

The Effects of the Confirmation Bias on Diagnostic Decision Making

A Thesis

Submitted to the Faculty

of

Drexel University

by

Meagan Carleton Parmley

in partial fulfillment of the

requirements for the degree

of

Doctor of Philosophy

October, 2006

Acknowledgements

First, I would like to acknowledge James D. Herbert, Ph.D. for his mentorship during the past five years. He has served as an exemplary professor, mentor, academic advisor, supervisor, and dissertation chair. He has provided consistent and tireless support during my graduate career.

I would like to acknowledge the dissertation committee: Pamela Geller, Ph.D., Kirk Heilbrun, Ph.D., Evan Forman, Ph.D., and Alexander Friedlander, Ph.D. for their time and assistance.

I would like to acknowledge the lab members of the Anxiety Research and Treatment Program at Drexel University as well as the 2005-2006 interns at the Southwest Consortium Pre-Doctoral Psychology Internship Program for their assistance both directly and indirectly related to this study.

I would like to acknowledge Clifford C. Qualls, Ph.D. for his assistance with the statistical analyses of this study.

I would also like to acknowledge my family for their support not only within the past five years, but throughout the past twenty-eight. Finally, I would like to acknowledge my husband, Jeremy, for his endless support since we met five years ago.

Table of Contents

List of Tables.....	vii
List of Figures.....	viii
Abstract	ix
1. Introduction.....	1
1.1 Early conceptualization and classification of mental illness.....	4
1.1.1 Renaissance and Enlightenment: The Gradual Shift to Naturalism...9	
1.1.2 Modern Classification of Psychopathology.....	12
1.1.3 Development of the Formal Diagnostic Classification Systems of Psychopathology.....	19
1.2 Development of the DSM-III.....	25
1.2.1 Utility of the DSM-III.....	29
1.2.2 Evolution of DSM -III to DSM-III-R, and III-R to IV.....	34
1.3 Alternatives to Traditional Psychodiagnosis.....	37
Treatment Utility.....	37
Functional Analysis.....	38
Expert Systems, Logical Functional Analysis, and the Functional Diagnostic Category Approach.....	39
1.3.1 Narrative Approach.....	41
1.3 Factors Impacting the Accuracy of Psychiatric Diagnosis.....	43
1.3.1 Cognitive Biases and Heuristics.....	46
1.4 The Confirmation Bias.....	50

1.4.1 The Confirmation Bias and Psychiatric Diagnosis.....	54
1.4.2 The Confirmation Bias in the Clinical Context.....	58
1.4.3 Real World Importance of Addressing the Confirmation Bias.....	62
1.5 Strategies for Debiasing.....	64
1.5.1 Structured Clinical Interviews.....	64
1.5.2 Perspective Taking.....	65
1.5.3 Computerized Programs.....	66
1.5.4 Decreasing Overconfidence.....	66
1.5.5 Delayed Decision Making.....	68
1.5.6 Considering the Opposite.....	69
1.5.7 Training.....	69
1.5.8 Considering Alternatives.....	70
1.5.9 Education.....	72
1.5.10 Additional Education Studies.....	75
1.6 Potential Errors Resulting from Debiasing.....	77
1.7 Statement of Purpose.....	79
2. Methods.....	85
2.1 Participants.....	85
2.2 Research Design.....	86
2.3 Materials.....	87
2.3.1 Case Vignettes.....	87
2.4 Procedure.....	88
2.5 Statistical Analyses.....	91

2.6 Statistical Power.....	92
3. Results.....	93
3.1 Demographics.....	93
3.2 Data Analytic Strategy.....	94
3.3 Main Analyses.....	94
3.4 Hypotheses.....	95
3.5 Additional Analyses.....	97
4. Discussion.....	99
4.1 Purpose.....	99
4.2 Summary of Results.....	100
4.3 Implications.....	101
4.4 Limitations.....	107
4.5 Strengths.....	110
4.6 Future Directions.....	111
List of References.....	117
Appendices	
A: Listserv Administrator Email.....	134
B: Recruitment Letter.....	135
C: Page One of Data Collection Website.....	136
D: General Instructions Provided to Both Instructional Groups at Time One....	139
E: General Instructions Provided to Both Instructional Groups at Time Two...	140
F: Bias Sensitization Instructions.....	141
G: Case Vignettes.....	142

H: Clinical Impression Record Form.....	148
I: Demographic Form.....	149
J: Confirmation Bias Information.....	150

List of Tables

Table 1. Gender.....	126
Table 2. Age.....	127
Table 3. Years in Practice.....	128
Table 4. Orientation.....	129
Table 5a. Percentage of Change in Each Condition.....	130
Table 5b. Percentage of Change in Each Condition by Vignette Type.....	131
Table 6. Accuracy in Diagnoses Across Conditions.....	132
Table 7. Accuracy in Diagnoses Across Conditions by Vignette Type.....	133

List of Figures

Figure 1. Research Design.....124

Figure 2. Accurate Diagnoses by Vignette.....125

Abstract

The Effects of the Confirmation Bias on Diagnostic Decision Making

Meagan Carleton Parmley, Ph.D.

Supervisor: James D. Herbert, Ph.D.

Research has shown that the confirmation bias affects judgments in a wide array of contexts, including clinical settings. Studies have demonstrated that the confirmation bias not only affects one's hypothesis testing strategy, but also how one interacts with others and how others in turn respond. In addition, several studies investigating the effects of diagnostic labels have found that such labels affect the way in which clinicians interpret information. Thus, once clinicians make a diagnosis, the confirmation bias has the potential to color their subsequent processing of new data, resulting in confirmation of initial diagnostic impressions even if these impressions are contradicted by subsequent data. In an attempt to mitigate the effects of the confirmation bias, several studies have investigated various debiasing techniques, including warning or educating subjects about biases. The results of these studies indicate that such techniques may be useful if subjects are adequately educated about a bias. The purpose of this study was to examine the possible effects of the confirmation bias in psychodiagnostic assessment, and to examine the effectiveness of educating clinicians specifically about the confirmation bias. Clinicians received two case vignettes, followed by additional information about each case one week later. The additional information was either consistent with the diagnosis indicated by the initial data, or inconsistent with it. Participants were asked to make a diagnosis after reading each vignette. Half of the participants received information about

the confirmation bias at time two before receiving additional information about the cases. Results suggest that clinicians did evidence the confirmation bias during performance of this task. Specifically, clinicians tended to remain consistent with their original diagnosis when a change in diagnosis was required from time one to time two. In addition, the confirmation bias instructions did not have a significant effect on participants' performance during this task. Participants did not overcorrect, or incorrectly alter their initial diagnoses, when they received confirmation bias instructions. Finally, age was a predictor of diagnostic accuracy; specifically, as age increased, performance decreased. There was also some evidence to suggest that those participants who identified their orientation as cognitive behavioral performed more accurately. The results hold important implications about commonly occurring biases in the psychodiagnostic assessment process.

Chapter 1: Introduction

Records of mental illness extend as far back as the earliest historical records. However, psychiatric diagnosis is a much more recent phenomenon, reflecting the conceptualization of psychology within a medical framework. One of the earliest conceptualizations of mental illness was known as demonology, or the notion that the devil resides within a person's body and controls their mind. The early Chinese, Egyptians, Babylonians and Greeks believed in demonology and regularly performed exorcisms in an attempt to cast the evil spirits from the afflicted body. In addition, extreme measures were taken at times to make the body uninhabitable to the spirits, including measures such as starvation or even beatings (Zilboorg & Henry, 1941).

Hippocrates (460 to 377 BC) was among the first to regard this bizarre behavior as an illness, which he likened to the common cold. He believed in the notion that something wrong with the soma, or body, interrupts thought and behavior. This idea is also known as somatogenesis. Therefore, mental illness was seen as a disease of the brain resulting in abnormal behaviors. He classified mental disorders into six categories (a) epilepsy, (b) mania (states of abnormal excitement), (c) melancholia (states of abnormal depression), (d) paranoia (an illness today that we would term mental retardation), (e) hysteria, and (f) phrenitis, also known as brain fever (Davison 1998; Zilboorg, 1941). He believed that a balance in the four humors (fluids of the body) blood, black bile, yellow bile, and phlegm was responsible for normal brain functioning and mental health. Any imbalance in these humors was thought to result in a mental disorder (Zilboorg, 1941). These early attempts at the psychodiagnostic process suggest

that categorizing mental disorders was an important first step in the treatment of these illnesses.

Although Hippocrates' ideas obviously did not hold up to later discoveries, the Greeks and Romans largely accepted them over the next seven centuries. Hippocrates' conceptualization of mental illness as stemming from a problem within the body was a very new and unique approach, especially compared to approaches like demonology where the problem was conceptualized as an uninvited spirit temporarily inhabiting the body. Although his theory of humors was not accurate, his conceptualization of mental illness as physical rather than supernatural was a significant step in understanding mental illness. Although his classification system was rudimentary by today's standards and involved only six diagnostic categories and four etiological sources (four humors), it did represent the phenomenon of psychodiagnostic classification even as early as the Greco-Roman period.

Following the demise of the Roman Empire, the church became the primary institution through which the mentally ill were treated. Physicians such as Galen (A.D. 130-200) of the Roman Empire were among the last physicians of the classical era (Zilboorg, 1941). A clear shift had occurred from a somatic conceptualization of mental illness toward more spiritual, supernatural explanations, representing a step backward, away from the naturalistic account of Hippocrates. This shift was a reflection of the influence of the church on all aspects of life during this time. As the influence of the church continued to grow, monasteries became the primary location for the treatment of the mentally ill. The early Christian authorities were often puzzled and frightened by the very idea of mental illness, as most disorders were thought to be the work of the devil

(Zilboorg, 1941). Monks would often pray over the mentally ill and have the patients drink potions during certain phases of the lunar cycle in an attempt to heal them (Davison, 1998). Clearly, Hippocrates' and Galen's humoral classification went into decline during this time, although it did not disappear completely. In actuality, their classification system began to coexist with the demonological paradigm (Schoeneman, 1977).

Beginning in the thirteenth century, people turned to witchcraft to explain the famines and plagues that killed a large number of the population. Because the church had such a great influence on all aspects of people's lives, the population turned to religion to explain the deaths of so many. Because death on such a massive scale could not be considered to be an act of God, people turned to the devil and believed the devil represented the foundation for witchcraft. Large hunts were made in search of these "witches" in an attempt to explain and understand what was inexplicable to them at the time. Although there was little logic to this idea, those who were accused of witchcraft and did not confess were tortured. Those who confessed and wanted to repent were sent to prison for life and those who were convicted and not willing to repent were executed. Although most of the accused were likely not mentally ill, some of the accused were targeted because they had allegedly made bizarre statements suggesting that they had seen or had intercourse with the devil. Today, professionals would likely consider someone like this to be suffering from hallucinations or delusions, and therefore some of these men and women known as "witches" are thought rather to have been psychotic (Zilboorg & Henry, 1941). Obviously, those who were tortured were likely confessing to accusations that were not true in order to alleviate at least some of their suffering.

Interestingly, torture was not allowed in England, and very few confessions there included psychotic-like symptoms (Schoeneman, 1977).

Early conceptualization and classification of mental illness

The *Malleus Maleficarum*, written by Henry Kramer and James Sprenger, was developed as a guide to seek out these “witches” (Anderson, 1970). This guide was known as the huntsman’s bible and described witches as “dissenters, schismatics and the mentally ill” (Schoeneman, 1977). There were two categories of women who were targeted throughout most of the witch hunts. The first was a melancholy group who were described as being in a depressed state. Other characteristics of this group included making statements that were obscure or threatening and any displays of odd behavior in general. The second category included women who had isolated themselves from the community (Schoeneman, 1977). Based on the information from these two categories, the descriptions potentially define a large percentage of women. Although they offered criteria to define what ‘melancholy’ means, these criteria were extremely vague and anyone working from these categories would likely be able to find these characteristics in a large part of the population.

In some of the witch hunts such as the Salem Witch Trials, some of the witches fit the stereotype of a “hag” or an older woman who was considered to be ugly or frightful, however of those who were killed, fourteen were women, six were men, and although some were very critical of the trials, others were merely disliked by the community (Schoeneman, 1977). Again, within this sample, it does not appear that there was any type of carefully considered classification system in place. At times it may have been that those individuals who the community feared, or those who spoke out against the

witch hunts, were among the accused. Again, these criteria defining “witches” were so broad that they included the mentally ill, anyone who did not believe in God and anyone displaying the least bit of bizarre behavior.

The *Malleus Maleficarum* listed two categories of etiological factors. The first was a supernatural intervention (“by the power of devils, with God’s permission”) and a natural defect (“as it is shown in the case of frantics and melancholy men, and in maniacs and some drunkards”) (Sprenger and Kramer, 1928, as cited in Zilboorg, 1941). This again, provides some evidence that accused ‘witches’ may have been anyone who did not follow the social rules of the community.

Finally, although there was little in the way of an official classification system, and the etiology was usually assumed to be the work of the devil, there were six categories identified with respect to how the devil injured humanity. Zilboorg (1941) describes these:

One is, to induce an evil love in a man for a woman, or in a woman for a man.

The second is to plant hatred or jealousy in anyone. The third is to bewitch them so that a man cannot perform the genital act with a woman, or conversely a woman with a man; or by various means to procure an abortion...The fourth is to cause some disease in any of the human organs. The fifth, to take away life. The sixth, to deprive them of reason. (p. 158)

Although there were some attempts at classification during the witch hunts, a major problem was the attempting to give explanations to events that were at the time unexplainable. The *Malleus Maleficarum* was based mostly on an emotional reaction of fear rather than an attempt to study the phenomenon from a scientific standpoint.

Anyone acting outside of the realm of normalcy at the time or anyone questioning the beliefs of the community was suspected of being a witch.

Clearly, these developments represented a great decline in the naturalistic approach to mental illness from the work of Hippocrates and Galen. However, this example again demonstrates an early desire for psychodiagnostic classification despite the unscientific nature of these classifications.

These examples also illustrate how biases played a vital role in determining who would be among those accused of being witches. Based on the criteria discussed above, if someone were somewhat depressed or otherwise mentally ill, did not believe in God, drank alcohol, had a disheveled appearance, or was simply somewhat isolated from the rest of the community, they risked being accused of being a witch. Clearly, the confirmation bias played a role in the mass killings of these men and women. One obvious example of this was the dunking and often subsequent drowning of witches in lakes. The belief was that if accused witches were dropped into a lake as they were tied to a chair and they sank, they drowned and were apparently guilty. Otherwise, the logic was that they would not have drowned. However, if they did float for a period of time, this was also a sure sign that they were guilty. Both outcomes were used to confirm their initial impression, and there was no opportunity for disconfirmatory evidence to be recognized.

By the fourteenth and fifteenth centuries, possession came to be viewed as an important etiological factor in insanity and supernatural explanations became prominent during the witch hunts of the 16th and 17th century (Schoeneman, 1977). However, supernatural explanations were not the only conceptualizations of mental illness from the

1200s to the 1600s. Many Europeans were kept in hospitals and stayed for the duration of their illness. Those who were considered incompetent as well as insane were hospitalized. Although there were very few mental hospitals in Europe in the 13th and 14th centuries, there were hundreds of hospitals for those with leprosy. At the end of the Crusades, in the fifteenth and sixteenth centuries, leprosy diminished, and the focus turned to the mentally ill. Many of the hospitals used for leprosy were turned into asylums for the mentally impaired. The main goal for patients confined in these asylums was to get them to work. During the same time period, however, hospitals emerged that focused on the confinement of the mentally ill in an attempt to isolate them from the rest of the community. One hospital became more specifically geared toward the confinement of the mentally ill in London, known as the Priority of St. Mary of Bethlehem (Davison, 1988; Zilboorg, 1941). The conditions were extremely poor in this hospital and soon after it opened, it became known as “Bedlam,” describing the confusing and chaotic environment within the hospital. Tickets were even sold for tourists to view the confusion in this hospital as late as the 19th century.

During the fifteenth and sixteenth centuries, Rosen (1968) noted that medical professionals usually assumed that illnesses had a natural cause. However, if the illness was bizarre or they had no experience treating it, they may accept a supernatural explanation. However, Rosen (1968) also noted that not all physicians were inclined to accept the possibility of possession and some denied it outright. Kroll (1973) cites the *Encyclopedia of Barthoemaeus* in which mental disorders were dealt with in a non-demonological manner. It was first written in the 13th century and was widely read in the

days of the witch hunts. This illustrates the gradual shift toward naturalistic explanations of mental illness in the waning days of the late Medieval period.

Schoeneman also cites Paracelsus, a major Renaissance figure who contributed to the fields of medicine, chemistry and pharmacy. Although it is not clear whether he obtained a medical degree, he was known to practice medicine and to give medical advice. Paracelsus wrote in the early sixteenth century about the five categories of permanently insane people. These included (a) Lunatici (Disturbed by the moon), (b) Insani (generally insane), (c) Vesani (poisoned or contaminated by food or drink), (d) Melancholici (insane by their nature) and (e) Obsessi (referring to those who are possessed by the devil). Although his classification system was rudimentary, he clearly believed in both a supernatural as well as a natural account of mental illness. As is clear from the five categories listed above, many of the explanations for mental illness included supernatural explanations during the 13th through the 16th centuries. Again, due to the rudimentary nature of this classification system, one can imagine the biases taking place during this psychodiagnostic process. Specifically, there may have been evidence to suggest that none of these hypotheses were accurate; however, information was interpreted to fit one of these categories.

From the fifteenth to the start of the seventeenth century, mental illness was largely conceptualized as a supernatural phenomenon, although some still believed in a naturalistic explanation. Examples include the descriptions of witches and their relationship with the devil as well as a few of the categories cited by Paracelsus including, Lunatici (disturbed by the moon) and Obsessi (possessed by the devil). However, not everyone during this time conceptualized mental illness from a

supernatural perspective. Paracelsus cited other categories which included naturalistic explanations for mental illness including Vesani (poisoned by food or drink) and Malancholici (insane by their nature). This classification system suggests that the understanding of mental illness was relied mainly on supernatural explanations, however, some categories reflect the naturalistic understanding such as Vesani (poisoned by food or drink). Therefore, although there were some naturalistic explanations for bizarre behavior, (such as being poisoned by food or drink) many chose to view this behavior as the result of supernatural phenomena.

Renaissance and Enlightenment: The Gradual Shift to Naturalism

In the seventeenth and eighteenth centuries there was again a shift in the conceptualization of mental illness. The first mental hospital in the United States was founded in Williamsburg, Virginia, in 1773 (Davison, 1998). Despite the shift to a somatic conceptualization, the treatment of the mentally ill at that time continued to be based on explanations that had little or no evidence. This was exemplified by the practices of Benjamin Rush, considered to be the father of American psychiatry. Although Rush was known for improving the treatment of the mentally ill, one of his theories was that the mentally ill had an excessive amount of blood in their brains (Holmes, 1967). He therefore drew large amounts of blood from his patients. In addition, he believed in frightening mental patients and convincing them that they were going to die. (Davison, 1998; Farina, 1976; Holmes, 1967). Although Rush's approach would be viewed as cruel by today's standards, his approach represented another shift in the conceptualization of mental illness toward a naturalistic account. Here again, the confirmation bias was likely evident. If Rush witnessed one of his patients improving,

that alone may have been enough for him to confirm that his approach was useful, despite the many who died during this practice. This suggests that he likely disregarded large amounts of evidence that suggested that this practice was harmful to his patients and he likely reinterpreted it to fit with his initial hypothesis.

Rush's earlier theories were based largely on the works of other writers at the time, specifically, William Cullen, who held that diseases could be classified according to their symptoms and causes. He wrote *Synopsis Nosologiae Methodicae*, where he constructed an extensive nosology similar to the Linnean system of classification in biology (Holmes, 1967). Cullen is known for transforming Von Haller's work on the nervous system into a general explanation of nervous disease. He developed a classification of diseases that he believed stemmed from "disordered motions of sensations of the nervous system," and this is where the term "neurosis" derives (Knoff, 1970), although many credit Freud with the term. Cullen believed that nervous energy that stems from the nervous system determined the normal state of the body and believed that movement of the nervous fluid was the cause of all diseases (Knoff, 1970). His term "neurosis" began to replace supernatural explanations such as humors, demons and spirits, representing a strong shift in the conceptualization of mental disorders back to a somatic approach. As mentioned above, the somatic conceptualization of mental illness never completely disappeared from the 13th to the 18th centuries, and now supernatural explanations were increasingly supplanted by naturalistic ones.

