	0.18	0.674
Cat# * Tot#	1.07	1	1.55	0.216
Typ * Cat# * Tot#	0.005	1	0.01	0.935
Error	93.93	136		

Table 9

Three-Way Analysis of Variance on Dependent Variable 2

$$\frac{\text{typicality rating for the modal dx}}{\text{mean of the typicality ratings for the nonmodal dx's}}$$

For Design Combination 2 + 3

Source	Sums of Squares	d.f.	F	p
Typicality	8.68	1	9.85	0.002
Dominance	10.88	1	12.34	0.001
Category Number	0.02	1	0.24	0.877
Typ * Dom	2.13	1	2.42	0.122
Typ * Cat#	0.04	1	0.04	0.837
Dom * Cat#	2.68	1	3.04	0.083
Typ * Dom * Cat#	1.51	1	1.71	0.193
Error	135.72	154		

Table 10

Three-Way Analysis of Variance on Dependent Variable 3

$\frac{\text{typicality rating for the chosen dx}}{\text{mean of the typicality ratings for the nonchosen dx's}}$

For Design Combination 1 + 3

Source	Sums of Squares	d.f.	F	p
Typicality	1.98	1	2.06	0.152
Dominance	1.66	1	1.73	0.190
Category Number	0.39	1	0.413	0.521
Typ * Dom	0.31	1	0.33	0.568
Typ * Cat#	0.01	1	0.01	0.928
Dom * Cat#	0.67	1	0.70	0.405
Typ * Dom * Cat#	0.09	1	0.10	0.757
Error	206.99	216		

Table 11

Three-Way Analysis of Variance on Dependent Variable 3

$$\frac{\text{typicality rating for the chosen dx}}{\text{mean of the typicality ratings for the nonchosen dx's}}$$

For Design Combination 1 + 2

Source	Sums of Squares	d.f.	F	p
Typicality	0.44	1	0.52	0.474
Category Number	0.004	1	0.004	0.948
Total Number	0.12	1	0.13	0.715
Typ * Cat#	0.40	1	0.47	0.493
Typ * Tot#	1.81	1	2.12	0.147
Cat# * Tot#	0.02	1	0.02	0.891
Typ * Cat# * Tot#	0.20	1	0.24	0.628
Error	180.79	211		

Table 12

Three-Way Analysis of Variance on Dependent Variable 3

$$\frac{\text{typicality rating for the chosen dx}}{\text{mean of the typicality ratings for the nonchosen dx's}}$$

For Design Combination 2 + 3

Source	Sums of Squares	d.f.	F	p
Typicality	0.005	1	0.005	0.945
Dominance	0.78	1	0.80	0.372
Category Number	0.57	1	0.58	0.446
Typ * Dom	0.61	1	0.62	0.431
Typ * Cat#	0.11	1	0.11	0.736
Dom * Cat#	0.45	1	0.47	0.495
Typ * Dom * Cat#	0.55	1	0.56	0.453
Error	201.84	207		

Table 13

Number of Chosen Diagnoses Matching the Intended and Modal Diagnosis for Each Subcondition

	LO DOM LO TOTAL#		LO DOM HI TOTAL#		HI DOM LO TOTAL#	
	LO CAT#	HI CAT#	LO CAT#	HI CAT#	LO CAT#	HI CAT#
MED TYP	A0-OC4 B0-SD3 <u>H4==H4</u> N0-OC2$	A1---- B0---- H1--N3# N1-PR3	A0---- B0-DY3 H2-SD4 N0-PR3	A0-PR3 B1-OC4 H0--N2# N2==N2#	A2---- B0-DY2# H1--N2 N2----	<u>A5==A5</u> B1-DP6 H3==H3 N1----
HI TYP	A0-OC4 B2---- <u>H5==H5</u> N2----	A2---- B2---- H3---- <u>N4==N4</u>	A1-SD4 B1---- H2--B3 <u>N5==N5</u>	A2----# <u>B4==B4</u> <u>H4==H4</u> N3==N3	<u>A7==A7</u> B1-PS2 H3==H3 <u>N6==N6</u>#	<u>A6==A6</u> <u>B7==B7</u> <u>H6==H6</u> <u>N5==N5</u>
	DESIGN 1		DESIGN 2		DESIGN 3	