Cullen classified all diseases by their symptoms into classes, orders, genera, and species. He classified man's diseases into four categories: (a) pyrexias, (b) neuroses, (c) cachexias, and (d) local diseases. "Neuroses" was further divided into four orders: (a)

comas (including apoplexy), (b) adynamias, (including autonomic dysfunctions and hypochondriasis), (c) spasms (including convulsions and hysteria), and (d) vesanias (the term for madness in Cicero's time)" (Knoff, 1970, p. 121). This classification system was the most detailed to date, and symbolizes the shift back to a naturalistic account of mental illness.

During the 18th century, the focus of treatment for the mentally ill kept in hospitals shifted from confinement to more humane treatment. Phillippe Pinel led the movement to treat the mentally ill in a more humane fashion. Pinel was also a critic of Cullen and developed his own system of classification consisting of five categories: (a) fevers, (b) inflammation, (c) hemorrhagic diseases, (d) neuroses (after Cullen), and (e) organic lesions. He then identified four subcategories of mania which included mania, melancholia, dementia and idiocy (Knoff, 1970). Pinel was in charge of an asylum in France, and during that time, removed the chains that were used to chain the patients to the walls. Although his focus was on improving the conditions for the mentally ill, his approach also relates to the broader view of treating mental illness. He believed that if these patients had become ill because of personal and/or social problems, he might be able to treat them by providing compassion, understanding and normal activity (Davison, 1998). This concern he had for his patients is consistent with other's accounts that he was more interested in a clinical description of his patients rather than a rigid system of classification (Knoff, 1970).

In the 1860s-1870s Louis Pasteur developed the germ theory of disease, which posited that disease is caused by microorganisms infecting the body. This provided a theoretical link between syphilis and general paresis. Further developments in the

understanding of syphilis arose in 1897 when Richard Von Krafft-Ebing inoculated patients with pieces of syphilitic sores, and they did not subsequently develop syphilis, suggesting that these patients had been infected earlier in their lives. It was not until 1905 that the specific organism that causes syphilis was discovered, which led to the idea that if one form of mental illness had a biological cause, then others likely did as well. Once again, Hippocrates' theory of somatogenesis increased in popularity, and dominated the field well into the 20th century.

Modern Classification of Psychopathology

During the late 19th century, the German Psychiatrist Emil Kraepelin, developed a classification system in an attempt to investigate the organic nature of mental illnesses. He suggested that a group of symptoms that often occur together might have an underlying physical cause. He believed that each mental disorder was unique with its own origins, symptoms, course and outcome. Kraepelin identified two major groups: dementia praecox (Schizophrenia) and manic-depressive psychosis. He believed that a chemical imbalance was the underlying cause of schizophrenia and an irregular metabolism was the cause of manic-depressive psychosis. This approach to classification became the basis for other diagnostic categories and Kraepelin's symptomatic approach to psychopathology was highly influential to the modern classification of mental disorders. His theories on the etiology and diagnosis of psychiatric disorders would form the theoretical basis of modern diagnostic schemes, including the *Diagnostic and Statistical Manual of Mental Disorders* and the World Health Organization's *International Classification of Disease* system.

Despite Kraepelin's work, still not enough was known to understand the origins of these disorders. Although more was understood about the nervous system in the 1800s, the somatic origins of mental disorders remained poorly understood. One discovery that was associated with senility and mental retardation was degeneration of brain cells, again supporting the idea that mental illness had a biological basis. It had been known that some mental patients suffered from a deterioration of their mental and physical abilities including delusions of grandeur and paralysis (Davison, 1998). In the 19th Century, this disease became known as general paresis, and while some knew that these patients had suffered from syphilis previously, there were many competing ideas for its origins. Griesinger (a German psychiatrist and neurologist) believed that liquor, tobacco and coffee could have been the cause, as the incidence was much higher in men.

Whereas this still did not explain why women were also diagnosed with the disease, the confirmation bias is again apparent within this approach. It appears that Griesinger developed the hypothesis that these illnesses were caused by liquor, tobacco and coffee simply because he observed this consumption frequently by men. It follows that each time he met with a male patient with general paresis who also endorsed the consumption of these substances, he gathered evidence to confirm his initial hypothesis. In addition, he was disregarding information that was inconsistent with this hypothesis, such as the fact that women were also suffering from this illness.

Whereas somatogenesis was the model that many had been using to explain mental disorders, psychogenesis, or the view that mental disorders were due to psychological malfunctions was alive and well in the 19th century in Austria and France. For example, Franz Anton Mesmer believed that a certain distribution of magnetic fluid

in the body was the cause of hysterical disorders. His treatment involved having patients sit around a covered tub. Bottles were placed at the bottom of the tub and rods were placed in the bottles, protruded through the covering. Mesmer would take the rods and touch certain parts of the patient's bodies. He believed that this would adjust the balance of the magnetic fluid and restore a person's sanity. Although this approach was largely somatogenic, Mesmer also practiced the technique of hypnosis, (a psychogenic approach to the treatment of the mentally ill) and the term "mesmerize" stems from Mesmer's work (Davison, 1998).

Jean Martin Charcot also studied hysterical states and once had a somatogenic belief about the origin of mental disorders. However, his conceptualization changed when he witnessed his students hypnotized a woman and had her fake hysterical symptoms. Charcot believed that she was truly hysterical, however his students showed him that they could remove the symptoms by waking her. It was at this time that Charcot began to conceptualize mental illness from a psychogenic point of view (Davison, 1998).

Throughout this and many of the treatments reviewed here, the confirmation bias was likely alive and well within the minds of these clinicians. If a clinician were to witness a patient's condition improving after the treatment, they were likely to continue with the treatment, despite evidence that suggested that it did not have any positive effect on their patients.

Stemming from Charcot's work, Josef Breuer a physician in Vienna, hypnotized Anna O. and repeated some of the words she spoke when she was hypnotized. It was shortly after that she was talking about events from her past with a great deal of emotion. He found that when she would wake from being hypnotized, she reported feeling better.

This led to the technique known as the cathartic method (Davison, 1998). Again, this reflected a shift from a somatogenic perspective to a perspective that suggested that mental illness stemmed from psychological problems or problems with the mind only. Throughout the works of Mesmer, Charcot and Breuer, mental illness was conceptualized from a psychogenic perspective. However, they differed from other supernatural explanations in that their explanations were somewhat more naturalistic. These approaches did not emphasize classification; rather the focus rested with the general principles of their theories and practices.

Just as Kraepelin was making progress with his classification system, Freud interpreted mental illness from a new vantage point, unconscious motivation (Knoff, 1970). It was in 1909 that Freud and Jung visited the United States lecturing at Clark University in Worcester, Massachusetts. It was shortly after this visit that an interest in psychoanalysis rose in the United States. Psychoanalysis was widely popular in the United States during the 1930's and 1940's when many psychoanalysts arrived from Europe. The psychoanalytic paradigm developed by Freud suggested that psychopathology results from unconscious conflicts. Freud classified the psyche into a tripartite model of the id, ego and superego. He believed that behavior results from a complex interaction of these three parts of the psyche. Freud asserted that there are five psychosexual stages that children enter as they develop. The oral stage, lasting from birth to 18 months, represents the time when the id is satisfied by feeding, and the sucking that accompanies the feeding. The anal stage, from 18 months to three years, represents the child's pleasure from passing feces. The phallic stage lasts from approximately age three to five or six when the child receives pleasure from stimulating the genitals. During the

next stage, known as the latency stage, the id plays a lesser role in motivating the child's behavior. Finally the genital stage is also known as the adult stage in which sexual interests are the primary source of gratification. Each person must resolve the conflicts between what the id desires and what the environment will provide during each stage of development. If an insufficient resolution is achieved, Freud believed this would lead to a fixation at that particular stage of psychosexual development. For example, this may take the form of an oral fixation that is due to an unresolved conflict between the id's needs and the environment. The child who sucks his or her thumb or the adult who smokes may be considered to have an oral fixation.

Although the psychoanalytic approach was later criticized for lacking a clear classification system, Freud did classify disorders especially with regard to presumed etiology, i.e., the psychosexual stages stemming from conflicts between the id and superego. In addition, the psychoanalytic approach distinguished between neurotic disorders and psychotic disorders. Neurotic disorders were thought to result from a neuronal problem, whereas psychotic disorders were thought to result from psychogenic factors. In addition, Freud asserted that people handle anxiety through the use of defense mechanisms, which are thought of as unconscious distortions of reality. Freud classified these defense mechanisms into seven categories: (a) repression, (b) projection, (c) displacement, (d) reaction formation, (e) regression, (f) rationalization, and (g) sublimation.

Whereas Freud's conceptualization of psychopathology was widely accepted during the 20th century, his approach is now increasingly regarded as lacking a scientific basis. Many of his key hypotheses are inherently unfalsifiable, and therefore cannot be

tested empirically. Those that can be tested have generally failed to garner support. (Crews & Bulkeley, 2001). It is especially in the case of unfalsifiable theories such as Freud's that clinicians may evidence the confirmation bias. In the case of Freud's theory, there is always room for the confirmation bias to occur, as there is always a way to reframe disconfirming evidence to confirm the initial theory. Freud's theory does this by utilizing an explanation of unconscious processes, which cannot be measured. Therefore, his theory can be relied upon to confirm ideas even when there is no outward supporting evidence.

Another influential figure in the field of psychology in the 20th century was Erik Erikson, who believed that people suffer from identity crises. These identity crises arise from challenges presented by one of Erikson's eight psychosocial stages of development. According to Erikson's theory if a person is able to meet the challenges presented by these stages, a crisis does not occur, however, if they are unable to learn and grow according to these stages, a crisis develops. Although Erikson did not directly address classification in his theory as his focus was on the identity crisis itself, he did classify the stages of psychosocial development and different types of identity crises that arise from each. His classification of psychosocial crises included: (a) trust vs. mistrust, (b) autonomy vs. shame, doubt, (c) initiative vs. guilt, (d) industry vs. inferiority, (e) identity vs. identity confusion, (f) intimacy vs. isolation, (g) generativity vs. stagnation, and (h) integrity vs. despair. One problem with his approach to psychopathology was that it was limited and narrowly focused on identity crises. It is not clear how other psychological disorders such as schizophrenia or bipolar disorder would fit into his theory of psychosocial development.

Several decades later, it was the interpersonal school that became widely popular. Harry Stack Sullivan, Frieda-Fromm-Reichmann, and neo-Freudians such as Fromm and Horney all contributed greatly to the development and evolution of the interpersonal perspective. The interpersonal approach focused on interactions among patients and their loved ones. The origin of psychopathology was thought to be found in interpersonal interactions (Klerman, 1984). Although Sullivan, Fromm-Reichmann and Horney's interpersonal contributions focused more on the principles of their theories rather than classification, Fromm's Sociological Theory included a classification of the ways in which people handle feelings of isolation. He asserted that people develop certain dominant strategies or personality styles that he identifies as different character types. These include: (a) the receptive character type, (b) the exploitative character type, (c) the hoarding character type, (d) the marketing character type, and (e) the productive character type (Carducci, 1998). Again, whereas there were some attempts at classification among this group, these theories were lacking a comprehensive and systematic diagnostic classification system.

Also during the early 20th century, behaviorists Pavlov, Watson, and Skinner made great strides with respect to human behavior, most notably, Pavlov's classical conditioning and Skinners operant conditioning. Although the behaviorists did not emphasize classification in the traditional sense, they did classify behavioral phenomena by their presumed functional etiology. One example of this classification is abnormal behavior presumably caused by classical vs. operant conditioning. Classical conditioning occurs when a neutral stimulus initially does not elicit a response. After being paired with an unconditioned stimulus, it functions as a conditioned stimulus and elicits a

response. The most common example of classical conditioning is the conditioned salivation response in Pavlov's dogs. Operant conditioning is the learning process by which the consequence of an operant response affects the likelihood that the response will occur in the future. Again, although these theories do not include classification systems per se, they do categorize behaviors according to their supposed etiology.

Despite Kraepelin's attempt to develop a standardized classification system, Freud's theory of psychoanalysis dominated the field until well into the twentieth century. Other schools of thought such as the interpersonal school, and behaviorism had their own approach to classification, however, a more formal classification system was still lacking.

Development of the Formal Diagnostic Classification Systems of Psychopathology

Although classification had been used in most of these earlier approaches to psychopathology, there was still no official list of diagnostic categories with operational criteria to use in assessment. There were attempts at developing classification schemes, however many failed. In 1882 the Royal Medico-Psychological Association developed a classification scheme, although it was never adopted by the members of the association (Fleming, 1933). This classification scheme included eight general classes of mental disorders: (a) oligophrenia (a mental deficiency, a condition of arrested or incomplete development of mind), (b) psychoneuroses (no differentiation was made between neuroses and psychoneuroses. The criteria include an abnormal mental state characterized by mental or motor fatigability and irritability and included generalized fear and phobias), (c) schizophrenic psychoses (including dementia praecox and paraphrenia), (d) psychopathic constitution (including paranoia), (e) affective psychoses, (f) epileptic

psychoses, (g) general paralysis, and (h) dementia (Fleming 1934). In 1932, the final revision of *The Classification of Mental Disorders* was published. However, it was not adopted by the Council of the Association at the time, and another revision was finally accepted in 1933 and served as the official classification of the Royal Medico-Psychological Association (Fleming, 1934). The original eight categories remained, with the major difference being the inclusion of three additional categories: confusional states, other psychoses associated with organic brain disease, and undetermined types. In addition to these three new categories, clinicians were asked to identify specifiers for each disorder. These included (a) heredity, (b) deprivation and special sense (hearing, sight), (c) critical periods, (puberty, adolescence, etc.), (d) child-bearing, (e) mental factors such as “sudden stress” (f) physiological disturbances such as malnutrition, (g) trauma such as injuries or operations, (h) toxic factors such as alcohol, drugs or poisons, (i) deficiency diseases such as pellagra, (j) diseases of the nervous system, (k) diseases of other systems such as the cardio-vascular or endocrine systems, (l) no factor ascertained, and (m) no history obtained (Fleming, 1934). At the time, this system was perhaps the most complex, not only with a classification system in place, but also a system to specify the suspected etiology of each disorder.

In 1840, the initial reason for developing a classification manual in the United States was for statistical purposes. As discussed by Frances, Pincus, First and Widiger in the introduction of the fourth edition of the DSM (American Psychiatric Association, 1994), the first attempt to collect information on mental illness in the United States was the recording of one category “Idiocy/Insanity” in the 1840 census. They go on to note

that by the 1880 census, there were seven recognized categories; (a) mania, (b) melancholia, (c) monomania, (d) paresis, (e) dementia, (f) dipsomania, and (g) epilepsy.

The Association of Medical Superintendents of American Institutions for the Insane, a forerunner of the American Psychiatric Association, adopted a revised version of the British system in 1886 (Davison, 1998). This group then adopted a new classification scheme in 1913 that included some of Kraepelin's work. However, consistency was still missing from these classification schemes. In 1917, the Committee on Statistics of the American Psychiatric Association (at the time it was called the American Medico-Psychological Association) and the National Commission on Mental Hygiene (which later became the National Association for Mental Health) developed a plan to collect statistical information from mental hospitals across the U.S. (Coolidge & Segal, 1999). Although the goal was clinical utility, the classification was still primarily a statistical one (American Psychiatric Association, 1994). There were many attempts at an official widely accepted classification scheme between 1882 and 1948. Contributors included the Royal Medico-Psychological Association, the Association of the Medical Superintendents of American Institutions for the Insane, American Medico-Psychological Association and the World Health Organization. However, the various classification systems differed in the number of recognized diagnoses. While some classified diagnoses for statistical use, others classified diagnoses for clinical use, and consistency was still lacking across these various classification schemes.

An *International List of Cause of Death* has been in existence for statistical purposes since 1893. This list was revised several times, and after World War II the World Health Organization adopted this list and produced the 6th revision which was then

titled the *International Statistical Classification of Diseases, Injuries and Causes of Death* known as the ICD (Dittman, 1991). The ICD listed all known diseases as well as descriptions of abnormal behavior. Mental Disorders had been added to the list in 1939, and although the WHO adopted this list, they did not accept the section on mental disorders. Even though psychiatrists had assisted in the development of the ICD and pushed for the list to be accepted by the WHO, the American Psychiatric Association eventually published its own *Diagnostic and Statistical Manual of Mental Disorders* in 1952 (Dittman, 1991).

The DSM-I included short descriptions of symptoms and briefly described each diagnostic category. The DSM-I was an advance over these other approaches to classification, as there was an official list of disorders with operational criteria for each disorder. However, the DSM-I was incompatible with the ICD-6, and there was a growing interest in developing an international classification of diseases in order to facilitate communication across countries (Spitzer & Wilson, 1968). Therefore, American representatives worked with these international committees to revise the mental disorders section of the ICD-6 which became effective in 1968 (Spitzer & Wilson, 1968). In 1965 the American Psychiatric Association set out to prepare a new edition of the DSM that would be compatible with the ICD-8 list of mental disorders. There were several ways in which the ICD-8 was modified for the development of the DSM-II. The ICD-8 included changes such as (a) changes in the organization and sequence of disorders, (b) certain diagnoses were omitted, (c) other diagnoses were added, and finally (d) additional coding digits were used in order to achieve greater specificity among categories.

The first change in the DSM-II was a modification of terms. For example, the DSM-I used the term Schizophrenic *Reaction* whereas DSM-II took the term reaction out and left it as “Schizophrenia.” The second difference was with regard to the organization of the manual. There were three major categories of mental disorders in the DSM-I whereas the DSM-II included ten categories. The DSM-I was also inconsistent regarding multiple disorders. Although it supported the general notion of multiple disorders, it did not allow certain combinations. For example, if one disorder was thought to be symptomatic of another, (such as alcoholism was thought to be symptomatic of another underlying disorder such as depression) the disorder was not listed. The DSM-II encouraged the identification of multiple disorders, even if one was symptomatic of another. The DSM-II also encouraged clinicians to identify physical conditions as a separate diagnosis, whereas the DSM-I did not. Finally, the DSM-II added three qualifying phrases to the already existing four phrases in the DSM-I. The qualifying phrases refer to phrases such as “acute,” “chronic,” “mild,” “moderate,” and “severe” (Spitzer & Wilson, 1968).

Despite the changes in the DSM-II there were still much needed improvements. The first two editions of the DSM reflected the dominant theoretical orientation at the time, which combined Freud’s theories of personality and Meyer’s psychodynamic theories (Grob, 1991). Another major problem with the DSM-I and II were the descriptions of mental disorders, which were in need of elaboration and operational definitions. Diagnosis was not reliable among clinicians and varied greatly from clinician to clinician. Scull (1989) reported that some research demonstrated that for even the most severe cases, reliability among clinicians for diagnostic categories

remained poor, with under 50% inter-diagnostician agreement (Gold, 2002). One example of the very low reliability was a study by Spitzer and Fleiss (1974). They reanalyzed the reliability of psychiatric diagnosis before the operational approach was used. Kappa coefficients ranged from .26 (neurotic depression) to .77 (organic brain syndrome). Although .77 is considered good, most other disorders did not reach above a kappa coefficient of .55, suggesting major problems with diagnostic reliability.

In the 1960s and 1970s there were four main issues regarding the conceptualization and categorization of psychopathology that were debated among mental health professionals. First, many said that psychiatry was not a legitimate part of medicine. Szasz and his followers argued that without some evidence of a biological or physiological abnormality, using the medical model to interpret mental illness was just a way to control the deviant rather than truly practice medicine (Klerman, 1984). Second, as mentioned above, there was low reliability among psychiatrists with respect to diagnosis. Third, Karl Menninger argued that there were social and psychological consequences to labeling a patient with a psychiatric diagnosis. One of the most influential studies regarding this criticism was conducted by Rosenhan (1973), "On Being Sane in Insane Places," described below. Finally, psychologists such as Lorr, Overall, and Eysenck believed in a dimensional approach to diagnosis and they disagreed with the categorical nature of the official classification system (Klerman, 1986).

From the 1940s to the 1970s there was relatively little attention paid to diagnosis and classification. This was in part due to the lack of connection between treatment decisions and diagnosis. This was mostly true when psychoanalytic therapy was the primary treatment available to clients, as formal diagnosis and classification were not part

of the psychoanalytic assessment strategy. Because this was most often the treatment of choice, it was not necessary to diagnose clients. By the end of the 1960s, however, there was a growing awareness that without a reliable system for diagnosing psychopathology, there would be little progress made in the field (Klerman, 1986).

Development of the DSM-III

During the 1970's Lehmann argued that there had been a long history of ignoring the need for diagnosis, nosology and classification (Lehmann, 1977). This sparked a new interest in classification and diagnosis. In the 1950's Robins, Guze and others at Washington University had concluded that low reliability was due in large measure to the inability of two clinicians to agree upon identified symptoms and behaviors for a particular disorder prior to meeting with a patient. They began developing a structured interview in an attempt to increase reliability among clinicians. They worked to define the exact criteria for each of the 16 DSM-II disorders. These criteria were published in individual papers and then integrated into one article in a classic 1972 paper by Feighner, et al. (1972) and led the way for improved reliability and empirical tests of validity (Matarazzo, 1983).

By the 1970s the National Institute of Mental Health had recognized that depression likely had a psychobiological component. It was later in the 1970s that Endicott and Spitzer, along with their colleagues at the New York Psychiatric Institute, Robins in St. Louis, and others at Columbia, Harvard, and Iowa medical schools joined together and conducted a federally funded study to investigate the psychobiology of depression. This study allowed them to attempt to improve upon the Feighner criteria (1972) not only for depression but for other disorders as well. Even though Feighner had

improved reliability, there was still a lack of knowledge among practicing clinicians in their understanding of how to elicit the information from a clinical interview (Helzer, 1977; Matarazzo, 1983). The tools they subsequently developed from this study were the manual of Research Diagnostic Criteria, (RDC) and the standardized Schedule of Affective Disorders (SADS) (Matarazzo, 1983).

The RDC is a manual of Research Diagnostic Criteria which provides operational criteria, including both inclusion and exclusion criteria, for the descriptions of psychiatric syndromes. For example, a diagnosis of depression required both the presence of sadness and sleep disturbances, but also a *lack of* symptoms such as delusions or hallucinations. However, after the development of these criteria, the group realized that this may present a problem for clinicians of different orientations. That is, if a psychoanalytic clinician and a nondirective clinician each assess their patients differently, this may lead to different information being elicited from the same client. The RDC criteria required a “yes/no” response to determine whether symptoms were present, and one potential problem was that different clinicians were not adept at interviewing patients in this manner (Matarazzo, 1983). Therefore, the Schedule for Affective Disorders and Schizophrenia was developed and employed as a standardized interview for clinicians to use when assessing their clients. The SADS provides questions, items and operationally defined criteria that rule in and rule out each of the of RDC diagnoses (Matarazzo, 1983). Endicott, Spitzer and Robins, (1978) reported high levels of reliability among clinicians using the RDC and the SADS.

In addition to these advances, there were advances with respect to statistical techniques including the ability to save large data sets and send them electronically.

Another breakthrough in the field was the emergence of pharmacotherapy and the implications that this had for the biological basis of mental illness. These studies quickly demonstrated that standardized interview techniques could provide better reliability among clinicians, as well as more reliable estimates in epidemiological community surveys. However, although reliability was improving, reliability was still far from optimal for some diagnostic categories. For example, in one study of lifetime SADS and RDC diagnosis, the initial vs. the consensus coefficients ranged from .06 for Bipolar 2, to .88 for Bipolar 1. Clearly, while there were improvements, there were still some categories needing improvement in reliability.

The Neo-Kraepelinians argued that diagnostic categories should be based on a description of course and symptoms without attempting to prove the etiology of the disorders, and this was the approach taken by the Washington University group. The group focused on the development of operational criteria in an attempt to keep the field moving forward and to increase reliability in psychological diagnosis that was lacking in the field until the late 1970s (Klerman, 1986). They were most interested in biological and genetic explanations for mental illness and focused on the categorical approach to psychiatric diagnosis. They believed that a focus on syndromes had the potential to uncover the etiology of some disorders; however, a comprehensive understanding of the etiology was not a necessity for reliability among clinicians.

It was Robins, Suze and Winokur who were known as the original Neo-Kraepelinians at Washington University. This group was highly influential and the DSM-III was developed based on many of their ideas. Spitzer describes the steps in the development of the DSM-III in Spitzer and Williams, (1980) and in Spitzer's

introduction to the DSM-III. First, The American Psychiatric Association Task Force on Nomenclature and Statistics was appointed with Spitzer as chairperson. The task force included psychiatrists, epidemiologists, and others as consultants. Additionally, the American Psychological Association appointed three psychologists to the committee. Eventually, 14 advisory committees were asked to support the development of the DSM-III. As the DSM-III began to emerge, there were many field trials taking place between 1976 and 1980, and each of the clinical descriptions of the DSM-II were developed to include operational criteria for the 16 major categories and the 187 specific disorders in the DSM-III. Because many of those researchers were involved in the development of the Feighner criteria, the RDC and the SADS were also involved in the development of the DSM-III, as there are many similarities among them. However, although some of the disorders listed in the DSM-III are very similar to those in the RDC, there are differences as well, including the criteria for disorders such as Schizophrenia and Major Depression (Spitzer & Williams, 1980). Two additional diagnostic interviews were developed after the publication of the DSM-III, known as the Renard Diagnostic Interview and the Diagnostic Interview Schedule, both of which were designed to improve reliability in assessment and diagnosis even with use by lay persons (Helzer, Robbins, Croughan & Welner, 1981; Robins, Helzer, Ratcliff & Seyfried, 1982).

With the recognition of the need for a system of diagnosis, nosology and classification came the Feighner criteria, a compilation of individual papers addressing the operational criteria for the DSM-II disorders. The Feighner criteria improved reliability among clinicians, however the RDC and the SADS further improved reliability and gave clinicians from different theoretical orientations a common ground to start from

when interviewing clients. All of these important developments contributed to the development of the DSM-III. Once the DSM-III was developed, an assessment of its utility was needed.

Utility of the DSM-III

There are several strengths of the DSM-III, including (a) provisions for the recognition of multiple disorders, (b) inclusion of the operational criteria with specific exclusion and inclusion criteria, (c) demonstrable reliability based on field testing, (d) a multiaxial system of diagnosis that allows accommodation of diverse aspects of patients' experiences, and (e) recognition of and provisions for the implicit necessity for further change directed by research based on evidence rather than assertions of competing ideological camps (Klerman et al. 1984). Klerman asserted that two of these strengths have had a major impact on clinical practice. The first is the use of operational criteria and the second is the multiaxial classification system. In response to the arguments about the need for increased reliability, many attempts have been made to understand the origins of mental illness and to increase reliability. The DSM-III encompasses these efforts. The multiaxial classification system addresses a problem that many have felt in the field. That is, the disorder is of scientific interest, but the unit of practice is the patient. This problem has contributed to the debate about the clinical relevance of diagnostic systems. Two arguments have been made regarding this topic. First, clinicians claimed that the diagnostic categories were inadequate for understanding the complex individual patient, and their interpersonal relationships (Frances, et al. 1991). Second, some argued that assigning patients to categories contributed to depersonalization and fostered a gap in the doctor-patient relationship (Klerman, 1986).

However, the five axial system helps to address syndromes, while also addressing individual aspects of the disorder (Klerman, 1986).

Although these strengths are promising, there have also been criticisms of the DSM-III that have led to questions about its reliability and utility. Constructing definitions of the basic terminology is one weakness in any classification system, and is a weakness of the DSM-III. The definition of a “mental disorder” or “disease” is one area in which the DSM has struggled, not only in the DSM-III but those that followed it as well. Rather than attempt to provide a definition for the term “mental disorder,” the DSM-III and those that followed elected to avoid defining the term (Gold, 2002). The DSM-III has also been criticized for lacking objective measures that might serve to establish the presence or absence of a mental disorder (Widiger, et al, 1991; Zarin & Earls, 1993). Wakefield (1992) introduced the notion that mental disorders can be best understood through scientific criteria as well as societal value. He argues that this can be done through the use of what he terms “harmful dysfunction.” He suggests that “harm” is a judgment about the desirability, or lack thereof, of a condition and “dysfunction” is the system’s failure to function as it is intended to by natural selection. He claims that this approach is a more objective way to approach mental disorders. However, others such as Lilienfeld and Marino (1995), have criticized Wakefield’s model for several reasons. First, they argue that there are mental functions that are not direct evolutionary adaptations, but instead are neutral by-products of adaptations. That is, many important mental and physical systems were not likely designed by evolution to perform a particular function. Rather, they are by-products of adaptations that have taken on functions different from their initial functions. They give the example of religion,

political beliefs, music, art, etc. They argue that these abilities are likely not the result of natural selection, but rather are an indirect consequence of natural selection for domain specific capacities, such as verbal and spatial ability. Second, they assert that Wakefield neglects the fact that natural selection almost always results in variability across individuals. They argue that there is usually great variability across different individuals for any given system response.

“...the most evolutionarily adaptive response for each system necessarily differs across individuals. In part, this is almost certainly because biological and psychological systems do not operate in a vacuum: The most adaptive response of a given system frequently depends on the functioning of other systems” (p. 6).

Finally, they argue that many consensual disorders represent evolutionarily adaptive reactions to danger or loss. Wakefield (1992) noted that dysfunction is a failure for a system to adequately perform its intended function. However, Lilienfeld and Marino (1995) argue that there are many reactions of a system to threat or bodily harm that can be seen as adaptive rather than a failure to follow intended functions. They give the example of flu symptoms where a fever, coughing and sneezing are adaptive responses to rid the body of an illness. They state that one may argue that these symptoms typically interfere with the body’s natural functioning, as someone with a flu usually has a decreased appetite. However, the system is responding appropriately given the presence of an illness. They also give the example of a specific phobia or blood phobia. Today, we have fewer reasons to fear the sight of blood than we did thousands of years ago due to medical advancements. Therefore counter to Wakefield’s theory of a dysfunction,

some phobias may be the product of systems that are performing *too* well to the way they were intended.

Other critics questioned how scientific it was to develop a classification system through the use of a committee. Some have argued that the process of consensus and compromise inherent in committees is unscientific and is a disadvantage in developing a classification system (Klerman, 1984). However, Gold (2002) argues that although there are some limitations to developing a classification system based on consensus (e.g., peer influence), there is a need for committee work in the development of science. Because classification systems are developed by humans, the use of committees is not beyond reason. Another common critique of the DSM-III was based on statements that the DSM-III claims to be atheoretical, but that there is no such thing as a truly atheoretical approach to diagnosis (Follette, and Houts, 1996; Rogler, 1997; Wakefield, 1999; Wilson, 1993). This criticism stems from the clear Neo-Kraepelin influence on the DSM-III. Gold (2002) asserts that incorporating a theoretical approach does not necessarily invalidate a classification system, as it is not possible to have an atheoretical classification system. Follette & Houts (1996) assert that the DSM claimed to be atheoretical for two pragmatic reasons. The first was that the task force wanted to avoid explicit statements about an underlying model in an attempt to appeal to the broader mental health community. Second, the psychoanalytic theory that influenced the DSM-II was scientifically inadequate, and rather than attempt to argue for another theory, the choice was made to avoid theory completely. Still another criticism is that the DSM-III has emphasized transient surface phenomena, and has not focused on clinical course and human development. The critics argue that there has been an emphasis on diagnostic

reliability at the expense of diagnostic validity (Klerman, 1984). It has been argued that there is an artificial quality of using methods based on observations to define mental illness, and they assert that this is demonstrated through the increased number of diagnoses of each revision of the DSM (Gold, 2002).

Arguments have also been made regarding the DSM's definition of what constitutes normality and abnormality, and that these judgments are based on the norms of a select group of mostly Caucasian male professionals. In addressing this point, several critics have pointed out the overrepresentation of women in a number of diagnostic categories such as Histrionic and Borderline Personality Disorders. However, although there may be an overrepresentation of women in some diagnostic categories, there are also categories in which men are far more frequently diagnosed than women, such as Antisocial Personality Disorder and substance use disorders such as Alcoholism (Turner & Herson, 1997). In addition, the DSM-III has been criticized for its European cultural focus, disregarding non-European cultures and ignoring the influence of the culture in the presentation of mental disorders. Therefore, many have argued that the DSM-III does not represent the countries or cultures outside of the North America (Fabrega, 1991; Mezzich, et al.1999; Widiger, et al. 1991).

Finally, many of the criticisms of the DSM have originated from the antipsychiatry movement. These critics argue that there are social and psychological consequences of labeling someone as "mentally ill." Szasz argued against the existence of mental illness, especially without some kind of biological or physiological evidence. These critics have suggested that there is a great potential for the social misuse of some

diagnostic categories. The most frequently cited example being that homosexuality was classified as a mental illness in the DSM-II (Gold, 346).

Whereas there have been many criticisms of the DSM-III, it nevertheless signified a great advance in the scientific approach to classification and diagnosis. However, further revisions were still needed.

Evolution of DSM-III to DSM-III-R, and III-R to IV

In 1983 the American Psychiatric Association began working on a revision of the DSM-III. This was done for several reasons. First, data emerged from newer studies showing inconsistencies with some of the DSM-III diagnostic criteria. Second, although there were extensive field trials before the publication of the DSM-III, the data had since shown to be unclear, inconsistent and even contradictory to the interpretations prior to the publication of the DSM-III. Because of this, the diagnostic criteria as well as the systematic descriptions of each disorder were in need of review for consistency, clarity, and conceptual accuracy, and then revision when it was necessary. Finally, the American Psychiatric Association was asked to aid in the development of the ICD-10 and the DSM-IV was expected to coincide with the ICD-10. However with the growing scientific literature on psychopathology, a revision of the DSM-III was needed before expected development of the DSM-IV in the 1990s (American Psychiatric Association, 1980).

The task-force working on the DSM-IV conducted a three stage process that included reviews of the literature, re-analysis of already-collected data sets and extensive issue-focused field trials. The domains considered in making decisions for changes included the clinical utility, reliability, descriptive validity, psychometric performance characteristics of individual criteria and a number of validating variables (American

Psychiatric Association, 1994). When a review of the literature suggested a lack of evidence for the resolution of an issue, two resources were used, namely, data re-analysis and field trials. Twelve field trials included more than 70 sites and evaluated more than 6,000 subjects. They collected information on the reliability and performance characteristics of each criteria set as a whole, as well as of the specific items within each criteria set. The authors also state that “the field trials also helped to bridge the boundary between clinical research and clinical practice by determining how well suggestions for change that are derived from clinical research findings apply in clinical practice” (American Psychiatric Association, 1994, p. xix). With all of the improvements of the DSM-III, some argued that there were new editions of the DSM being published, such as the DSM-IV from the DSM-III-R, without any real gains in knowledge (Gold, 2002). However, because the tenth version of the ICD was published, the DSM task force attempted to resolve the inconsistencies that remained between them.

The revision of the DSM-IV to the DSM-IV-TR was a task that was undertaken in an attempt to bridge the span between the publication of the DSM-IV and the DSM-V. The following goals of the DSM-IV-TR were as follows: (a) correct factual errors identified in the DSM-IV, (b) review the DSM-IV to ensure that all information continued to be up-to-date, (c) make any revisions necessary based on additions to the literature since 1992, (d) make improvements to enhance the educational value of the DSM-IV and (e) to update the ICD-9-CM codes that were changed since the DSM-IV 1996 Coding Update (American Psychiatric Association, 1994).

In evaluating the various revisions of the DSM, it is important to consider the psychometric status of the diagnoses. Specifically, the reliability and validity of the disorders in the DSM are of great importance.

Reliability was assessed during the DSM-IV field trials, and the results revealed a mixed pattern, with excellent inter-diagnostician reliability for some disorders, and only modest reliability for others. Reliability for distinguishing between autism and another pervasive developmental disorder was assessed through pairs of experienced raters. They exhibited excellent agreement ($\kappa = .85$). Reliability for Attention-Deficit/Hyperactivity Disorder for test-retest agreement and agreement with the clinician's validation diagnosis for DSM-IV was .61 for combined type, .64 for predominantly inattentive type, and .51 for predominantly hyperactive-impulsive type. Reliability for Schizophrenia was generally high, with kappa coefficients ranging from .79 to .97. For the Mood Disorders, reliability ranged from .43 to .82 in the past month, and reliability for lifetime mood disorders ranged from .36 to .81. Obsessive-compulsive disorder reliability was somewhat lower, with kappa coefficients ranging from .55 to .88. The inter-rater reliability for a lifetime Post Traumatic Stress Disorder (PTSD) diagnosis was .85, and .86 for current PTSD. The median kappa value for the five sites investigating the reliability of sleep disorders ranged from .30 for all listed diagnoses to .38 for the first diagnosis only. Clearly, there are several disorders for which reliability is excellent; however, there are still others in need of improvement.

Validity is inherently a much more complex issue than reliability, as there are several different types of validity. In addition, validity is more complex because most types of validity are difficult to assess. In addition, validity is not as well established for

many disorders and appears to differ by disorder. For example, the distinction between unipolar Major Depressive Disorder and Bipolar Disorder has more validity than the distinction between Social Anxiety Disorder (generalized subtype) and Avoidant Personality Disorder (Herbert, Hope & Bellack, 1992).

The DSM-IV and DSM-IV-TR represent improvements on the DSM-III based on the research performed especially with respect to the field trials. However, they do not represent a fundamental departure from the basic logic and structure of the DSM-III. Consequently, the problems and criticisms identified with the DSM-III are still relevant with the DSM-IV-TR. Although the use of the DSM has become widespread, there remain important unresolved problems, and scholarly debate has continued with respect to alternative classification schemes.

Alternatives to Traditional Psychodiagnosis

Treatment Utility. The validity of the DSM-III and DSM-IV is judged primarily on the basis of traditional psychometric theory. Some have questioned this approach and have proposed an alternative method to evaluate the quality of diagnosis known as treatment utility (Bisset & Hayes, 1999; Hayes, & O'Brien, 1990; McKnight, Nelson, Hayes, & Jarrett, 1984). Treatment Utility has been proposed as an alternative to traditional psychometrics as a means for assessing the quality of diagnosis (Bissett & Hayes, 1999; Hayes & O'Brien, 1990; McKnight, et al. 1984). Treatment Utility is defined as “the degree to which an assessment method contributes to positive treatment outcomes” (Hayes, Nelson & Jarrett, 1987). An assessment is thought to have treatment utility when positive outcomes occur during the treatment process, which are linked to the method of assessment (Hayes, et al. 1987). Some argue that the DSM-IV currently

has a limited ability to use syndromal classification to lead to etiology, choice of treatment, and treatment response (Bissett & Hayes, 1999). There has been some research to support this view (Kratochwill & McGivern, 1996; Kratochwill & Plunge, 1992). Bissett & Hayes argue that,

Logically, no diagnostic system can make much of a difference at the level of treatment outcome until it produces reliably differential treatment linked to the diagnostic categories. The poor correlation between syndromes and treatment is demonstrated by the application of essentially the same treatment technologies across a broad range of syndromal disorders, such as the application of cognitive therapy to mood (Beck, Rush, Shaw & Emery, 1979) anxiety, (Michelson & Ascher, 1987) and personality (Beck & Freeman, 1990) disorders, with only minimal change (p. 381).

Although, treatment utility is not widely used today, it nevertheless represents a potentially viable alternative to psychometric theory. However, treatment utility is not the only alternative to the current psychometric standards.

Functional Analysis. Some researchers have suggested alternatives to the syndromal classification of the DSM such as a functional analytic approach. Whereas the DSM primarily takes a nomothetic approach to diagnosis, a functional analytic approach takes an idiographic perspective, focusing on the individual client. Nelson-Gray and Farmer, (1999) argue that it is possible for both approaches to be used simultaneously. That is, they suggest that the DSM may serve as a valuable starting point, possibly acting as a guide for a functional analytic approach to diagnosis. However, Bissette and Hayes (1999) argue that this is unlikely without an interactive effect of functional analysis on

the syndromes themselves. They also assert that the DSM system is closed to a more idiographic approach, and therefore suggest that what is needed is a nomothetic, syndromal level of classification that is also functionally derived. They suggest that a true functional analysis would have the ability to re-categorize the current classification system through a functional approach. They argue that this may be the only possible outcome if a functional analytic approach were to be nomothetically successful. “A large set of functional analyses may reveal common functional processes. This could lead to a nomothetic level of analysis that is also functionally derived- a classification system based on functional categories” (Bissett & Hayes, 1999, p. 381). However, as they acknowledge, the DSM’s syndromal approach is not based on a functional approach and may be resistant to reorganizing the existing categories into functional categories based on a behavioral analysis. In addition, there is very little research on inter-rater reliability of functional analysis, making this a difficult alternative to the traditional psychodiagnostic approach.