Each of the four subconditions within each cell presents the following information in the following order: the intended diagnosis, the number of chosen diagnoses that matched the intended diagnosis, the modal diagnosis if one occurred, and the number of chosen diagnoses that matched the modal diagnosis. There was a maximum of eight possible matching chosen diagnoses in each subcondition. Due to missing data, however, six of the 48 subconditions had only seven valid cases (indicated by a "#"), and one subconditon had only six valid cases (indicated by a "$"). The 13 <u>underlined</u> subconditions are those in which the chosen diagnoses (a) matched the intended diagnosis, and (b) were agreed upon by at least half of the clinicians in that subcondition.

A = Antisocial	OC = Obsessive-compulsisve PD
B = Borderline	PR = Paranoid PD
H = Histrionic	DY = Dysthymic Disorder or "Depressive PD"
N = Narcissistic	SD = Schizoid PD
	PS = Passive-aggressive PD
	DP = Dependent PD

Table 14

Number of "Correct" Chosen Diagnoses Matching the Intended Diagnoses for Each Condition

	LO DOM LO TOTAL#		LO DOM HI TOTAL#		HI DOM LO TOTAL#	
	LO CAT#	HI CAT#	LO CAT#	HI CAT#	LO CAT#	HI CAT#
MED TYP	4 13% (n=30)	3 13% (n=31)	2 6% (n=31)	3 10% (n=30)	5 16% (n=30)	10 31% (n=32)
HI TYP	9 28% (n=32)	11 34% (n=32)	9 8% (n=32)	13 13% (n=31)	17 55% (n=30)	24 75% (n=31)
	DESIGN 1		DESIGN 2		DESIGN 3	

n = # of valid cases out of 32 possible cases per cell.

Table 15

Number of "Correct" Chosen Diagnoses Matching the Intended Diagnosis Collapsed Across Designs 1 and 3

LO AND HI DOM COLLAPSED

LO TOTAL#

	LO CAT#	HI CAT#
MED TYP	9 (n=60)	13 (n=63)
HI TYP	26 (n=62)	35 (n=63)

n = # of valid cases out of 64 possible cases per cell.

Table 16

Number of "Correct" Chosen Diagnoses Matching the Intended Diagnosis Collapsed Across Designs 1 and 2

LO DOM

LO AND HI TOTAL# COLLAPSED

	LO CAT#	HI CAT#
MED TYP	6 (n=61)	6 (n=61)
HI TYP	18 (n=64)	24 (n=63)

n = # of valid cases out of 64 possible cases per cell.

Table 17

Number of "Correct" Chosen Diagnoses Matching the Intended Diagnosis Collapsed Across Designs 2 and 3

LO AND HI DOM COLLAPSED

LO AND HI TOTAL# COLLAPSED

	LO CAT#	HI CAT#
MED TYP	7 (n=61)	13 (n=62)
HI TYP	26 (n=62)	37 (n=62)

n = # of valid cases out of 64 possible cases per cell.

Table 18

McNemar Tests on Number of "Correct" Chosen Diagnoses Matching the Intended Diagnosis for Typicality

Design Combination 1 + 3

MED TYP	HI TYP incorrect	HI TYP correct	
correct	A 10	B 12	X = 22.74
incorrect	C 52	D 47	p < .001

Design Combination 1 + 2

MED TYP	HI TYP incorrect	HI TYP correct	
correct	A 10	B 2	X = 14.38
incorrect	C 72	D 37	p < .001

Design Combination 2 + 3

MED TYP	HI TYP incorrect	HI TYP correct	
correct	A 6	B 14	X = 31.13
incorrect	C 52	D 48	p < .001

Table 19

McNemar Tests on Number of "Correct" Chosen Diagnoses Matching the Intended Diagnosis for Dominance

Design Combination 1 + 3

		HI DOM incorrect	HI DOM correct	
LO DOM	correct	A 9	B 17	X = 15.85
	incorrect	C 57	D 37	$p < .001$

Design Combination 2 + 3

		HI DOM incorrect	HI DOM correct	
MED DOM	correct	A 11	B 12	X = 16.98
	incorrect	C 54	D 42	$p < .001$

Table 20

McNemar Tests on Number of "Correct" Chosen Diagnoses
Matching the Intended Diagnosis
for Category Feature Number

Design Combination 1 + 3		HI CAT# incorrect	HI CAT# correct	
LO AT#	correct	A 17	B 17	X = 1.84
	incorrect	C 59	D 27	p = n.s.

Design Combination 1 + 2		HI CAT# incorrect	HI CAT# correct	
LO CAT#	correct	A 14	B 8	X = 1.36
	incorrect	C 72	D 22	p = n.s.

Design Combination 2 + 3		HI CAT# incorrect	HI CAT# correct	
LO CAT#	correct	A 11	B 21	X = 5.92
	incorrect	C 60	D 27	$p < .02$

Table 21

McNemar Tests on Number of "Correct" Chosen Diagnoses
Matching the Intended Diagnosis
for Total Number of Features

Design
Combination
1 + 2

		HI TOT# incorrect	HI TOT# correct	
LO TOT#	correct	A 20	B 4	X = 0, df =1
	incorrect	C 74	D 21	p = n.s.

Table 22

Number of Chosen Diagnoses Matching the Modal Diagnoses in Each Condition

	DESIGN 1: LO DOM LO TOTAL#		DESIGN 2: LO DOM HI TOTAL#		DESIGN 3: HI DOM LO TOTAL#	
	LO CAT#	HI CAT#	LO CAT#	HI CAT#	LO CAT#	HI CAT#
MED TYP	13 43% (n=30)	6 19% (n=31)	10 31% (n=32)	11 37% (n=30)	4 13% (n=31)	14 44% (n=32)
HI TYP	9 28% (n=32)	4 13% (n=32)	12 38% (n=32)	11 35% (n=31)	18 58% (n=31)	24 75% (n=32)

n = # of chosen diagnoses per cell. The maximum n per condition is 32, corresponding to the eight clinicians within that condition who were asked to diagnosis each of the four hypothetical clients associated with each intended diagnosis.

Table 23

Number of Chosen Diagnoses Matching the Modal Diagnosis Collapsed Across Designs 1 and 3

LO AND HI DOM COLLAPSED

LO TOTAL#

	LO CAT#	HI CAT#
MED TYP	17 (n=40)	20 (n=38)
HI TYP	27 (n=36)	28 (n=39)

n = # of valid cases out of 64 possible cases per cell

Table 24

Number of Chosen Diagnoses Matching the
Modal Diagnosis Collapsed Across
Designs 1 and 2

LO DOM

LO AND HI TOTAL# COLLAPSED

	LO CAT#	HI CAT#
MED TYP	23 (n=49)	17 (n=45)
HI TYP	21 (n=38)	15 (n=32)

n = # of valid cases out of 64 possible cases per cell

Table 25

Number of Chosen Diagnoses Matching the Modal Diagnosis Collapsed Across Designs 2 and 3

LO AND HI DOM COLLAPSED

LO AND HI TOTAL# COLLAPSED

	LO CAT#	HI CAT#
MED TYP	14 (n=35)	25 (n=53)
HI TYP	30 (n=44)	35 (n=55)

n = # of valid cases out of 64 possible cases per cell

Table 26

McNemar Tests on Number of Chosen Diagnoses
Matching the Modal Diagnosis
for Typicality

Design Combination 1 + 3

		HI TYP incorrect	HI TYP correct	
MED TYP	correct	A 5	B 19	X = 4.05
	incorrect	C 14	D 15	p < .05

Design Combination 1 + 2

		HI TYP incorrect	HI TYP correct	
MED TYP	correct	A 7	B 13	X = 1.71
	incorrect	C 19	D 14	p = n.s.