Expert Systems, Logical Functional Analysis, and the Functional Diagnostic Category Approach. Hayes & Follette (1992) have argued that functional analysis has failed primarily because of its lack of replicability, making it difficult to study in a scientific way, as is so easily done with the current DSM syndromal classification system. They do offer several alternatives including the development of expert systems, logical functional analytic systems, and functional diagnostic categories. In their discussion of expert systems, they argue that when clinicians argue about how to conduct a functional analysis of a case, there is no clear answer as to who is more accurate. They suggest that a therapist could conduct a number of functional analyses under controlled

conditions. Then the kind of principles used, information collected and treatment recommendations made could be specified. Eventually, several algorithms would be identified and when enough are identified, an “expert system” or “a coherent and specified system of decision rules- could be developed that systematizes, at least to some degree, what the clinicians does” (Hayes & Follett, 1992, p. 356). In addressing their second alternative of logical functional analytic systems, Hayes and Follett suggest the use of a system in which there are specified treatment rules depending on the results of each stage of assessment. They give the example of taking 80 clients with social skills problems and randomly assigning them to two conditions. One group could be a logical functional system and the other could be either another logical system, or an expert system. With a series of these studies, they believe that it is possible for stable systems of functional analysis to emerge as an alternative method for evaluating categories. Finally, they present the functional diagnostic category approach.

If one were to perform many individual functional analyses from a particular point of view, a variety of patterns might emerge. By arranging these into functional categories with assessment methods and treatment recommendations related to them, a kind of diagnostic system would result. The guiding principle behind the categorization would not be simply agreement, but utility. Many different and competing functional diagnostic systems could result, but the treatment utility of these systems would allow competitive systems to be compared. Such functional categories would provide a general framework that could complement the individual application of classic functional analyses (p. 358).

Functional analysis is still debated as a viable alternative to the current syndromal classification system. Some argue that functional analysis is the clear alternative, whereas others argue that it is impossible to study functional analysis scientifically, as it is not possible to replicate an idiographic behavioral interpretation. Hayes and Follette (1992) offer potentially viable alternatives to traditional diagnostic categorization system based on the general idea of functional analysis and utilizing treatment utility as the core evaluative tool.

Narrative Approach. Goncalves, Machado, Korman and Angus (2002) argue that a narrative approach to diagnosis should be considered. Their argument stems from the results of the classic Rosenhan study (1973). In this study, eight pseudo-patients checked themselves into a mental hospital complaining of schizophrenic like symptoms. Specifically, they complained of hearing voices. They then behaved as they normally would and did not present with any additional psychotic symptoms. Rosenhan found that most of the pseudo-patients' actions were interpreted by the psychiatrists as being abnormal. Goncalves et al. argue that this is one example of how psychopathology is a system of meanings that expresses itself through the organization of language such as narratives, or stories that people tell to create consistency and meaning. Although they find that narratives are of great interest by themselves as an indicator of how clients organize their experience, they believe that aspects of the narrative can provide important information on client problems. The staff in the Rosenhan study never attempted to talk with the pseudo-patients, but rather they observed their behaviors from afar and interpreted them to fit into their original hypothesis. Goncalves et al. argue that narratives offer more information and provide a more sophisticated insight into the

problems a patient is dealing with. They have developed a coding system for narratives and assert that diagnosis should involve analysis of a narrative in order to choose an appropriate treatment.

In summary, the need for revisions of the DSM-III stemmed from studies showing inconsistencies in the studies conducted during the course of the development of the DSM-III. The criteria were reviewed for consistency, clarity and conceptual accuracy. The DSM-IV revisions were based primarily on data reanalysis and field trials. Finally, the revisions of the DSM-IV for the DSM-IV-TR were made in an attempt to bridge the span between the DSM-IV and the DSM-V. The DSM-IV field trials have shown that reliability is good for most of the categories; however, improvement can still be made. Validity is more complex because most types of validity are difficult to assess. In addition, validity is not as well established for many disorders and appears to differ by disorder. Finally, several alternatives have been proposed to either replace the existing classification system, or improve the evaluation of the system. These include (a) focus on treatment utility rather than reliability and validity (Hayes & O'Brien, 1990; McKnight, Nelson, Hayes & Jarrett, 1984), (b) a functional analytic approach (Nelson-Gray & Farmer, 1999), (c) experts systems, (d) logical functional analysis, (e) functional diagnostic category approach (Hayes & Follette, 1992), and (f) a narrative approach (Goncalves, Machado, Korman & Angus, 2002).

Whereas the alternatives proposed to replace or improve the DSM-IV, there are problems with each alternative. First, treatment utility as a way to assess the quality of the current diagnostic system remains a viable alternative. However, it has been argued that using the same treatment technologies for a broad range of syndromal disorders is a

weakness with the current system. This is not necessarily the case as the application of the cognitive model to a broad range of disorders has proven to be highly effective. With respect to a functional analytic approach, one problem is that there may be several analyses for the same case and there is no way to determine a clear answer when two analyses conflict. Second, a complete reorganization would be needed for the existing categories of the DSM into functional categories. Finally, idiographic behavioral interpretations will likely prove difficult to replicate. The expert systems, logical functional analysis and functional diagnostic categories also hold similar problems, such as the need for a complete reorganization of the DSM. In addition, these approaches are extremely cumbersome to conduct. Problems with the narrative approach include factors such as a lack of consistency. Narratives may also lack the necessary relevant information for clinicians to make a diagnosis, information that can be obtained with the use of a structured clinical interview. However, narratives may be useful as a supplement to the structured clinical interview, and likely, patients construct narratives on their own within the clinical setting just by explaining their presenting problems. Therefore, although these alternatives may improve the current diagnostic system in the future, for now they are in need of revision before they will likely impact on future editions of the DSM.

Factors Impacting the Accuracy of Psychiatric Diagnosis

Despite whatever diagnostic system is chosen, research has identified a variety of factors that impact the accuracy of the diagnostic process. Several factors have been identified that impact the accuracy of psychiatric diagnosis. Some of these include overconfidence (Oskamp, 1965), reliance on memory (Arkes, 1986), lack of use of a

structured interview (Fals-Stewart & William, 1994), lack of a multi-modal assessment (Cumpsty, 1996), confusing medical symptoms with psychiatric symptoms (Scheinbaum, 1979), and several information processing biases and heuristics (Ashcraft, 2002; Friedlander & Phillips, 1984; Silverman, 1992).

Disclosure within the therapeutic context may vary greatly from one client to another. The amount of information that a client discloses during his or her first session may have a significant impact on the accuracy of diagnosis. New clients may not feel comfortable disclosing very personal information prior to developing a rapport with their therapist, and a therapist will likely not make an accurate diagnosis if important relevant information is missing. However, although a lack of information would clearly impact the accuracy of diagnosis, there is some research that asserts that as more information is available to clinicians, their diagnostic accuracy does not improve (Gambara & Leon, 1996). Gambara and Leon (1996) found that confidence increased as clinicians receive more information. However, at some point in the information gathering process, the accuracy of diagnosis did not improve, regardless of the amount of information (Oskamp, 1965). Oskamp (1965) also found that overconfidence in clinical judgments may impact the accuracy of diagnosis. That is, if clinicians were not overly confident, they may take a more careful look at the information presented to them without assuming that their diagnosis is correct. Arkes (1986) found similar results, suggesting that lack of awareness, overconfidence, and reliance on memory instead of progress notes add to errors in diagnostic judgments.

Fals-Stewart and William (1994) argue that at times, errors in diagnosis result from not using a structured diagnostic interview. They recognize that although it is time

consuming to use a structured interview, accuracy drops significantly when clinicians do not use them. In addition, Spitzer (1983) found that clinicians were far less accurate in assessing clients without a structured interview than assessing with a structured interview. He argued that that the accuracy was so poor for clinicians without a structured interview and so great for lay people using a structured interview that clinicians may need to show that advances in technology have not rendered them superfluous in the task of clinical assessment.

Cumpsty (1996) conducted a study assessing supervisor's perception of supervisee burnout. She found that the supervisors' accuracy of diagnosis was low to moderate. She suggests that the use of a multi-modal assessment would likely improve the accuracy of diagnosis and multi-modal assessments would likely improve the accuracy of diagnosis in many settings.

Elstein, Shulman and Sprafka (1978) found that clinicians' accuracy was affected when they made diagnoses very early in the assessment process. These authors found that the most accurate clinicians were those who waited until later in the assessment process to make their final diagnosis, suggesting that perhaps clinicians should defer their diagnostic judgments until they have sufficient information from their clients.

Childs, Mercer and Childs (1993) assessed the accuracy of diagnosis in patients referred for inpatient brain injury rehabilitation. They found that errors in diagnosis may result from confusion in terminology, lack of extended observation of clients and lack of skill or training in assessment. Although this study was conducted on the accuracy of neurological assessment, the specific factors highlighted, including lack of training in assessment and confusion with terminology apply to the general clinical setting as well.

Scheinbaum (1979) investigated the effects of medical disorders such as hypoglycemia and delirium that produce psychiatric symptoms and are often confused with mental illness. Due to the rates of misdiagnosis and the costs of prolonged psychiatric hospitalization, the author suggests that more sophisticated diagnostic measures are needed. Some of these include (a) obtaining a detailed history, (b) a careful review of the symptoms checklist, (c) a detailed physical examination, (d) biochemical evaluations, (e) routine urinalysis, (f) urine drug screen, and (g) a complete blood count.

Finally, many studies have investigated the impact of cognitive biases on the accuracy of decision making (Ashcraft, 2002; Friedlander & Phillips, 1984; Silverman, 1992). These biases and their effects on decision making are discussed below.

Many factors contribute to the inaccuracy of diagnosis. Some of these include overconfidence (Oskamp, 1965), reliance on memory (Arkes, 1986), lack of use of a structured interview (Fals-Stewart & William, 1994), lack of a multi-modal assessment (Cumpsty, 1996), confusing medical symptoms with psychiatric symptoms (Scheinbaum, 1979), and various cognitive biases and heuristics (Ashcraft, 2002; Friedlander & Phillips, 1984; Silverman, 1992). Some of these factors have relatively obvious remedies, such as using a structured interview and a multi-modal assessment. Others are more difficult to address, such as overconfidence in clinical and diagnostic skills. This is more indicative of a cognitive bias that may require a more sophisticated approach to improve accuracy in diagnosis.

Cognitive Biases and Heuristics

Cognitive biases and heuristics can significantly impact a clinician's ability to make objective decisions. As defined by Garb (2003), "biases are beliefs or

preconceptions that adversely influence clinicians' interpretations of available data" (p. 24). Heuristics can be thought of as shortcuts that describe how clinicians make judgments, treatment decisions, or both (Garb, 2003). Generally, biases stem from the use of heuristics. That is, biases exist within the context of cognitive short-cuts. This is because anything that leads to storage of information in memory can influence reasoning because judgments are based on recalled information (Ashcraft, 2002).

Examples of common biases include (a) the confirmatory bias, (b) the overconfidence bias, (c) the hindsight bias, (d) the race bias, (e) the gender bias, (f) the familiarity bias, (g) the anchoring or primacy bias, and (h) the vividness bias. The confirmatory bias occurs when individuals knowingly or unknowingly review information in a way that they search out and highlight information that is consistent with an initial hypothesis or judgment, and ignore or de-emphasize inconsistent information. Overconfidence occurs when individuals are more confident in their hypotheses than is warranted based on the amount or quality of information they have received. The hindsight bias refers to the tendency for knowledge of a particular outcome event to influence the perceived likelihood of that event, leading to incorrect explanations for the occurrence of the event. Essentially, this describes the mental processes that occur when individuals create explanations for events that have already occurred. Race and gender biases occur when judgments are considered more valid for one racial or gender group than another. Familiarity bias refers to the notion that familiar information such as famous names are more easily recalled (Ashcraft, 2002). The anchoring or primacy bias occurs when people are exposed to identical information, but in varying order, resulting in significantly different judgments. For example, a client's diagnosis may be related to

the time a clinician is exposed to pathognomonic information. A different diagnosis may result from presenting information earlier in treatment versus later. Finally, Vividness biases occur when the information stored in memory is more vivid than other information. For example, even though statistically air travel is safer than car travel, the report of an airplane crash is more vivid, thereby leading to greater fears about air travel relative to car travel (Ashcraft, 2002). Although other biases have been identified, these represent some of the most common examples of cognitive short-cut processes that influence decision making.

There is some evidence to suggest that the use of heuristics can be functional by increasing the efficiency of decision making, especially compared to the use of an algorithm (Beckman & Kuhl, 1984). “An algorithm is a specific rule or solution procedure, often detailed and complex, that is guaranteed to furnish the correct answer if it is followed correctly” (Ashcraft, 2002, p. 464). However, whereas an algorithm is theoretically error free, it is often not the most efficient approach to problem solving, whereas heuristics tend to be much more efficient. Nevertheless, cognitive heuristics are fallible, and as noted by Garb, (2003) they can lead clinicians to fail to learn from their clinical experience. Heuristics allow people to make judgments efficiently and at times accurately, however, they also have the potential to lead to errors in judgment (Garb, 2003). Whereas a heuristic is a cognitive short-cut and may or may not lead to error in decision making, biases often stem from the use of these cognitive short-cuts, which then lead to errors.

As mentioned above, biases are often specific types of heuristics. One of the most common examples is the availability heuristic. This refers to the tendency to be

influenced by the ease with which individuals remember situations or aspects of situations. Biases that can potentially stem from the availability heuristic are biases in general world knowledge, familiarity biases, and salience and vividness biases. Other heuristics identified by Ashcraft (2002) include the representativeness heuristic and the simulation heuristic. “The representativeness heuristic is a judgment rule in which an estimate of the probability of an event is determined by one of two features: how similar the event is to the population of events it came from or whether the event seems similar to the process that produced it” (Ashcraft, 2002, p.468). One bias that may stem from this heuristic is a stereotype. Another is the gambler’s fallacy (thinking that the next toss of a coin is more likely to be either heads or tails based on the previous tosses, when really the previous tosses have no effect on the next). Finally, the simulation heuristic refers to a “mental construction or imagining of outcomes, a forecasting of how some event will turn out or how it might have turned out under another set of circumstances” (Ashcraft, 2002, p. 475). The most common bias that stems from this heuristic is the hindsight bias described above.

Although there are many biases and heuristics, Nickerson (1998) wrote that “if one were to attempt to identify a single problematic aspect of human reasoning that deserves attention above all others, the confirmation bias would have to be among the candidates for consideration” (p. 175). He asserts that the confirmation bias appears to be pervasive enough that it may by itself account for a significant number of the disputes and misunderstanding that occur among individuals, groups, and even nations.

The Confirmation Bias

The confirmation bias has been studied in a wide array of contexts and situations (Frey & Thelen, 2001; Garb, 2003; Huebner, 1990; Loslowksi & Maqueda, 1993; Nickerson, 1998; Snyder & Campbell, 1980). As Nickerson (1998) defines it, the confirmation bias, as the term is typically used in the psychological literature, connotes the seeking or interpreting of evidence in ways that are partial to existing beliefs, expectations, or a hypothesis in hand. Whereas the confirmation bias is studied in a wide variety of contexts, it is most often studied by social psychologists.

Wason (1960) developed a task that asks subjects to discover a rule that determines which of a series of three-number strings were correct and which were incorrect (Gorman & Gorman, 1984). Subjects started with the number string '2-4-6' and were informed that this string was "correct." They were then asked to generate different number strings. After each string, they were told whether it was correct or incorrect. When they believed that they had discovered the rule, they would announce their hypothesis. For example, "numbers increasing in order of magnitude." After hearing many of his subject's responses, Wason discovered that many of his subjects displayed a strong confirmation bias and neglected to use disconfirmatory testing strategies. Replicating his findings were Mahoney & DeMonbreun (1978) who found that when testing psychologists with this task, all groups showed a tendency to confirm their hypotheses rather than disconfirm. Based on these studies using Wason's task (1960), subjects are much more likely to use confirmatory strategies than disconfirmatory strategies unless told to do otherwise.

In one famous study, Snyder Tanke and Berscheid (1977) investigated the confirmation bias in social interactions and assessed the belief that physically attractive people have more appealing personalities than less physically attractive people. College age men were given a picture of their assigned partner before talking with them over the phone. The men who believed that their partner was physically attractive had expectations that they would be socially adept, and therefore conversed in a more warm and friendly way than men who believed that their partner was unattractive. The women then responded in a similar manner (whether they were physically attractive or not). In addition, women who were thought to be less attractive were thought to be aloof during the phone conversation.

In another study by Snyder and Swann, (1978), the confirmation bias was investigated with respect to testing hypotheses about others. Each student was given a “target person” and told that this person was either introverted or extroverted. They were asked to test this idea by selecting questions to ask their target person. Results showed that the students tested their hypotheses in a confirmatory way by asking questions that would elicit either extroverted or introverted behavior. That is, they acted as if their hypotheses were truth, rather than as an idea to be tested. In summary, these results suggest that the confirmation bias can alter social interactions as well as influence the testing of hypotheses about others.

In another study by Fugelsang, Stein, Green and Dunbar, (2004), the confirmation bias was evidenced in an observation study. Three leading molecular biology laboratories at a prominent U.S. university were analyzed. One researcher interviewed the scientists and attended the laboratory meetings which were audiotaped and

transcribed. They were then coded along a number of dimensions, including the types of interactions between speakers such as clarification, agreement and elaboration, disagreement, and questioning. First, the researchers found that participants discounted data that was inconsistent with their initial theory as not being “real” however, over time, they found that participants altered their approach and began to consider information that was *inconsistent* with their initial hypothesis and to modify that hypothesis to account for the new information. This study provides further support for the existence of the confirmation bias, although, it extends a step further to include the notion that the confirmatory strategies that are often used initially may dissipate with time and repeated experience when the data does not fit the initial hypothesis.

This phenomenon not only occurs in social or laboratory contexts, however, and awareness of the confirmation bias is of particular importance within the clinical context. Every time clinicians interact with their clients they have the opportunity to gather information and form new or different impressions. Clinicians often form these impressions through several means. Clinicians may form impressions based on the verbal content of their clients’ comments, the paralinguistic features such as tone of voice or voice volume, or their clients’ nonverbal behaviors such as body posture, facial expression, etc. In addition, these observations are often made simultaneously and very quickly (Elstein, Shulman & Sprafka, 1978; Guaron & Dickinson, 1969; Sandifer, Hordern & Green, 1970). As an example, the impression formation literature in social psychology attends to how individuals get to know others and how people often make inferences about the characteristics or traits of people in a brief amount of time. Whereas this may be a useful strategy as it allows people to get beyond the specific behavior of an

individual and generalize to similar events in the future (Van Overwalle & Labiouse, 2004), it may also lead clinicians to form inaccurate and premature impressions. This can be problematic because once an initial diagnostic impression is formed, it can be very difficult to alter this impression. (Elstein, et al., 1978; Guaron & Dickinson, 1969; Oskamp, 1965; Sandifer, et al., 1970). However, clients typically do not seek treatment for one session only and they may be exhibiting behaviors over subsequent sessions that are inconsistent with the clinician's initial diagnosis. If the clinician is influenced by the confirmation bias, he or she may search for information consistent with his or her original impression, and disregard subsequent inconsistent information.

Eva (2003) also found that participants in a medical diagnostic study were prone to the confirmation bias and results revealed that diagnosticians did not seriously consider alternative diagnoses when they were not explicitly asked to evaluate them. Even when subjects were given direction to "consider alternatives," they continued to focus on their initial hypothesis, despite evidence to the contrary.

Another study by Jonas, Schulz-Hardt and Frey (2001) supports this idea. These researchers found that the way in which information is presented greatly influenced whether the participants demonstrated the confirmation bias. That is, they found that participants are more prone to the confirmation bias when information was presented and processed sequentially instead of simultaneously. The authors hypothesize that this increase in the confirmation bias was due to the heightened commitment that participants made to their initial hypotheses when the information was provided sequentially. This study provides useful information as to how the confirmation bias may unfold, and may

hold important implications for gathering as much data as possible and reserving decision-making until all of the data is collected.