Design Combination 2 + 3

		HI TYP incorrect	HI TYP correct	
MED TYP	correct	A 6	B 24	X = 5.76
	incorrect	C 19	D 19	p < .02

Table 27

McNemar Tests on Number of Chosen Diagnoses Matching the Modal Diagnosis for Dominance

Design Combination 1 + 3

		HI DOM incorrect	HI DOM correct	
LO DOM	correct	A 6	B 10	X = 0.94
	incorrect	C 9	D 11	p = n.s.

Design Combination 2 + 3

		HI DOM incorrect	HI DOM correct	
MED DOM	correct	A 10	B 21	X = 4.97
	incorrect	C 16	D 24	p < .05

Table 28

McNemar Tests on Number of Chosen Diagnoses
Matching the Modal Diagnosis
for Category Feature Number

Design Combination 1 + 3

		HI CAT# incorrect	HI CAT# correct	
LO CAT#	correct	A 7	B 19	X = 0.23
	incorrect	C 13	D 10	p = n.s.

Design Combination 1 + 2

		HI CAT# incorrect	HI CAT# correct	
LO CAT#	correct	A 12	B 16	X = 0.19
	incorrect	C 20	D 9	p = n.s.

Design Combination 2 + 3

		HI CAT# incorrect	HI CAT# correct	
LO CAT#	correct	A 11	B 25	X = 0.05
	incorrect	C 19	D 9	p = n.s.

Table 29

McNemar Tests on Number of Chosen Diagnoses
Matching the Modal Diagnosis
for Total Number of Features

Design Combination 1 + 2

		HI TOT# incorrect	HI TOT# correct
LO TOT#	correct	A 8	B 9
LO TOT#	incorrect	C 19	D 11

X = 0.21
p = n.s.

Table 30

Number and Percentage of Chosen Diagnoses Within Each Condition that Matched the Intended Diagnosis And Were Agreed Upon by at Least Half of The Clinicians in Each Subcondition

	DESIGN 1: LO DOM LO TOTAL#		DESIGN 2: LO DOM HI TOTAL#		DESIGN 3: HI DOM LO TOTAL#	
	LO CAT#	HI CAT#	LO CAT#	HI CAT#	LO CAT#	HI CAT#
MED TYP	4 13% (n=30)	0 0% (n=31)	0 0% (n=32)	0 0% (n=30)	0 0% (n=31)	5 16% (n=32)
HI TYP	5 16% (n=32)	4 13% (n=32)	5 16% (n=32)	8 26% (n=31)	13 42% (n=31)	24 75% (n=32)

n = # of chosen diagnoses per cell. The maximum n per condition is 32, corresponding to the eight clinicians within that condition who were asked to diagnosis each of the four hypothetical clients associated with each intended diagnosis.

Table 31

Summary Chart Comparing the Results Using the -7 to +7 Measure with the Results Using the Proportion Measure

Measure	Factor	Design Combination 1 + 3	1 + 2	2 + 3
DV1	Typ	***	***	****
-7 to +7	Dom	***	.	***
	Cat#	-	-	*
	Tot#	.	-	.
DV1	Typ	****	****	****
Propor.	Dom	****	.	****
	Cat#	-	-	*
	Tot#	.	-	.
DV2	Typ	*****	*****	*****
-7 to +7	Dom	*****	.	****
	Cat#	*****	****	*****
	Tot#	.	-	.
DV2	Typ	**	*	**
Propor.	Dom	*	.	***
	Cat#	-	-	-
	Tot#	.	-	.

KEY:
***** = p < or = .00001
**** = p < or = .0001
*** = p < or = .001
** = p < or = .01
* = p < or = .05
- = not significant
. = not applicable

DV1 = Measure using intended diagnosis.
DV2 = Measure using modal diagnosis.