In addition to the short amount of time it takes clinicians to make their initial judgments of clients, another problem is the amount of information that clients reveal during the first session. There is a great deal of information for a client to reveal to their therapist and often, it is not possible for the client to reveal all of the information in the first session due to lack of time. In addition, the client may want to develop a strong rapport with the therapist before they will reveal highly personal information. This alone is not necessarily problematic. However, it does become problematic when a therapist makes an initial impression and then disregards inconsistent information in later sessions. Clearly, the confirmation bias should be of concern to clinicians, as accuracy in diagnosis is an essential first step in treatment.

The Confirmation Bias and Psychiatric Diagnosis

Some research shows that when a client is assessed, diagnostic impressions are typically made very quickly and it is often difficult to set aside these initial impressions to allow for new and or inconsistent information to be considered (Guaron & Dickenson, 1969; Houts & Galante, 1985; Langer & Abelson, 1974; Sandifer, et al., 1970; Temerlin, 1968). Herbert, Nelson, and Herbert, (1988) examined how a diagnosis assigned to a client affects the accuracy of the assessment of normal and abnormal behaviors that were either consistent or inconsistent with the initial diagnosis. Practicing clinical psychologists were randomly assigned to two groups, one of which was told that the clients they were about to see had been diagnosed with Dysthymic Disorder at a local clinic, and the other of which was not provided a diagnostic label. The psychologists

then watched 12-minute videotapes of three clients. Each psychologist observed a client displaying “normal” behavior, another displaying symptoms characteristic of Dysthymic Disorder, and a third displaying symptoms characteristic of a Major Depressive Episode. The results revealed that the assessment of normal behaviors exhibited by clients was impaired by the label they received. In other words, when given a diagnosis of “Dysthymia,” psychologists were more likely to diagnose a normal “client” with Dysthymia than when no such label was given. This is but one example of how the confirmation bias can affect clinical judgment.

One of the most famous studies relating to the clinical effects of the confirmation bias was conducted by Rosenhan and his colleagues in 1973. Eight pseudo-patients checked themselves into a mental hospital complaining of schizophrenic like symptoms, specifically, they complained of hearing voices. They then behaved as they normally would and did not present with any additional psychotic symptoms. However, what they found was that most of their actions were interpreted as being abnormal. The effect of the initial judgment of pathology here is powerful as the normal behaviors exhibited subsequently were either ignored or reinterpreted to be those of someone who was mentally ill, reflecting the confirmation bias.

Langer and Abelson (1974) studied the effects of labels on clinicians’ judgments. Both behavior and psychoanalytic therapists viewed a videotaped interview between a man who had just applied for a job and one of the authors. Half of each group was told that the interviewee was a job applicant, and the other half was told that he was a patient. All clinicians were then asked to complete a questionnaire to evaluate the interview. The behavior therapists described the interviewee as “well adjusted” regardless of the label

they were given. However, the psychoanalytic therapists judged the interviewee as more disturbed when they were told that he was a patient rather than a job applicant. This is another example of the confirmation bias, as the interviewee's "normal" behaviors were reinterpreted to fit the initial idea that he was a patient. The authors offer several interpretations. There is the possibility that the interviewee did have deep underlying problems about which the behavior therapists were unaware of; however, because the interviewee was able to cope with his environment and was not really a patient, the authors were not satisfied with this interpretation. They also suggest that it is possible that the behavior therapists were so focused on the behaviors of the interviewee that the label did not affect them. The authors believe that the most likely explanation is that the behavior therapists noted the label and actively tried to discount its relevance because training from a cognitive perspective encourages such discounting.

In yet another labeling study, Temerlin (1968) studied the suggestion effect in psychiatric diagnosis. He investigated the interpersonal influences that may influence diagnostic decision making. Psychiatrists, clinical psychologists, and graduate students in clinical psychology listened to an audiotaped interview of a "normal, healthy man." They then offered a diagnosis. Prior to listening to the interview, each subject heard a highly prestigious clinician say that the individual was "a very interesting man because he looked neurotic but was actually quite psychotic." They found that the subjects' diagnoses were influenced by the presence of the label "psychotic". No subjects in the control condition diagnosed the individual with psychosis, whereas a diagnosis of psychosis was made by 60 percent of the psychiatrists. Temerlin's work indicates that the confirmation bias affected the subjects' impressions and potentially altered their

interpretation of his behaviors, leading them to interpret his “normal, healthy” behaviors as psychotic.

Making the process of diagnosis even more complicated is the work of Arkes and Harkness (1980), who found that the very act of making a diagnosis changes the way that a clinician remembers the case. They found that a clinician is more likely to remember a symptom that is consistent with their diagnosis, even if the client did not present with that symptom.

Clearly, the assessments made by these clinicians were influenced by the presence of a label. Not only did the label “psychotic” influence the clinicians, but merely the understanding that an interviewee was a “patient” influenced the clinicians to interpret normal behaviors as abnormal. Again, this is of concern for clinicians who may be interpreting the behavior of their clients in an abnormal way, without objectively considering their behaviors. Whereas these studies focus primarily on labeling, they relate to the confirmation bias in that the confirmation bias may stem from the presence of a label. As an example, if a clinician assessing a patient finds that a previous therapist labeled the patient as depressed, the new therapist may be affected by the presence of the label and form hypotheses consistent with that label. He or she may then judge the patient’s behaviors to be consistent with someone who is depressed and may also disregard behaviors that are inconsistent with the label and miss important diagnostic information. Whereas these studies have indicated the importance of the confirmation bias with respect to psychodiagnosis, the following studies relate to the confirmation bias in more general clinical contexts.

The Confirmation Bias in the Clinical Context

Although the previously mentioned studies are of importance to clinicians, the focus of these studies was on the labeling effects of a diagnosis, and the subsequent confirmation bias that may result from a label. However, there are several studies investigating the confirmation bias more directly within the clinical context. Haverkamp, (1993) assessed the responses of 65 counseling trainees after they viewed a videotape of a therapeutic interaction. The hypothesis testing strategies of the trainees were assessed using hypotheses provided by the client. The hypothesis testing strategies were also assessed for those hypotheses generated by the counselor. The trainees were asked to generate a list of questions they would like to ask the client. In assessing these questions, the results suggested that their style of hypothesis testing was confirmatory. Questions were confirmatory 64% of the time, (meaning that their questions focused on confirming their hypotheses, rather than refuting them), neutral 21% of the time, and disconfirmatory 15% of the time. This strongly suggests the presence of the confirmation bias with respect to hypothesis testing. Similarly, Snyder and Swann, (1978) found that students tested their hypotheses in a confirmatory manner. Students were told that a “target person” was either introverted or extroverted. The students then asked questions of this target person, and results showed that the students tested their hypotheses in a confirmatory manner by asking questions that would elicit either an extroverted or introverted answer.

Whereas the confirmation bias has been shown to affect the style of hypothesis testing, Guaron and Dickenson (1969) studied the confirmation bias in relation to the amount of time it takes clinicians to make a diagnosis. They conducted a study in which

psychiatrists were asked to watch videotaped interviews and to give a diagnosis based on their observations. Results indicated that many of the psychiatrists made a diagnosis within 30-60 seconds of watching these videotapes. Additionally, after the initial diagnosis was made, the psychiatrists did not change their diagnostic impression even when contradictory information was given. These psychiatrists were told that they would first watch a videotape of a patient and then would complete a Patient Evaluation Form, following which they would be asked to make a diagnosis. At the time that the psychiatrists made a diagnosis, they were asked to provide ratings of the likelihood of the diagnosis and their satisfaction with the diagnosis they provided. After this, they were asked to report what items of information were most important in making the diagnosis. The examiner then explained that information pertaining to the patient had since been obtained and the examiner read aloud to the subject that piece of information from the case history categories that most closely approximated the informational item rated of greatest importance. This continued until the entire list had been covered. The subjects were then asked to make a final diagnosis. Results showed that the psychiatrists did not actively seek out disconfirming information, which likely led them not to consider changing their initial diagnoses, suggestive of the confirmation bias. The psychiatrists either ignored the disconfirming information, or reinterpreted the new information to fit with their initial diagnostic impression.

Similar to the results of Guaron and Dickinson (1969), Sandifer, Hordern and Green (1970) found that clinicians make judgments and diagnoses very quickly and may not consider additional information. They investigated diagnostic decision making by having psychiatrists watch films of diagnostic interviews. They then asked them to

record their observations while watching the film. They found that the first three minutes of observation have a significant and sometimes decisive impact on the final diagnosis. This again suggests the presence of the confirmation bias because clinicians in this study made diagnostic decisions very quickly and appeared to disregard additional information once their initial impression was formed.

These results are consistent with the work of Elstein, Shulman and Sprafka (1978), who found that clinicians formed diagnostic impressions very early in the assessment process. They also found that the most accurate clinicians arrived at a final diagnosis later than those who were less accurate. This research further supports the notion that the confirmation bias is not only present within the clinical context, but also suggests that it affects the accuracy of diagnostic impressions.

In 1965, Oskamp conducted a study investigating the effects of the confirmatory bias and overconfidence in clinical decision making. The relationship among incremental data gathering, validity of personality judgments and the confidence in the accuracy of these judgments was assessed. As additional information was presented to the clinicians, confidence in their accuracy of the personality judgments increased. Interestingly, accuracy remained at the same level while confidence continued to rise. One possible explanation is that overconfidence causes clinicians to focus on information that confirms their hypotheses and to ignore or reinterpret disconfirming evidence.

Another study investigating the confirmation bias and clinician confidence within the therapy session found similar results. Martin (2001) conducted a study in which mock therapy clients participated in a clinical interview. The clinicians were asked for their diagnosis and confidence rating, in addition to listing questions that they wished to

ask to clarify their hypothesis. Over time, participants became more confirmatory and less disconfirmatory in their questioning of the pseudo-client. In addition, those who used more disconfirmatory questioning strategies were less confident in their diagnoses than were those who used confirmatory strategies.

This overconfidence can be especially important with regard to clinical versus statistical prediction research. In the process of making clinical judgments, therapists tend to ignore base rate information and assume that their clinical judgment is more accurate than statistical predictions. However, this is usually not the case. Paul Meehl (1957) asserts that the evidence for the superiority of actuarial prediction is far too convincing to ignore. Studies investigating clinical vs. statistical prediction on everything from medical diagnosis to sports suggests that neither is highly reliable, however clinical prediction is far less reliable (Meehl, 1957; Stanovich, 2001). This inaccuracy in clinical prediction is potentially hazardous in the clinical setting. If clinicians are affected by the confirmation bias and their overconfidence leads them to disregard inconsistent information such as base rate information, this makes it even less likely that a clinician will consider alternative diagnoses. In other words, if clinicians are overconfident and ignore base rate information, assuming that their clinical judgment is more accurate than statistical predictions, the effects of any confirmation bias could be magnified.

The studies discussed here suggest that the confirmation bias affects several aspects of the therapeutic process. Haverkamp (1993) found that the style of hypothesis testing was confirmatory most of the time and only disconfirmatory 15% of the time. The questions asked of clients were not the only confirmatory practice, however. Gauron

& Dickenson (1969), Elstein et al. (1978), and Sandifer, et al. (1970) found that diagnostic impressions are made very early on during the clinical interview, sometimes within the first 30 seconds, and additional information was subsequently ignored. Finally, Oskamp (1965) reports that overconfidence may cause clinicians to focus on information that confirms their hypotheses and either to ignore or to reinterpret information that might disconfirm their hypotheses. Clearly, the confirmation bias affects several aspects of the diagnostic decision making process.

Real-World Importance of Addressing the Confirmation Bias

The studies discussed above are alarming, and illustrate the need for the development of assessment practices designed to counteract powerful cognitive biases that impact clinical decision making. These studies are clear examples of how biases and heuristics, especially the confirmation bias, apply to the clinical realm. One example of how biases impact real world decision making is a study of Schizophrenia among African Americans. Research has shown that clinicians are more likely to assign a diagnosis of Schizophrenia to an African American client than to a Caucasian client (Neighbors, Trierweiler, Ford & Muroff, 2003; Trierweiler, Neighbors, Munday, Thompson, Binion & Gomez, 2000). There are other explanations for the difference in diagnostic rates, including differences in presentation across cultures, and the possibility that there really are differences in the rates of psychiatric disorders between races (Trierweiler et al., 2000). However, biases, such as the race bias, have been found in addition to these explanations. Clearly, if clinicians are expecting that African Americans have a higher rate of Schizophrenia and they look for consistent symptoms, this is another example of how the confirmation bias can potentially affect clients in a negative way.

The confirmation bias is important for clinicians to be aware of for many reasons. Consider, for example, the scenario in which a therapist believes that symptoms such as feelings of depression, shame, worthlessness, perfectionism, and powerlessness are suggestive of previous sexual abuse. Authors such as Ellen Bass and Laura Davis, authors of the highly popular *Courage to Heal* (1988), suggest that survivors of sexual abuse are highly likely to experience such symptoms. Therefore, when a client reports these feelings, and her therapist suggests that she may have been sexually abused, this is another subtler example of the confirmation bias. Loftus (1998) and McNally (2001) have found that even if a patient cannot remember any abuse, some suggest that it nevertheless occurred, and they she simply repressed the memory. Through the use of guided imagery, hypnosis, or dream interpretation, therapists have searched for information that confirms their hypotheses. Once the therapist finds some information believed to be consistent with sexual abuse, this only strengthens his or her hypothesis, and as previous studies have shown (Gauron & Dickinson, 1969; Haverkamp, 1993), therapists rarely look for disconfirming evidence without being prompted to do so. Within this context, the confirmation bias has led to false memories in clients, while therapists believe they have revealed what were previously repressed memories. The harm done by so-called “recovered memory therapy” includes the severing of relationships with between clients and their fathers, uncles, and other innocent friends and family members. Several researchers including Loftus (1998, 2003) and McNally (2001, 2003) have refuted the research on recovered memories; however, some therapists continue to rely on the confirmation bias and actively search out these “repressed memories.”

The previous studies demonstrate how the confirmation bias has the potential to impact clinical decision making in a negative way. Knowing this, several researchers have attempted to “debias” subjects through a number of strategies. Instead of merely investigating whether biases exist, these researchers use several strategies in an attempt to actively decrease bias.

Strategies for Debiasing

The studies discussed above suggest that biases, especially the confirmation bias, are clearly problematic for clinicians. Therefore, several researchers have attempted to “debias” subjects through a variety of strategies including (a) the use of structured clinical interviews, (b) decreasing overconfidence, (c) delaying decision making, (d) training, (e) considering the opposite, (f) considering alternatives, and (g) education or warning the subject about the bias (Arkes, Christensen, Lai & Blumer, 1987; Eva, 2003; Galinsky & Moskowitz, 2000; Hirt & Markman, 1995; Lord, Pepper & Preston, 1984; Mumma & Steven, 1995; Mynatt, Doherty & Tweney, 1978; Newstead, Pollard, Evans & Allen, 1992; Renner & Renner, 2001; Spengler, Strohmer, Dixon & Shivy, 1995). One of the most common strategies is to educate or warn someone about a bias and suggest they be aware of it while performing their respective tasks. Although this is a common approach, the research is somewhat contradictory with regard to the effectiveness of warning a subject about a potential bias. Some argue that it is worthless (Arkes, 1981), however many others have found that educating or warning subjects is an effective means to combat biases.

Structured Clinical Interviews. Whaley and Geller, (2007, in press) discuss cognitive biases within two models of information processing. The first is cognitive bias

stemming from prototype models of information processing. The other incorporates the structured clinical interview to assess the decision making process. Specifically, they address racial biases and misdiagnoses of African Americans. They suggest that comprehensive training in the administration of the *Structured Clinical Interview for the DSM-IV* (First, Spitzer, Gibbon & Williams, 1996) is a key component in decreasing these biases, although they also acknowledge that accurate diagnosis requires knowledge of variability in symptoms presentation between different cultures and ethnic groups.

Perspective Taking. Galinsky and Moskowitz (2000) investigated perspective taking as a debiasing technique for stereotypes. Subjects were shown a black and white photograph of an older man sitting next to a news stand. Subjects were then asked to write a short essay about a typical day in his life. Before writing the essay, one third of the subjects were assigned to a control condition and received no further instructions. One third were assigned to a suppression condition and told that they should try to avoid thinking about the photograph when writing their essays. The final third were instructed to adopt the perspective of the man and imagine a day in the life of this individual. Subjects then completed a lexical decision task and after this, they were shown a picture of a second elderly man and asked to write a second essay. Finally, subjects were shown a picture of an African American male in his early twenties and asked to write one final essay. Results from the essays showed that perspective taking successfully decreased stereotypic biases, as subjects' essays did not include information about stereotypes. This research holds important implications for clinicians who knowingly or unknowingly are biased by their clients' race. As Trierweiler et al. (2000) and Neighbors et al. (2000) found, the race bias is a very real concern for clinicians.

Computerized Programs. Silverman (1992) investigated whether computers can interact with subjects through modeling and detecting their confirmation biases during a problem solving task. He investigated whether computers can have an effect on decreasing the confirmation bias by offering criticism in response to the subject's use of the confirmation bias. The results of his work suggest that computers likely can decrease the use of the confirmation bias, although the improvement is sensitive to the strategy used. The results suggest that the most successful strategy was the use of an "influencer" that tutors subjects about the confirmation bias with an example before they attempt the task. This is similar to the work on educating subjects before starting a task. These results suggest that offering a warning about the confirmation bias along with an example can be an effective debiasing strategy.

Decreasing Overconfidence. Arkes, Christensen, Lai, and Blumer (1987) attempted to reduce overconfidence using two strategies across two different experiments. In experiment one, half of the subjects answered five practice questions that appeared to be difficult. The remaining subjects answered questions that appeared to be easy, but were just as difficult as the other questions. Half of each group received feedback on the accuracy of their answers, whereas the other half received no feedback. All four groups then answered an additional 30 questions, and rated their confidence in their answers. The group that thought they were answering easy questions and then received feedback became less confident in their answers on the last 30 questions. Subjects in a second experiment anticipated a group discussion of their answers after they completed answering all questions. Not surprisingly, these subjects took a longer amount of time to answer their questions and were less confident than a control group. Koriat,

Lichtenstein and Fischhoff (1980) suggest that one reason for overconfidence is that subjects can more easily generate supporting reasons for their decisions compared to contradictory reasons. Therefore, if the evidence that people are likely to retrieve is supportive of their hypotheses rather than representative of all available evidence, confidence increases. This process again reflects the confirmation bias, as people will seek out confirmatory evidence (thereby increasing confidence) and disregard disconfirmatory evidence.

Koriat, et al. (1980) suggest two strategies for decreasing overconfidence. In one experiment subjects were given two alternative questions and were required to list reasons for and against each of the alternatives before choosing an answer and assessing the likelihood of its being correct. This decreased confidence to a much more appropriate level. In a second experiment, subjects were asked first to choose an answer and then to list (a) one reason supporting that choice, (b) one reason contradicting it, or (c) one reason supporting and one reason contradicting it. In this study, they found that the only group that was able to accurately rate confidence was the group that listed contradicting reasons.

Renner and Renner (2001) assessed whether decreasing overconfidence can be done in a classroom setting, as well as whether this has practical value for students. They found that providing confidence estimates as questions were answered was enough to debias students in a series of quizzes and resulted in better performance on quizzes. In a second experiment, quizzes were taken on a computer. Students were given feedback for each question as it was answered. Again, results indicated that overconfidence decreased and students performed better on the quizzes.

The results of several studies indicate that decreasing overconfidence is an effective means of debiasing subjects. Given the relationship of overconfidence and the confirmation bias, this may also hold important implications for debiasing subjects who are using confirmation bias. That is, decreasing overconfidence may also lead to a reduction in the confirmation bias.

Delayed Decision Making. Spengler, Strohmmer Dixon and Shivy (1995) investigated the effects of delaying judgment on the accuracy of diagnosis. They found that when therapists delay final judgments, they have the time to reflect on and even extend the assessment. In addition, therapists were able to test alternative hypotheses, and subsequently became more open to alternative explanations. In addition, Elstein, Shulman and Sprafka (1978) found that the most accurate clinicians arrived at a final diagnosis later than those who were less accurate. Based on this work, Hill & Ridley (2001) investigated whether counselors put this debiasing technique into practice. That is, do they delay their diagnostic impressions when in a setting that allows them to do so, without pressure from insurance companies to submit a diagnosis right away? Results suggested that delayed diagnostic decision making occurred more frequently than immediate diagnostic decision making. They are unsure if these results can be generalized, due to the small number of participants, and the fact that the counselors were all still in training. Because of the inconsistency with other studies finding that clinicians make judgments very quickly (e.g., Elstein, et al., 1978; Guaron & Dickenson, 1969; Sandifer, Hordern & Green, 1970) it may be that these students were recently trained to avoid a premature diagnosis until they believed they had acquired a sufficient amount of information. Nevertheless, this appears to be a useful debiasing technique for clinicians,

as Spengler, et al., (1995) found that delaying final judgments increased the accuracy of diagnosis, and Hill and Ridley, (2001) found that clinicians were in fact able to implement this technique.

Considering the Opposite. Lord, Pepper, and Preston (1984) investigated the notion that considering the opposite (beliefs, perceptions, outcomes, etc.), would lead subjects to display less bias in social judgment. In two experiments, they found that the introduction of a “consider the opposite” strategy had a stronger corrective effect than giving general debiasing instructions to “be as fair and unbiased as possible.” They believe that this is an effective method for retraining biases in social judgment. These results support the notion that a general statement about avoiding bias does not work, although the general statement they used did not attempt to educate subjects about any one bias in particular or provide examples of when and how biases interfere with decision making.

Mumma and Steven (1995) used the “consider the opposite” strategy as well in an attempt to decrease the primacy effect (also known as the anchoring effect) in clinical judgment. They found that a “consider the opposite” procedure, involving cue sorting by diagnosis, decreased the primacy effect. However, they also found that a “bias inoculation” procedure also reduced the primacy effect. However, the bias inoculation training involved more than simply informing subjects about the bias and warning them against it. This approach involved more detailed and consistent training throughout the task.

Training. The approach used by Mumma and Steven (1995) involved informing subjects about the primacy effect, and why it may occur. They were then given an

example demonstrating how they might adjust for the primacy effect and practiced this several times. Each trial involved a description of a person who made a personality judgment about someone after receiving an initial set and a subsequent set of cues about them. The subject was asked to indicate how the initial judgment should have been adjusted to avoid the anchoring bias. Subjects who gave incorrect answers to a scenario were corrected, and given an explanation for why they were incorrect. The trials were then repeated. Mumma and Steven found this to be marginally successful in decreasing the primacy effect. They suggest incorporating specific or non-specific debiasing strategies into clinical training so that they become a routine component in the process of assessment and diagnosis.

Considering Alternatives. Hirt and Markman (1995) propose that considering *any* plausible alternatives, not just the opposite as suggested by Lord et al. (1984), will also lead to debiased judgments. They recognize that there are instances in which only two alternatives can be considered. However, very often, as with the case of psychodiagnosis, there are many alternatives. Subjects in this study (students at Indiana University) were told that the experiment would investigate how well subjects are able to discover relationships between personality traits and behavior. Participants were asked to explain either one hypothetical outcome or two hypothetical outcomes for a particular event. The results suggested that debiasing occurred in all multiple explanation conditions, not only those that involved one opposite outcome. The results support their hypotheses that considering any alternative may be a useful debiasing tool. This may have important implications for the clinical realm, as there are many alternative explanations or diagnoses rather than just one opposite alternative explanation.

Slovic and Fischhoff (1977) found that the hindsight bias could be reduced when people were forced to consider how the results of a research study could have otherwise turned out. They investigated whether there is a difference between the perceptions of those in a foresight condition relative to a hindsight condition. That is, some subjects were asked to consider alternatives to a *hypothetical* outcome whereas others were asked to consider alternatives to an *already existing* outcome. They found that both groups were able to consider alternatives, suggesting that asking subjects to consider alternatives leads to a decrease in the hindsight bias, both prior to and after a particular outcome.

Gorman and Gorman (1984) investigated the effects of group decision making using the Wason (1960) task. Three groups were evaluated: (a) confirmatory, (b) disconfirmatory, and (c) a control group. They found that subjects in the disconfirmatory condition predicted and obtained a significantly higher proportion of incorrect strings and tested significantly more of their hypotheses with a string they predicted would be incorrect than did subjects in other conditions, suggesting the benefit of instructions to search for disconfirming information.

Gorman, Gorman, Latta and Cunningham (1984) also assessed the effectiveness of confirmation and disconfirmation instructions. Their study involved the use of the card game “New Eleusis” which is “designed to simulate the scientific, mathematical or metaphysical search for truth” (Gorman & Gorman, 1984). In playing this game, subjects were instructed to determine a rule that governs how the cards can be played. Subjects were randomly assigned to one of three groups: (a) disconfirmatory, (b) confirmatory, or (c) a mixed group using elements of both. They found that the disconfirmatory group solved more of the rules than the other two groups. They believe that the results are

explained by the increased number of attempts (of the disconfirmatory group) to search for disconfirming evidence, thereby falsifying their hypotheses. These studies provide support for the notion that considering another perspective, namely the disconfirmatory perspective, allows subjects to solve the task.

Education. Whereas some believe that educating subjects about a bias is not a practical approach, others have found this approach to be highly effective (Arkes, 1981; Kurtz & Garfield, 1978; Mynatt, Doherty, & Tweney 1977; Tweney, Doherty, Worner, Pliske, Mynatt, Gross & Arkkelin, 1980). Arkes (1981) argued that it “absolutely worthless” to tell subjects what a particular bias is and tell them not to be influenced by it, citing one study by Kurtz and Garfield, (1978). However, there are several other studies that offer evidence in support of education as a debiasing technique.

Kurtz and Garfield (1978) investigated the illusory correlation and whether educating can have an effect on decreasing the illusory correlation. Four groups of 15 subjects were told that the authors needed the help of an untrained population to study basic patterns of test interpretation. Thirty Rorschach cards were placed on transparencies along with one response that was paired with two statements about emotional problems of a supposed patient who allegedly had given that response. Groups 3 and 4 received special pre-training instructions, warning them against the illusory correlation. Despite a warning to subjects in the experimental groups, the illusory correlation was not reduced. The authors suggest that perhaps more intensive training with feedback and numerous examples may have a greater success in combating the hindsight bias. They also suggest that the results may be due to a response bias that may have been embedded in the form of the questionnaire used. For example, one question

states “Did you notice any kind of general thing that was seen most often by men with this problem? If your answer is ‘yes’ name that kind of thing and give one example of that kind of thing.” Clearly, the demand characteristic may have led subjects to think they should have something to list in response to this question, even if they did not truly see a correlation. In addition, the authors suggest that perhaps the social pressure of working in a group setting with 15 students may have affected the responses. For example, if some students are writing down responses, they may have thought that they too should find something to write down, even when they did not make an association themselves. In addition, this study may not be generalizable to an individual clinician working with an individual client.

Newstead, Pollard, Evans and Allen (1992), and Newstead, Allen, and Pollard (1994) studied the effects of instructions given to subjects aimed at avoiding the belief bias. “The belief bias is the finding that people are more likely to accept the conclusion to a syllogism if they believe it than if they disbelieve it, irrespective of its actual logical validity” (Newstead, et al. 1992). Taken together, the results of these studies suggest that the bias was reduced, although not completely eliminated. However, the instructions used in this study warned subject to avoid bias in a general sense, without specifically addressing what the belief bias is. In addition, they do not give examples of how one may fall prey to its effects or examples of how to avoid it.

Mynatt, Doherty, and Tweney (1977) investigated whether subjects who received information about the importance of disconfirming evidence when solving a computerized task would evidence decreased use of the confirmation bias. The task involved firing particles at several objects on a computer screen. Subjects were randomly

assigned to one of three groups: (a) a control group, (b) a confirmatory group, or (c) a disconfirmatory group. Interestingly, they discovered that when *all* groups received falsifying information, they recognized its significance and used it to solve the task. They also found that those subjects who used disconfirming strategies performed better than the remaining subjects. However, they also discovered that when subjects had already begun to use the confirmation bias, it was difficult to decrease the bias, even when they were instructed to do so. This suggests that perhaps addressing the effects of the confirmation bias prior to beginning a task may be more useful as an attempt to decrease bias.

Mynatt, Doherty, and Tweney (1978) revised the computer task for another study, however the task was so difficult that no subject was able to solve it in the ten hours allotted to them. Again, they found no differences between the disconfirmatory and control groups with regard to their understanding of the rules. In some cases, subjects omitted disconfirmatory rules when they had misunderstood the experiment. However, there were very few subjects in this experiment and given the difficulty of the task, results may not generalize to other problems solving tasks.

In an attempt to use a more basic and established task, Tweney, Doherty, Worner, Pliske, Mynatt, Gross, and Arkkelin (1980) tested subjects using Wason's (1960) task (discussed above). They found that the group told to search for disconfirmatory information made many more attempts to falsify their hypotheses than those in a confirmatory group, suggesting that educating subjects about the confirmation bias was in fact successful during this task.

Whereas Arkes (1981) states that educating subjects about a bias is “worthless” there are several studies suggesting just the opposite. Arkes (1981) cites a study by Kurtz and Garfield (1978) to support his statement; however, there are several problems with this study, including the presence of demand characteristics and group influence. Newstead et al. (1992) found a decrease in the belief bias when subjects were told to be aware of bias in a general sense. That is, the subjects were not given information specifically about what the belief bias is and they were still able to decrease their bias. Mynatt, Doherty and Tweney (1977) found that subjects who received information about the importance of disconfirming evidence showed a decrease in their use of the confirmation bias. Tweney, Doherty, and Worner et al. (1980) also found that educating subjects about the confirmation bias lead to a decrease in the bias. Based on the results of these studies, it is clear that educating subjects about a bias can decrease the bias and therefore is not completely worthless, as Arkes (1981) believed.

Additional Education Studies. Still, other studies have investigated the effectiveness of educating subjects about a bias in an attempt to reduce that bias. Highhouse and Bottrill (1995) have found that people who were warned about other cognitive biases, such as the misinformation effect can resist the effect when warned. The misinformation effect occurs when misleading post-event information is remembered as having been part of an event. They investigated the effects of misinformation on memory for behavioral information in an employment interview. In an initial study, subjects watched a videotape of an employment interview, and then provided ratings of the interview. They were then given evaluations made by others containing either misinformation, (mislead group) or no misinformation (control group). They were told

that they should take this information into account with their own ratings to provide a group rating. Finally, they completed a memory test of the tape. Results showed that subjects in the mislead group were accepting of the misinformation and were confident that the misinformation was accurate. The authors then conducted a follow-up experiment, in which other subjects received the same misinformation as the mislead group. However, they were told that they had received some misinformation before they completed their memory test. The results showed that this group was just as accurate as the control group. They report that these results are interesting, especially given that the subjects report that a high level of confidence in their memories. However the misinformation effects are explained, these studies suggest that when warned, subjects were able to resist the misinformation effects.

In a similar study, Greene, Flynn, and Loftus (1982) found that as long as subjects were not primed to seek out misinformation, they would likely be accepting of this information. They assessed 216 students participating in four experiments in an attempt to understand whether warning them about the possibility of future misinformation would lead to a more careful assessment of the information. The results suggest that giving a subject warning about misinformation just before the misinformation is presented may lead to a more careful analysis of the information and result in a slightly greater resistance to the misinformation. This study suggests that it may be possible to combat biases that clinicians use, potentially leading to more accurate diagnosis and assessment. These studies provide additional support for the effectiveness of educating subjects about a bias in an attempt to reduce that bias.

Potential Errors Resulting from Debiasing. Wilson, Centerbar, and Brekke (2002) discuss the importance of debiasing, and the possible errors in judgment that can result from debiasing strategies. They argue that when people are exposed to “mentally contaminating information,” (and are aware of it) they may make an attempt to correct for any known or unknown biases. However, they argue that people are often not very successful in correcting for these biases. They give the example of a hiker at a crossroads.

There is a sign pointing toward one of the paths indicating that it leads to the hiker’s destination. However, the hiker has been told that a mischievous boy often moves the sign so that it points in the wrong direction, and she has just seen a boy running away from the sign, laughing gleefully. She is pretty sure that the sign is pointing to the wrong path (i.e., that it is “contaminated”), but how should she correct for this fact? Given that the sign is pointing to the trail on the far left, should she take the one on the far right? Or should she assume that the boy only had time to move the sign a little, and therefore take the middle trail? Or that he is a clever boy who tried to convince her that he moved the sign, when in fact he did not? (p. 191).

They argue that people who are exposed to “contaminating information” and engage in debiasing may not arrive at a similar conclusion to those who were not exposed. They report finding three types of errors: (a) insufficient correction (debiasing in the direction of accuracy that does not go far enough), (b) unnecessary correction (debiasing when there was no bias to start with), and (c) overcorrection (too much debiasing, so judgments end up biased in the opposite direction). For example, Stapel, Martin and Schwartz

(1998) studied subjects' ratings of the weather in tropical places and then rated the weather in Midwestern cities. People in the experimental group, who rated the tropical locations first, and who were warned to avoid bias, overcorrected for their biases and rated the weather in the Midwestern cities as more desirable than the tropical locations.

In addition to providing suggestions to avoid bias, this study suggests that when warned about a potential bias, subjects may overcorrect under some conditions, or attempt to correct when there is no reason to, still leading to inaccurate judgments. This suggests that certain debiasing strategies may be associated with their own inherent biases, and may not necessarily lead to more accurate judgments. However, despite the unnecessary correction and overcorrection found in this study, several others have found that warning subjects about biases does lead to more accurate judgments (Mynatt, Doherty, & Tweney 1977; Tweney, Doherty & Worner, 1980). One reason for this is that the study that found that subjects overcorrected for bias involved studying the subjects' personal preferences. In contrast, these studies did not find an overcorrection effect during a problem-solving task. It may be that subjective preference is more easily swayed compared to evidence used during a problem-solving task.

Whereas some research has shown that subjects may overcompensate for a bias when warned about the bias, others have shown that subjects reach more accurate conclusions when educated about a bias. Despite the overcompensation effect found in several studies, education appears to be a potentially useful tool to use when attempting to decrease bias.

Many strategies have been offered in an effort to reduce cognitive biases. These include; (a) taking another's perspective, (b) use of computerized programs, (c)

decreasing overconfidence, (d) delaying decision making, (e) considering the opposite, (f) training, (g) considering alternatives, and finally, (h) educating subjects about biases and warning them not to succumb to their effects. Of particular interest in the present study are the attempts to debias through the use of educating subjects about a bias prior to undertaking a task. Based on the studies discussed above, education appears to be the most generally promising approach to debiasing subjects. Whereas some studies did not find a high success level, many others did find a decrease in bias. In addition, education is of particular relevance because of its high level of practical utility, as it does not involve extensive training or elaborate procedures. Subjects can simply be told about the bias, without any excessive training on how to avoid the bias or how to correct for it. This may be especially true for educating subjects in a clinical setting. A therapist may need merely to remind him or herself about bias before meeting with a client in order to decrease the effects of the bias.

Based on the studies reviewed, the results suggest that education is an effective way to decrease bias. In addition, the relationship between overconfidence and the confirmation bias is of interest, as some studies have shown that simply asking subjects to rate their confidence level leads to a more accurate level of confidence, and a decrease in bias. (Arkes, et al. 1987; Renner & Renner, 2001).

Statement of Purpose

Research has shown that the confirmation bias affects judgments in a wide array of contexts, including clinical settings. To start, the confirmation bias affects the way that individuals test their hypotheses. That is, they tend to test hypotheses in a confirmatory way, rather than attempting to search out information that may contradict

their hypotheses (Synder & Swann, 1978). Studies have further demonstrated that the confirmation bias not only affects one's hypothesis testing strategy, but also how one interacts with others and how others in turn respond.

In addition, several studies investigating the effects of diagnostic labels have found that such labels affect the way in which subjects interpret information. For example, Herbert, Nelson, and Herbert, (1988) found that assessment of "normal" behavior was affected by the provision of a clinical diagnosis. Interestingly, Langer and Abelson, (1974) found similar results, although they found that behavior therapists were not affected by the presence of the label "patient" whereas psychoanalytic therapists were susceptible and interpreted "normal" behaviors as disordered. Temerlin (1968) also found similar results when subjects interpreted "normal, healthy" behaviors as psychotic after hearing a prestigious clinician make a statement that the client was "psychotic." Making the problem even more complicated are the findings of Arkes and Harkness (1980), who found that clinicians remember a case differently when they make a diagnosis. For example, if a diagnosis of depression is made, they may remember symptoms of depression being present that the client never actually endorsed. The studies collectively demonstrate that the confirmation bias is a potentially serious problem for clinicians.

Other researchers have found that the confirmation bias affects clinical decision making in several ways. First, Haverkamp (1993) found that subject's questions were confirmatory 64% of the time and disconfirmatory only 15% of the time. This is a potential problem for clinicians if they are only seeking to confirm their initial hypotheses, and disregarding later contradictory information. Second, clinicians may

only take a *very* brief amount of time to make a diagnosis. In some cases, a clinician may only take 30 seconds to make a diagnosis (Gauron & Dickinson, 1969). Other researchers have found that it takes clinicians three minutes, (Sandifer, Hordern, & Greene, 1970), still a very short period of time to gather information. In addition to the short amount of time it takes clinicians to make their initial judgments of clients, another potential problem is the amount of information that clients reveal during the first session. Because it is often not possible for the client to reveal all of the necessary information for an accurate diagnosis in the first session, clinicians may form incorrect diagnostic impressions based on insufficient data. In addition, the client may want to develop a strong rapport with the therapist before revealing highly personal information. This alone is not necessarily problematic. However, it does become problematic when a therapist makes an initial impression and then disregards inconsistent information in later sessions. Again, the confirmation bias is a concern for clinicians, as accuracy in diagnosis is an essential first step in treatment.

Given all of this information, the confirmation bias has the potential to significantly impact a clinician's diagnostic impressions. First, a client may not feel comfortable revealing highly personal information during the first session, or even the first several sessions. Or, they may simply not have enough time to disclose all of the relevant information necessary for a clinician to make an accurate diagnosis. As suggested by the work of Gauron and Dickenson (1969) and Sandifer et al. (1970), the clinician may form a hypothesis well before the necessary information is revealed. If the clinician then follows in line with Haverkamp's (1993) research and questions the client in a way that confirms his or her initial hypotheses, he or she may make a diagnosis

prematurely. Following the results of Arkes and Harkness (1980), the clinician may then remember information consistent with the diagnosis and include symptoms that the client never actually endorsed.

Cognitive biases and heuristics are clearly a potential problem within the clinical domain. This is especially true when the confirmation bias is made even stronger through lack of awareness and overconfidence in one's clinical skills. Research has suggested that clinicians are highly susceptible to making errors in their clinical judgment (Kahemmann & Tversky, 1974). However, Greene et al. (1982) and Highhouse and Bottrill (1995) and others have found that people who were warned about cognitive biases were able to resist the bias when warned prior to reviewing the information. This work also suggests that clinicians are not as aware of these biases unless they are educated or warned prior to making an assessment, and some may continue to evidence cognitive biases.

Many suggestions have been made regarding the question of how to reduce cognitive biases. Some of these include considering alternatives, perspective taking, and educating subjects about biases, in particular warning them not to be influenced by a certain bias. Of particular interest in the context of the present study are the attempts to debias through the use of education or warning subjects about a bias prior to undertaking a task. Based on the studies mentioned, the results are not entirely conclusive, with some authors suggesting that warning subjects about a bias is inadequate, whereas others have found great success in warning subjects. (Highhouse & Bottrill, 1995; Green, Flynn & Loftus, 1982; Newstead & Pollard et al. 1992; Tweney, Doherty, & Worner, et al. 1980). It appears that education may be an effective technique, and may be especially effective

when used in conjunction with an attempt to decrease overconfidence. It appears that this may be the most promising technique as it has a high level of practical utility. That is, it does not involve any extensive training or elaborate procedures, as some other approaches do, and may be readily applicable to subjects in the clinical setting.