APPENDIX C

Experimental Design

Experimental Design

	LO DOM LO TOTAL#		LO DOM HI TOTAL#		HI DOM LO TOTAL#	
	LO CAT#	HI CAT#	LO CAT#	HI CAT#	LO CAT#	HI CAT#
MED TYP	$\frac{1}{3}$	$\frac{3}{9}$	$\frac{2}{6}$	$\frac{6}{18}$	$\frac{2}{3}$	$\frac{6}{9}$
HI TYP	$\frac{1}{3}$	$\frac{3}{9}$	$\frac{2}{6}$	$\frac{6}{18}$	$\frac{2}{3}$	$\frac{6}{9}$
	DESIGN 1		DESIGN 2		DESIGN 3	

KEY: Low Dominance proportion = 1/3
High Dominance proportion = 2/3

Within each cell,
Numerator = # of Category Features
Denominator = # of Total Features

APPENDIX D

Results of the -7 to +7 Dependent Measure

The -7 to +7 Dependent Variable with Corresponding Results

Boykin (1987) used a dependent variable that combined accuracy and typicality by assigning a positive value to the 1-7 typicality ratings of accurate diagnoses, and a negative value to the 1-7 typicality ratings of inaccurate diagnoses. Thus, an inaccurate diagnosis with the highest possible rating of 7 would be represented as a single data point of -7, whereas an accurate diagnosis with the highest possible rating of 7 would be represented as a single data point of +7. Given its range from -7 to +7, this dependent variable will henceforth be referred to as the -7 to +7 measure. The accuracy of the diagnosis chosen by each clinician was determined by its correspondence with the intended diagnosis, that is, with the diagnostic category that each profile was intended to describe. In addition to a discussion of the -7 to +7 measure, this appendix includes the experimental results of the present study using the -7 to +7 dependent measure.

In the present study, the -7 to +7 measure was judged to be somewhat incompatible with the prototype model of categorization. This dependent veasure places the inaccurate diagnosis on a -7 to +7 continuum with the correct diagnosis, suggesting that the more correct or typical a given diagnosis is (e.g., Borderline), then the less accurate or typical another diagnosis must be (e.g., Histrionic). More consistent with the prototype model of psychodiagnosis is the assumption that two forms of psychopathology, such as the two personality disorders cited above, are not mutually exclusive; they can be considered along two different continua, such that a personality profile may be equally reflective of both disorders. In the present study, a measure judged to be more compatible with the prototype model was used, as discussed in Chapter III. Unlike the -7 to +7 measure, the alternative measure devised examined the clinician's typicality rating for the chosen diagnosis within the greater context of that clinician's typicality ratings for the other personality disorder diagnoses that were not chosen.

Two Dependent Variables Based Upon the -7 to +7 Measure

In order that the present results could be compared with earlier studies that employed the -7 to +7 measure, two variations of this dependent measure were created. The first variable was created using the intended diagnosis as the standard for comparison, whereas the second variable used the modal diagnosis as the standard for comparison. For the first variable, a positive value was assigned to the 1-7 typicality ratings of the diagnoses that corresponded to the intended diagnosis, and a negative value was assigned to the 1-7 typicality ratings of diagnoses that did not correspond to the intended diagnosis. This resulted in an "accuracy-typicality score" for each diagnosis chosen by the clinicians. For the second variable, a positive value was assigned to the 1-7 typicality ratings of the chosen diagnoses when they corresponded to the modal diagnosis, and a negative value was assigned to the 1-7 typicality ratings of the chosen diagnoses when they did not correspond to the modal diagnosis. This resulted in a "reliability-prototypicality score" for each diagnosis chosen by the clinicians.

Results

The -7 to +7 data were bimodally distributed, thereby violating the required assumptions for parametric statistics. Therefore, a nonparametric test was used. The Wilcoxon Matched-Pairs Signed-Ranks Test was judged to be the most appropriate nonparametric test for two related samples of ordinal data (Segal, 1956).