This research raises interesting questions about cognitive biases and whether clinicians are able to resist biases when assessing and diagnosing clients. We believed that if the confirmation bias could be resisted by subjects during a laboratory clinical task (Highhouse & Bottrill, 1995), perhaps the confirmation bias could be resisted within a clinical setting as well. The present study examined these questions. First, this study examined the existence of the confirmation bias among clinicians receiving inconsistent information in two case vignettes across two different time points. This study has also assessed whether clinicians who were educated about the confirmation bias prior to making a diagnosis were able to guard against the bias. In addition, this study took steps beyond the education and warning studies discussed above. First, the case vignettes provided to clinicians were separated by a one week interval, simulating the typical outpatient psychotherapy setting, thus providing greater ecological validity. The procedures remained as close as possible to actual clinical contexts, while continuing sufficient experimental control. In addition, given the research on overcorrection and unnecessary correction, (Stapel, Martin & Schwartz, 1998) a control group was included to determine whether merely telling subjects about a bias would oversensitize them to alter a diagnosis when no change is necessary.

The present study therefore addressed the following research questions:

1. Do clinicians evidence a confirmation bias in an ecologically valid analogue diagnostic assessment task? That is, do they fail to alter their initial diagnostic judgment based on new data when the latter could disconfirm the initial diagnosis? It was hypothesized that a significant number of clinicians would evidence the confirmation bias and fail to alter their initial diagnosis even when new information presented at time two had disconfirming evidence.

2. Do explicit instructions regarding the confirmation bias mitigate its effects? It was hypothesized that the presence of instructions regarding the confirmation bias would mitigate the effects of the confirmation bias. Specifically, for those participants who received confirmation bias instructions in a condition in which a diagnostic change was required, it was expected that they would alter their diagnosis at a significantly higher rate than those who did not receive confirmation bias instructions.

3. Do clinicians who receive information about the confirmation bias engage in unnecessary correction of the confirmation bias? That is, do they alter an initial diagnosis when the information provided does not indicate that a change in diagnosis is indicated? It was hypothesized that the presence of the confirmation bias instructions would not lead clinicians to inaccurately alter their initial diagnosis, or to over-correct. Specifically, it was expected that clinicians who received the confirmation bias instructions would not alter their diagnosis when no change was required.

Chapter 2: Methods

Participants

A total of 102 participants completed this study at both time points and data collection took place over a period of nine months. Participants were recruited from several psychology listservs including 1) The Acceptance and Commitment Therapy (ACT) listserv, 2) The Association for Behavioral and Cognitive Therapy (ABCT) listserv, 3) The Counselor Education and Supervision (Cesnet) listserv, as well as, 4) The Society for a Science of Clinical Psychology (SSCP) listserv. In addition, requests were sent to several institutions such as practicum sites through Drexel University's Psychology Department, and approximately twelve rotation sites through the Southwest Consortium Pre-doctoral Psychology Internship Program in Albuquerque, New Mexico. It was also requested that clinicians who received the study information distribute it to other clinicians who fit the inclusion criteria. The inclusion criteria were: (a) licensure as a psychologist in any state or Canadian province, and (b) current practice of psychodiagnosis and psychotherapy. There were no exclusion criteria.

In addition, 62 clinicians took part at time one and not at time two. The reasons for this are unclear due to the anonymity of participation. However, it is possible that because each participant was asked to take part at two different time points, the commitment was more difficult to follow through with. In addition, it is possible that clinicians may have completed the task at time one, forgotten their subject number and started again. Therefore, there is a possibility, because the study is anonymous per IRB requirements, that some clinicians took part at time one and then repeated participation at time one in order to complete the study. Finally, general reminder emails were sent out

to participants who may have participated, i.e., those clinicians at practicum sites and rotations sites, requesting that anyone who began the study at time one log on to the website to complete step two. The specific participants who took part at time one were not identified through the emails; rather, the initial recruitment emails that were sent out to groups of clinicians were followed up with and all responses continued to remain anonymous. It may be that those clinicians were more likely to follow through, compared to those clinicians that received the information about the study via listservs, and did not receive any reminders.

Research Design

The research design was a 2 (debiasing instructions vs. no debiasing instructions) by 2 (diagnostic change vs. no diagnostic change vignette) by 2 (time: baseline vs. 1 week retest) mixed factorial design, with the instructions factor being between-subjects and with repeated measures on the last two factors (see Figure 1). Participants were randomly assigned to one of two groups. One group received instructions regarding the confirmation bias with particular emphasis on its potential impact on the psychodiagnostic process, whereas the other group received no such instructions. Participants in each group were presented with two clinical vignettes at baseline (time 1), then additional information on each vignette at 1-week retest (time 2). In the diagnostic change condition, the information presented at time 2 required a change in the diagnosis according to DSM-IV criteria. In the no change condition, the information provided at time 2 was fully consistent with the information provided at time 1, and did not require a change in diagnosis.

In addition, the specific vignettes were counterbalanced across condition, such that participants were able to rate different vignettes in each of the two diagnostic conditions. That is, a total of four vignettes were used, as described below (see Figure 2).

The primary dependent variable was a categorical variable representing whether participants accurately changed or did not change their diagnosis from time 1 to time 2. Thus, for purposes of data analysis, the design can be conceptualized as a 2 (debiasing instructions vs. no debiasing instructions) by 2 (diagnostic change vs. no diagnostic change vignette) mixed factorial design, with the first factor being between-subjects and the second being within-subjects. The time factor was not analyzed as a factor per se, but instead represents the categorical dependent variable (i.e., whether or not participants accurately modified their diagnosis).

Materials

Case vignettes. Four case descriptions based on examples from the *DSM-IV Case Book* (Spitzer, Gibbon, Skodol, Williams, & First, 1994) and the *DSM-IV-TR Case Studies* (Frances & Ross, 2001) were used (see Appendix G). Materials were designed so that, per strict adherence to DSM-IV criteria, one could reach a specific diagnosis after reading the material. This was confirmed through pilot work with experienced diagnosticians and is discussed further below.

Each case description was divided into two parts. The first part was provided to the clinicians at time 1, and the second at time 2, one week later. Two of the case vignettes involved an anxiety disorder, and two involved a psychotic disorder. Vignette A began with a description of symptoms characteristic of Panic Disorder with Agoraphobia. The second part of vignette A, provided at time 2, continued with a

description of Panic Disorder with Agoraphobia. Vignette B likewise began with a description of Panic Disorder with Agoraphobia. The second part of vignette B, however, consisted of symptoms that modify the diagnostic picture to be consistent with Social Anxiety Disorder. Vignette C began with a description of Schizophreniform Disorder. The second part of vignette C, provided at time 2, continued with a description of Schizophreniform Disorder. Finally, vignette D started with a description of Schizophreniform Disorder. However, the information provided at time 2 modified the diagnosis to an Alcohol Induced Psychotic Disorder.

Procedure

In an attempt to ensure that the clinicians would assign the correct diagnosis to each vignette independent of the other vignettes, a pilot procedure was conducted with experienced clinicians not taking part in the study. Twenty clinicians including senior level clinical psychology graduate students at Drexel University read these vignettes independent of the other vignettes and assigned diagnoses to each vignette they read. The results suggested that these clinicians were accurate 96% of the time.

In addition, in order to investigate the effect of the confirmation bias instructions on clinicians' diagnostic decision making, an assurance was needed that these vignettes would elicit the confirmation bias for some clinicians. A second pilot study was conducted with another set of practicing clinicians, also not taking part in the study, to determine whether or not the existing vignettes would elicit the confirmation bias. Twenty-five participants participated in this pilot study. Results of the pilot study suggested that clinicians evidenced the confirmation bias approximately 23% of the time,

and therefore the vignettes did not require modification. Data collection for both pilot studies took place over the course of 3 months.

The procedural design was developed in a way that incorporated the IRB requirements for strict anonymity for all participants. First, several listserv administrators were contacted via email requesting permission to advertise for participants through their listserv (Appendix A). Once we received permission, another email was sent to the administrators that they then forwarded to the members of the listserv. This email included a brief overview of the general purpose of the study, the study procedures, and the approximate amount of time required to participate in the study (Appendix B). If clinicians were still interested in participating, they clicked on a link that directed them to the first page of the website where they read additional information about the study and were given the option to participate. In an attempt to eliminate those clinicians who were not willing to complete the study, it was also on this page that each reader was encouraged not to participate unless they were willing to take part at both time points (Appendix C).

If the reader agreed to participate, he or she would randomly click on either A, B, C or D, which were all links that provided the participant with access to the first two vignettes. However, prior to reading the vignettes, each participant was directed to a page that required them to assign themselves a five digit subject number and which gave more specific instructions (Appendix D). This allowed each participant and their responses to remain completely anonymous. After reading each of the first two vignettes, each participant completed a form asking for their clinical impression ratings such as a DSM-IV diagnosis as well as what they may have thought to be potential obstacles to

treatment. Once participants read the first two vignettes and completed the clinical impression rating forms (Appendix H), they were asked to complete a demographic form (Appendix I). Once the form was complete and the participant clicked on another link, the information was sent to a secure database. At this time, each participant's subject number appeared on the screen. This screen reminded each participant to write down his or her subject number as well as a reminder to log on to the same website in approximately one week. One to two weeks later, the participants logged onto the website. At the bottom of the page, an option was provided for those participants logging on at time two. After clicking on that link, they were required to enter their five-digit subject number. If they did not have that number, they were not able to participate at time two and the data from time one was not included in the data analysis. If they were able to enter their subject number, they were given the final instructions (Appendix E) and were given access to a) the first two vignettes they read at time one, b) the initial clinical impression rating forms with their responses from time one, c) two additional vignettes, and d) two new blank clinical impression rating forms. Once the participants read the new vignettes, and completed the new forms, they submitted them to the same secure database by clicking on another link. Finally, participants had the option of clicking on one final link to receive several suggestions for how to avoid making the confirmation bias when practicing psychodiagnosis and psychodiagnostic decision making (Appendix J). They were also given the website address where the results of the study will be posted once the study is complete. The web address is

<http://www.psychology.drexel.edu/anxietyresearch/pages/>.

Prior to presentation of the vignettes at time 2, participants in the bias sensitization condition, sets B and C were provided with a description of the confirmation bias, a brief discussion of its relevance to the psychodiagnostic process, and instructions to be aware of the bias when completing the task (Appendix F). Those in the no bias sensitization instruction condition, sets A and D, did not receive these instructions. Instructional condition was determined through random assignment when each participant selected the A, B, C or D condition. To ensure that each experimental condition had an equal number of participants, conditions were blocked one at a time on the website once 25 participants had taken part at both time points so that new participants would not select a condition which already had 25 subjects.

Regardless of instructional condition, at time two, all participants were provided with the same vignettes they read at time one, plus additional information on each vignette. For one of the vignettes, the additional information was consistent with the original diagnosis. For the other vignette, the additional information required a change in diagnosis. The specific disorders displayed in the vignettes (i.e., an anxiety disorder vs. a psychotic disorder) were counterbalanced across condition. In addition, the order of presentation of the vignettes (i.e., diagnostic change vs. no change condition) was also counterbalanced.

Statistical Analyses

The principal dependent measure was a binomial category reflecting whether participants accurately changed or did not change their diagnosis from time 1 to time 2. These data were analyzed by a modified form of logistic regression known as generalized estimating equations (GEE). In standard logistic regression, a categorical dependent

variable is predicted by either categorical or continuous predictor variables, but the latter must represent independent observations (i.e., they must be between-subjects variables). GEE permits one or more of the predictors to be a repeated-measures variable. The first predictor variable is instructional condition, and the second is diagnostic change condition, both of which were dummy-coded. The effect of interest is the interaction between these two variables, which was entered into the equation following the two predictors.

Statistical Power

As GEE is a relatively new technique, the methods for estimating power for interaction effects are not yet established. Thus, standard multiple regression was used to estimate the number of participants required for the present study. For a medium effect size based on Cohen's (1988) conventions for a multiple regression with two predictor variables and their interaction, 68 participants are required to achieve a power of .80 and alpha of .05. Because logistic regression is generally less powerful due to the dependent variable being dichotomous rather than continuous, a minimum of 100 participants were required to be enrolled and a total of 102 clinicians took part in the study at both time points. Due to the significant main effect described below, it is clear that 102 participants were sufficient to have an adequate level of power. Counting each trial as independent, (which is valid given the significant difference between participants' accuracy between the psychotic and the anxiety vignettes) the sample size is adequate to detect a 20-percentage point difference in accuracy, e.g., 50% vs. 70%. For the worst case, where we consider a sample size of 102 participants split 51 vs. 51, we could detect a 28% point difference.

Chapter 3: Results

Demographics

A total of 102 participants completed the study at both time points, totaling 25 in two of the experimental conditions, and 24 and 28 in the two remaining conditions. An additional 62 participants completed the study at time one, but did not complete the study at time two. These participants were flagged two and a half to three weeks after they participated at time one to ensure that an appropriate and realistic time line was followed to mirror real world time between psychotherapy sessions. If any of these participants attempted to complete the study, their data would not have been used. However, none of the participants that allowed two and a half weeks to lapse attempted to complete the study at time two.

Gender: A total of 61 women and 38 men took part in the study (See table 1). Three participants did not identify their gender.

Age: The age of participants ranged from 24 to 79 with a bimodal distribution. A large number of participants were in their mid thirties and mid fifties (See table 2).

Years in practice: The participants' number of years in practice ranged from 1 year to 52 years, with a large number of the participants practicing between 2 and 6 years (See table 3).

Theoretical Orientation: Participants' primary theoretical orientation ranged over nine different orientations including psychodynamic, cognitive-behavioral, family systems, eclectic, biological, ecological, bio-psycho-social, humanistic and existential. Participants were not provided with a pre-determined list of orientations and were free to

write in their orientation. The majority of the participants (57.8%) reported that they were cognitive-behavioral in their approach to psychotherapy (See table 4).

Data Analytic Strategy

The principal dependent measure was a binomial category reflecting whether participants accurately changed or did not change their diagnosis from time 1 to time 2. The first predictor variable was instructional condition, and the second was diagnostic change condition, both of which were dummy-coded. The primary effect of interest was the interaction between these two variables, which was entered into the equation following the two predictors.

Of note, tables 5a and 5b reflect change, or consistency, whereas tables 6 and 7 reflect accuracy. Specifically, being correct in changing one's diagnosis does not necessarily imply diagnostic accuracy. For example, a clinician may accurately change a diagnosis in the change condition, but may inaccurately change the diagnosis from panic disorder to generalized anxiety disorder, rather than change it to the correct diagnosis of social phobia. Analyses were run using both consistency and then accuracy as standards. As the pattern of results was identical for these two standards, only the accuracy results are reported below. In addition, note that the results presented do not include vignette type as a covariate unless otherwise specified, as results were run both ways and results were consistent.

Main Analyses

Main Effect #1. Results suggest that the diagnostic change condition led to poorer performance overall relative to the no-change condition (*GEE Parameter Estimate* = -2.01, *SE* = .72, *p* = .005). That is, when the vignettes required a diagnostic change,

participants did not perform as well compared to when the vignettes required the diagnosis to remain consistent.

Main Effect #2. Results suggested that overall, the presence of confirmation bias instructions did not affect participants' performance (*GEE Parameter Estimate* = -.08, *SE* = .84, *p* = .92). Therefore, independent of the diagnostic change condition, the presence of the confirmation bias instructions did not affect performance by participants.

Interaction Effect. When the interaction of the main effects were analyzed, no significant interaction effect was noted (*GEE Parameter Estimate* = .42, *SE* = 1.01, *p* = .68). Specifically, the interaction of the diagnostic change condition and the confirmation bias instructions did not lead to a change in accuracy by participants.

In summary, the diagnostic change condition led participants to perform poorly relative to the no change condition. Second, the confirmation bias instructions neither hindered nor helped the participants in their performance. Finally, the interaction of the two main effects did not affect participants' performance in either a positive or negative way.

Hypotheses

Hypothesis I: A significant number of clinicians will evidence the confirmation bias and fail to alter their initial diagnosis even when new information presented at time two has disconfirming evidence.

Results of Hypothesis I: Overall, clinicians evidenced the confirmation bias 33% of the time in this condition. In examining more closely each vignette, clinicians evidenced the confirmation bias 16% of the time when evaluating the psychotic vignettes, and 50% of

the time when assessing the anxiety vignettes. See tables 6 and 7 for descriptive information.

Hypothesis II: Presentation of instructions regarding the confirmation bias will mitigate the effects of the confirmation bias. Specifically, for those participants who received confirmation bias instructions in a condition where a diagnostic change was required, it was expected that they would alter their diagnosis at a significantly higher rate than those who did not receive confirmation bias instructions.

Results of Hypothesis II: Results showed that the percentage of clinicians who evidenced the confirmation bias even when provided with the confirmation bias sensitization instructions was 23%, compared to 33% in the condition where no instructions were received (see table 6). As noted above, this difference did not reach statistical significance.

Hypothesis III: The presence of the confirmation bias instructions will not lead clinicians to inaccurately alter their initial diagnosis, or to overcorrect. Specifically, clinicians who receive the confirmation bias instructions will not alter their diagnosis when no change is required.

Results of Hypothesis III: Results suggest that the confirmation bias instructions did not lead clinicians to inaccurately alter their diagnosis in the conditions in which it should not have changed. Over-correction occurred 5.7% of the time in the confirmation bias sensitization instructions condition, compared to 6.1% in the condition with no confirmation bias instructions (see table 6).

Examination of table 6 indicates the percentage of clinicians who were accurate in each condition collapsed across vignette type. It is interesting to note that it appears as

though the confirmation bias sensitization instructions resulted in a trend for participants to perform more accurately when a change in diagnosis was required. Specifically, within this condition, accuracy reached 77%, compared to 67% without confirmation bias instructions. However, despite the appearance of slightly more accurate performance in this condition, this effect did not reach statistical significance (*GEE Parameter Estimate* = .75, *SE* = 1.04, *p* = .47).

Table 7 reports the percentage of clinicians who were accurate in each condition broken down by type of vignette, i.e, psychotic vignettes versus the anxiety vignettes. Interestingly, 16% of clinicians evidenced the confirmation bias without the bias sensitization instructions when evaluating the psychotic vignette, whereas 50% of clinicians evidenced the confirmation bias without the instructions when assessing the anxiety vignette.

Finally, despite receiving confirmation bias sensitization instructions, clinicians continued to evidence the confirmation bias 12% of the time when evaluating the psychotic vignette and 32% of the time when evaluating the anxiety vignette.

Additional Analyses

Predictive Factors

Gender: Results indicate that gender was not a significant predictor of diagnostic accuracy (*GEE Parameter Estimate* = .44, *SE* = .37 *p* = .24); gender was therefore not included as a factor in subsequent analyses.

Vignette Type: All analyses were run with and without vignette type, (the psychotic vignette versus the anxiety vignette) as a covariate, given the significant differences between participants' performance across vignette type (*GEE Parameter*

Estimate = -25.61, *SE* = .68, $p < .0001$). As discussed above, note that the results presented do not include vignette type as a covariate unless otherwise specified, as results were consistent across both analyses.

Age: Results indicate that age was negatively correlated with performance (*GEE Parameter Estimate* = -.04, *SE* = .02, $p = .02$).

Years in Practice: Results indicate that years in practice was not a significant predictor of accuracy (*GEE Parameter Estimate* = -.02, *SE* = .02, $p = .28$).

Orientation: When all orientations were analyzed as a predictor of performance, the results were insignificant. When CBT was analyzed alone compared to all other orientations, the results again did not indicate a significant difference between orientations as a predictor of overall accuracy (*GEE Parameter Estimate* = -.01, *SE* = .37, $p = .99$). However, given that there was a significant difference in participants' performance between the psychotic vignettes and the anxiety vignettes, (*GEE Parameter Estimate* = -25.61, *SE* = .68, $p < .0001$) vignette type was added in as a predictor variable for this analysis. These results suggest that those who identified their orientation as cognitive-behavioral performed better on the psychotic vignettes (*GEE Parameter Estimate* = -2.43, *SE* = 1.24, $p = .05$) although no difference in orientation was noted for the anxiety vignettes.

Finally, results suggest that age was not a predictor of orientation (*GEE Parameter Estimate* = .02, *SE* = .03, $p = .53$).