For each dependent variable, separate analyses were performed for the three different combinations of designs (1 + 3, 1 + 2, and 2 + 3). For each analysis, the data for the four cells corresponding to a given condition, such as high typicality, were combined and then ranked.

The -7 to +7 Dependent Measure Using the Intended Diagnosis as the Standard of Comparison. When the intended diagnosis was used as the standard for determining the accuracy of the chosen diagnosis, the clinician's chosen diagnoses were associated with significantly higher accuracy-prototypicality scores when the intended category features of the hypothetical client were of high typicality compared to when they were of medium typicality. The z-scores and p values corresponding to each design combination are as follows:

Design Combination 1 + 3, $z = -3.70$, $p = .0002$;

Design Combination 1 + 2, $z = -3.56$, $p = .0004$;

Design Combination 2 + 3, $z = -4.03$, $p = .0001$.

Across both of the two design combinations that included dominance as a factor (1 + 3 and 2 + 3), a main effect for dominance was revealed. The clinician's chosen diagnosis were associated with a significantly higher accuracy-prototypicality score when the category features represented a highly dominant proportion of all the features presented relative to when the category features represented a less dominant proportion of all the features presented. The z-scores and p values for the dominance main effects are as follows:

Design Combination 1 + 3, $z = -3.54$, $p < .0004$;

Design Combination 2 + 3, $z = -3.39$, $p < .0007$.

A main effect for the number of category features was revealed only for Design Combination 2 + 3. For this combination of designs, the subject's chosen diagnoses were associated with significantly higher accuracy-prototypicality scores when the profile contained a high number of category features, compared to when the profile contained a low number of category features ($z = -1.95$, $p = .0507$). No main effects for category feature number were revealed for Design Combination 1 + 3, nor for Design Combination 1 + 2 ($z = -1.1562$, $p = .2476$; and $z = -0.8947$, $p = .3709$, respectively). Consistent with predictions, no main effect was revealed for total feature number, which was included as an experimental control variable in Design Combination 2 + 3 ($z = -0.1402$, $p = 0.8885$).

<u>The -7 to +7 Dependent Measure Using the Modal Diagnosis as the Standard of Comparison</u>. Across all three Design Combinations (1 + 3, 1 + 2, and 2 + 3), the Wilcoxon tests revealed a main effect for feature typicality. As predicted, the clinician's chosen diagnoses were associated with significantly higher reliability-prototypicality scores when the features in the profile were of high typicality, compared to when the features were of medium typicality. The z-scores and p values for the typicality main effects are as follows:

Design Combination 1 + 3, $z = -4.94$, $p < .00001$;

Design Combination 1 + 2, $z = -4.86$, $p < .00001$;

Design Combination 2 + 3, $z = -4.86$, $p < .00001$.

Across both design combinations that included dominance as a factor (1 + 3 and 2 + 3), a main effect for dominance was revealed. The clinicians' chosen diagnoses were associated with significantly higher reliability-prototypicality scores when the category features represented a highly dominant proportion of all the features presented relative to when the category features represented a less dominant proportion of all the features presented. The z-scores and p values for the dominance main effects are as follows:

Design Combination 1 + 3, $z = -4.18$, $p = .00001$;

Design Combination 2 + 3, $z = -3.81$, $p = .0001$.

Finally, across all three design combinations, a main effect for the number of category features was revealed. Subject's chosen diagnoses were associated with significantly higher reliability-prototypicality scores when the profile contained a high number of category features, compared to when the profile contained a low number of category features. The z-scores and p values corresponding to each of the design combinations are as follows:

Design Combination 1 + 3, $z = -4.31$, $p < .00001$;

Design Combination 1 + 2, $z = -3.83$, $p < .0001$;

Design Combination 2 + 3, $z = -4.06$, $p < .00001$.

Consistent with predictions, no main effect was revealed for total feature number, which was included as an experimental control variable in Design Combination 2 + 3 ($z = -0.1568$, $p = 0.8754$).

Comparison of Results Using Other Dependent Variables. A summary chart comparing the experimental results using the -7 to +7 measure with the results using the proportion measure is presented in Table 31. When the intended diagnosis was the standard, analyses of the two types of dependent measures yielded a similar pattern of results. In this context, however, the proportion dependent variable was associated with more significant results relative to the -7 to +7 dependent variable.

When the modal diagnosis was examined, analyses of the two types of dependent measures also revealed a similar pattern of results. Within this context, analyses performed on the -7 to +7 measure yielded more significant results than did those performed on the proportion measure. Additionally, whereas analyses on the proportion dependent measure revealed no significant effects for category feature number, analyses on the +7 to -7 dependent measure yielded a highly significant main effect for category feature number across all three design combinations.

Note that whereas the proportion dependent variable examined the prototypicality of the intended or modal diagnosis in the context of the prototypicality of the other diagnoses, the -7 to +7 measure examined the prototypicality of only the intended or modal diagnosis in isolation. The most likely explanation of the discrepancy in the effects of category feature number on the proportion measure vs. the -7 to +7 measure rests on the fact that the latter measure does not take into account the ratings made of the nontarget diagnoses. It appears that although category feature number did influence the ratings of the modal diagnosis, this factor also had a corresponding similar influence on the ratings of the nonmodal diagnoses. When ratings of the modal diagnosis are examined in isolation, they appear to vary as a function of the number of category features presented. When these ratings are adjusted to take into account the overall ratings made of the other diagnoses, however, this effect disappears.

Thus, the results of the proportion and the -7 to +7 dependent variables are quite similar across the majority of the analyses conducted. The discrepancies that do occur, however, illustrate the weakness of the -7 to +7 measure, and underscore the importance of considering prototypicality ratings of a target item in the context of corresponding ratings of other items.

APPENDIX E

Procedures Used in McNemar Tests

Procedures Used in McNemar Tests

In preparation for each McNemar test, each subject's two observations corresponding to the two levels of a given factor, within each level of the other factors, were paired together. A 2 by 2 matrix, such as that depicted below, was then constructed for each test.

		HI TYP	
		incorrect	correct
MED TYP	correct	A	B
	incorrect	C	D

$$X = \frac{[(A - D) - 1]}{A + D}$$

$$d.f. = 1$$

This matrix resulted in four cells that represented the four possible outcomes of the paired observations given the possibility of either correct or incorrect diagnoses for one or both levels of the factor in question. The paired observations were then represented as a single data point and categorized into one of four cells, as illustrated in the following example comparing medium versus high typicality:

Cell A - the number of paired observations in which the diagnosis in the medium typicality condition was correct, but the corresponding diagnosis in the high typicality condition was incorrect;

Cell B - the number of paired observations in which the diagnosis in the medium typicality condition was correct and the corresponding diagnosis in the high typicality condition was correct;

Cell C - the number of paired observations in which the diagnosis in the medium typicality condition was incorrect, and the corresponding diagnosis for the high typicality condition was incorrect;

Cell D - the number of paired observations in which the diagnosis in the medium typicality condition was incorrect, but the corresponding diagnosis in the high typicality condition was correct.

The formula for calculating the chi-square statistic for the McNemar test is presented above (Segal, 1956). For each design combination, a McNemar test compared the number of cases in Cell A versus Cell D under the null hypothesis that there were no differences in the number of cases in each cell. When typicality was analyzed, for example, there were significantly more cases in Cell D than in Cell A. That is, compared with the number of paired observations in Cell A, there was a significantly greater number of paired observations in which the subject's chosen diagnosis in the medium typicality condition was incorrect, but the that subject's corresponding chosen diagnosis in the high typicality condition was correct. Thus, as predicted, the clinicians were significantly more likely to give diagnoses that corresponded to the intended diagnosis when the features of the hypothetical client were of high typicality, relative to when they were of medium typicality. A further discussion of the results of the McNemar tests is presented in Chapter III.