Chapter 4: Discussion

Purpose

The purpose of this study was to investigate the effects of the confirmation in the process of clinical psychodiagnosis, as well as the possibility that instructions regarding such a bias might mitigate its effects. The study investigated whether clinicians would fail to alter an initial diagnosis even when additional information warranted a change. The effects of information about the confirmation bias were hypothesized to correct the bias. However, there was a possibility that participants might evidence an over-correction when they received information about the confirmation bias. Specifically, this study investigated whether participants who received confirmation bias instructions would change their initial diagnosis, even when the symptoms presented at time two did not require a change. Finally, this study investigated the possible effects of gender, age, years in practice, and theoretical orientation on diagnostic accuracy. Specifically, the study investigated whether these variables were significant predictors of the tendency to demonstrate the confirmation bias in the process of psychodiagnosis.

In sum, the overall purpose of this study was to determine whether the confirmation bias occurred during an analogue study and whether the confirmation bias instructions would mitigate the confirmation bias effects in any way. It was hoped that the results might be applicable to actual clinical practice, in which clinicians may be evidencing this bias with their patients. Given the potential problems caused by the confirmation bias, decreasing this bias among clinicians is a high priority.

Summary of Results

Demographics

A total of 102 clinicians participated at both time points. Sixty-two clinicians took part at time one and not at time two. Thirty-eight men and 61 women took part in the study, and three participants did not identify their gender. The age of participants ranged from 24 to 79. Years in practice ranged from 1 to 52. Finally, participants identified themselves as belonging to one of nine different orientations, although the majority identified themselves as having a cognitive behavioral orientation.

Main Analyses

First, the results suggest that when a change in diagnosis was required in the vignettes from time one to time two, regardless of the presence of confirmation bias instructions, participants tended to perform poorly. Overall, 32.6 % of the clinicians that took part in this study evidenced the confirmation bias. Second, the results suggest that regardless of diagnostic change condition, the presence of the confirmation bias sensitization instructions did not affect participant's performance. Overall, 23% of participants in the change condition continued to evidence the confirmation bias even with instructions warning them to be aware of this bias, compared to 33% that evidenced the bias in the change condition without instructions. Moreover, 5.7% of participants changed their diagnosis at time two when they should have remained consistent when they received confirmation bias instructions, compared to 6.1% of participants who changed their diagnosis in the no change condition, when they did not receive confirmation bias instructions. The results did not indicate any significant interaction between the two main effects, i.e., the diagnostic change condition and the confirmation

bias instructions condition. Finally, as noted above, participants' performance was significantly different between the psychotic vignettes and the anxiety vignettes. Specifically, participants performed more accurately when taking part in the anxiety vignettes (*GEE Parameter Estimate* = -25.61, *SE* = .68, *p* < .001).

Predictor Variables

Results suggest that neither gender nor years in practice played a role in whether participants evidenced the confirmation bias. However, the results suggest that age was negatively correlated with accuracy in performance.

Finally, there was no significant difference between each orientation when all were included as distinct predictors. In addition, there was no significant difference between CBT and all other orientations combined in predicting accuracy in diagnosis, except when vignette type was factored in as a covariate. Here, those who identified themselves as cognitive behavioral in orientation performed more accurately on the psychotic vignettes.

Implications

It appears that gender was not a significant predictor variable for evidencing the confirmation bias. Therefore, it is reasonable to suggest that both male and female clinicians of any training level are equally susceptible to making the confirmation bias. This is not surprising, as there is no a priori reason to suspect that either gender would be more or less prone to the confirmation bias, although it provides additional evidence that this is a problem that all clinicians need to be aware of.

It also appears that years in practice was not a significant predictor variable for evidencing the confirmation bias. This suggests that clinicians do not become less (nor

more) prone to the bias as they continue in their practice of psychodiagnosis. Therefore, experienced and less experienced clinicians are equally susceptible to evidencing this bias. This highlights the need to determine a way to counter the bias as additional experience does not protect against it.

The results also suggest that age was negatively correlated with accuracy in performance. This may be due in part to the notion that newer training programs may tend to be more focused on empirically supported assessment strategies, and therefore may also focus more extensively on issues such as cognitive biases. Although this may not be the case across all programs, especially when free-standing professional schools are considered, it may be the case with the majority of the training programs represented by the present sample. Recall that many of the listservs through which participants were recruited were comprised mainly of clinical psychologists who generally adhere to a scientific-practitioner training model. It is also possible that younger clinicians who were trained with the DSM-III (or subsequent editions of the DSM) tend to take a more systematic approach to psychodiagnosis compared to those clinicians who were trained earlier.

Orientation was not a predictor of accuracy when vignette type was not factored into the equation. However, when analyzing orientation as a predictor of accuracy by vignette type, those who identified their orientation as cognitive behavioral performed more accurately for the psychotic vignettes than those in all other orientations combined. However, there was no significant difference in performance for the anxiety vignettes. This result is somewhat confusing as there is no a priori reason for those clinicians who identify themselves as cognitive behavioral to perform more accurately on one vignette

but not another. However, this may be due in part to the fact that the majority of those who participated identified themselves as cognitive behavioral and the majority of errors took place within the anxiety disorder vignettes. Therefore, this may be coincidence and not a true difference in ability.

Overall, a comparison in participant's performance between the psychotic vignettes and the anxiety vignettes suggested that participants made more errors when completing the anxiety vignettes. The psychotic vignette was always presented first in each set and the anxiety vignette was always presented second. Although the diagnostic change condition was counterbalanced across each set, the results revealed that participants made more errors for the anxiety vignettes. It is possible that these errors were made due to the order in which vignettes were given, or it may be due to the similarities in symptoms across anxiety disorders. For example, the evidence is presented more clearly in the psychotic vignette regarding the etiology of the pseudo client's psychosis. That is, the vignette clearly states that she had been drinking for the same amount of time that she had experienced the psychotic symptoms and a large part of the vignette is devoted to making that point clear. The symptoms for the anxiety vignettes are more similar in nature as they both focus on symptoms of anxiety. Clinicians reading the anxiety vignettes may have focused more on the pseudo-patient's physical symptoms rather than the fact that these symptoms were elicited by social situations. Specifically, clinicians may have overlooked the fact that many people suffer from panic attacks when they suffer from social phobia, and the participants may have focused on the fact that the client was still experiencing anxiety attacks. Finally, another explanation is that this difference in performance stems from variability in clinicians' diagnostic skills. For

example, some clinicians may not be aware of the precise nuances between the various anxiety disorders. Specifically, they may not understand that panic attacks commonly occur within the context of other anxiety disorders aside from panic disorder itself, and in fact do occur within the context of social phobia. Therefore, this result may be based on knowledge of diagnostic standards rather than the confirmation per se.

Interestingly, the majority of those who participated in this study identified themselves as having a cognitive-behavioral orientation. This fits with the fact that there were a large percentage of younger clinicians taking part in the study, and many had been practicing for only 2-6 years, as newer training programs are frequently cognitive-behavioral in focus. In addition, several of the listservs that were used to recruit participants were cognitive-behavioral in nature and focused on empirically supported treatments, e.g., SSCP, ACT and ABCT. Additionally, many clinicians may claim to be cognitive-behavioral in orientation due to the gradual shift in focus over the last twenty years from psychodynamic psychotherapy to cognitive-behavioral therapy, but may not truly have the appropriate training in cognitive-behavioral therapy, or may not truly practice from a cognitive-behavioral orientation. Finally, many clinicians may consider themselves to be *primarily* cognitive-behavioral, but may truly practice from an eclectic approach.

Given that participants performed poorly overall when they completed a vignette set that required a change in diagnosis compared to their performance when completing a vignette set that did not require a change in diagnosis supports the notion that clinicians often like to remain consistent even when additional disconfirming evidence is provided, and suggests a confirmatory approach to psychodiagnosis. In addition, this was true

regardless of whether or not the confirmation bias instructions were present, providing evidence for the strength of the confirmation bias. This result supports the work of Elstein et al. (1978), Gauron and Dickinson (1969), Sandifer et al. (1970), and Synder and Swann (1978), and as well as many others who also found that the confirmation bias was evidenced in their work.

Interestingly, the presence of the confirmation bias instructions, regardless of diagnostic change condition, did not significantly affect participants' performance. This suggests that the confirmation bias instructions did not help clinicians to perform more accurately in any significant way, although it also did not hinder clinicians' performance in a significant way either. Specifically, the confirmation bias instructions did not significantly help participants to accurately change their initial diagnosis when the vignette called for such a change, and therefore, participants continued to evidence the confirmation bias. In addition, the confirmation bias instructions did not help participants to remain consistent when the vignettes required no change in diagnosis as there was no significant difference between the groups in the no change condition.

However, based on the results discussed from table 6, it may be that the confirmation bias instructions did positively affect participants' performance, although not significantly and a type II error may have been made. When evaluating the difference in percentages between those who altered their diagnosis when a change was required and received confirmation bias sensitization instructions relative to those who altered their diagnosis when a change was required and did not receive confirmation bias instructions, there appears to be a trend, albeit not statistically significant (see table 6). Specifically, those in the change condition who received the confirmation bias

instructions performed with 77% accuracy whereas those in the change condition who did not receive confirmation bias instructions performed at 67% accuracy. These results suggest that the confirmation bias instructions may have in fact contributed to a trend toward more accurate performance when a change in diagnosis was warranted. This non significant result may be due to limited power; perhaps with a larger sample size the difference would have reached statistical significance. In addition, despite the nonsignificant result, this should be considered in future studies, especially given the work of Mynatt et al. (1977), Newstead et al. (1992), Newstead et al. (1994) and Tweney et al. (1980) who found that educating subjects about biases was somewhat successful. This insignificant trend is consistent with this literature, as one might expect that the confirmation bias instructions would improve performance in the change condition, e.g., if a clinician is uncertain about changing a diagnosis, the confirmation bias instructions may lead the clinician to alter their diagnosis in an attempt to avoid making the confirmation bias.

One explanation for why some participants continued to show the confirmation bias despite the presence of confirmation bias sensitization instructions is that they were eager to complete the task and did not read the instructions carefully, or truly consider them when completing part two of the task. Another explanation is that clinicians are aware that biases such as the confirmation bias exist, but do not believe that they apply to themselves. They therefore may have continued to quickly dismiss disconfirmatory information and feel justified in seeking confirmatory information. A third possibility is that these clinicians are not completely confident in their clinical skills and questioned their accuracy. Finally, the most parsimonious explanation is that the confirmation bias

instructions are simply not powerful enough to significantly counteract such a powerful cognitive bias. This is especially alarming given the frank nature of the instructions directly informing clinicians about the bias and warning them not to evidence this bias.

Overall, the results suggest that the confirmation bias is evident in the process of psychodiagnosis, and that instructions to beware of the confirmation bias did not significantly mitigate the effects of the bias.

Limitations

There are several limitations of the design, data collection and external validity of this study. Potential remedies for many of these limitations in future studies are addressed below under future directions. One limitation is that no information was collected on participants' type of degree or professional discipline. For example, there may have been significant differences in accuracy between psychiatrists, psychologists, counselors, and social workers, although it was not possible to conduct an analysis of this kind.

A second limitation is that there is no data about how the clinicians arrived at their diagnoses. Of course, direct inquiry into one's clinical decision making may not yield an accurate account of the process, as individuals may not be able to report accurately on their cognitive processes. In addition, attending to this process in order to report it might very well have resulted in reactance.

A third limitation is that the entire study was completely anonymous. Therefore, we were unable to attempt to retain the 62 clinicians that took part at time one, and not at time two. It may have been that they would have completed the study had a simple email reminder been sent to them. In addition, it is possible that because the study was

completely anonymous, that some clinicians took part at time one and did not complete their initial task at time two, and therefore started over again with a new subject number. Although there is no way to determine whether or not this occurred, this may have influenced how these clinicians responded to the vignettes the second time that they participated. That is, they may have had even more exposure to the initial vignettes, potentially making it more likely that they would ignore disconfirmatory data at a later time.

In addition, clinicians were not only responsible for reminding themselves to complete the study, but they were also responsible for remembering their five-digit user identification number. It is possible that if an email reminder were sent out with their user identification number, that more clinicians would have taken part at time two.

Fourth, as with all analogue studies, external validity is limited. Specifically, in the real world, clinicians would have the opportunity to ask questions to clarify any questions or concerns they may have had and patients would be able to fully report on their symptoms. In support of this, there were several responses from participants that indicated the desire for more information about the client and the client's symptoms in the vignettes. In addition, there are many qualities attached to the real world interactions such as facial expression, tone of voice etc. that could not be incorporated into this study. Therefore, clinicians were not only limited by the lack of this information, but they may have also completed this task in a hurried manner, as the vignette likely did not hold any personal relevance to them, as a client might. However, as there is always a balance between external and internal validity, we needed to continue sufficient experimental control while also replicating the real world time between sessions.

Fifth, the anxiety vignettes were always presented second in each set of vignettes. Given this, we are not able to rule out order effects as the psychotic vignette was always presented prior to the anxiety vignette.

Sixth, this study did not inquire about participants' confidence level. This information would have been interesting, but simply asking about their confidence level may have influenced their responses. That is, it may have caused the clinicians to read more thoroughly, focus more on their responses, and to possibly perform more accurately.

Next, there were sample limitations. Specifically, the listservs selected for recruitment were primarily those that attract scientifically minded clinicians. For example, SSCP, ACT and ABCT are all listservs whose members are likely to endorse more empirically supported treatments. This was evident in that nearly 58% of participants identified themselves as cognitive-behavioral in orientation. Had this study recruited participants from other listservs that attract clinicians who identify themselves as primarily psychodynamic in nature, the sample would likely have been more balanced in terms of orientation.

Finally, this study was not able to determine which clinicians spent a great deal of time reading the vignettes and which clinicians merely skimmed the material. This would have provided useful information regarding whether the amount of time reading vignettes and amount of time spent in assigning diagnoses contributes to the confirmation bias.

Whereas there are several weaknesses of the design of this study, the design is still able to assess the hypotheses of interest. In addition, many of these weaknesses may be corrected in future studies, as discussed below.

Strengths

One strength of the study is that the clinicians were instructed to complete the vignettes with a one-week time interval between each time point, simulating the typical outpatient psychotherapy setting, thus providing greater ecological validity. The procedures have remained as close as possible to actual clinical contexts, while continuing sufficient experimental control.

In addition, given the research on over-correction, (Stapel, Martin & Schwartz, 1998) a control group was included to determine whether merely telling subjects about a bias would oversensitize them to alter a diagnosis when no change was necessary. The inclusion of these various experimental conditions within one study is one of the greatest strengths of the study.

Finally, all responses remained anonymous which holds the potential for clinicians to be more honest in their responses. This also provides the possibility for clinicians to be more likely to participate. Had the study been in person, clinicians may have been less likely to participate.

Overall, there are several strengths and weaknesses. Some of these weaknesses have been addressed below with suggestions for future studies and some may remain weaknesses in an attempt to balance internal and external validity. It does appear that the study has been effective in assessing the hypotheses at hand in a valid way and therefore, the results should not be dismissed based on the aforementioned weaknesses.

Future Directions

The results presented in this study lead to several possible follow-up studies. A study incorporating the use of a structured clinical interview would be useful to determine whether such an interview is able to counter the biases that tend to occur during open ended interviews. Based on the work of Whaley and Geller (in press), this may be a useful intervention to decrease the confirmation bias. Structured interviews such as the *Structured Clinical Interview for the DSM-IV* (First et al., 1996) are quite lengthy, and their routine use in many clinical settings is unrealistic. An alternative approach might be shorter, semi-structured interviews.

Next, a study investigating the effects of confirmation bias instructions that are more strongly worded may be more effective in leading participants to a more accurate direction. That is, instructions that are more strongly worded may lead clinicians to change their diagnosis when necessary, but not to over-correct and change their diagnosis when they should not. Therefore, a study investigating several different forms of confirmation bias instructions would be useful in determining whether the bias instructions used in this study were not strongly worded enough. However, as part of this study, the confirmation bias instructions were purposefully written in a way that was direct, although not too strongly worded, in an attempt to avoid overcorrection, or changing a diagnosis when there is no information that requires a change. In addition to using different forms of the confirmation bias instructions, it would be useful to investigate more extensive training with clinicians as a way to determine whether clinicians would decrease their use of the confirmation bias following this training. For example, it would be useful to use specific examples of how to avoid the confirmation

bias before taking part in a study like this one to determine whether this decreases the incidence of the confirmation bias.

Another potential follow-up study would be to ensure that, as part of the demographic questions, participants are asked about their professional training and discipline, in order to determine what percentage of clinicians taking part have a Masters of Social Work, a Counseling or Clinical Psychology degree, or a Psychiatry degree. This would also determine whether professional discipline is a predictor of accuracy. In addition, it would be useful to obtain information about the participants' training programs and the amount of training that they received on biases and heuristics.

In addition, based on the work of Martin (2001) and Oskamp (1965), it would be important to include a query about the participants' confidence level in a future study. It would also be useful to request information about how they arrived at their diagnosis. This may provide important information about the process that participants go through in making the confirmation bias and it may alter participants' responses, either increasing or decreasing the rate of the confirmation bias. For example, it may signal clinicians to re-read the information provided, or to revisit their diagnoses before deciding how confident they are in their responses.

Based on the work of Jonas et al. (2001), who point out the increase in the rate of the confirmation bias when information is presented and processed sequentially instead of simultaneously, it might be useful to conduct a study comparing the sequential versus simultaneous presentation of clinical data to clinicians. However, this does not simulate the real world therapeutic environment where clinicians necessarily receive information in a sequential way from patients.

Another suggestion for a future study is to present the information from time one and time two with a shorter time period in between. For example, leaving one to three days between time points may reduce the incidence of the confirmation bias, as clinicians may not be so attached to their initial diagnosis. These studies may aid in learning more about pathways for clinicians to avoid making the confirmation bias.

Fugelsand et al. (2004) demonstrated that with continuous experience whereby new data does not fit with an initial theory, participants changed their initial theory to include the new inconsistent data. Another logical follow up study would extend the time points at which clinicians have the opportunity to alter their initial diagnoses to determine whether or not the same trend would occur among clinicians at one, two and three weeks re-test.

Next, a study conducted with clinicians and client-actors would be useful to determine whether the confirmation bias exists when the “client” is sitting face to face with the clinician. In this scenario, the clinician would have the advantage of asking additional questions rather than being limited to the information provided by a short vignette, as this scenario is more akin to real life clinical practice. In addition, if the participants in this study evidenced the confirmation bias in part due to reading the vignettes very quickly, this face-to-face scenario would likely correct for these errors, as they would have the time to ask more in depth questions.

Additionally, as clinicians appeared to be more accurate with the psychotic disorder vignette as compared to the anxiety disorder vignette, another study may incorporate several different vignettes which cover several categories of mental illness to

determine whether one type of disorder leads clinicians to evidence the confirmation bias more often.

Next, it would be useful for a follow up study to incorporate a more balanced sample. Specifically, as this study recruited clinicians from listservs that tend to attract more scientifically minded clinicians, another study could recruit clinicians from listservs that attract a more balanced sample of clinicians who are likely to be both cognitive behavioral as well as psychodynamic in orientation.

Finally, it would be important to include technological advances such as computerized audio or video media that are now available to track things such as the amount of time clinicians spend reading the clinical information as well as the amount of time clinicians take to assign a diagnosis. This may provide insight as to whether or not time spent reading clinical notes and assigning diagnoses was related to the confirmation bias.

Most importantly, there is a great need for future studies to focus on finding remedies for the confirmation bias that can be easily applicable to the real world. Given the work of Neighbors et al. (2003) and Trierweiler et al. (2000), who have identified some of the problems of the confirmation bias within the clinical context such as over-diagnosis of African Americans with Schizophrenia, it is important that a remedy be found quickly. This is not merely a problem of misdiagnosis or a labeling dilemma. The consequences of the confirmation bias lead to misdiagnosis. However, often that misdiagnosis has its own ramifications such as receiving different or lesser treatment from health care professionals, never mind the exclusionary treatment from the general society. In addition, what may have been a straightforward treatment for one disorder

may be convoluted by ignoring disconfirming evidence and making a commitment to confirmatory information. It is especially important given the frequency with which the confirmation bias is being made. In this study alone, the confirmation bias was evidenced 33% of the time which is far too frequent.

To summarize, the options for changes to future analogue studies are to include a.) using a structured clinical interview, b.) using various forms of the confirmation bias instructions, c.) assessing further information regarding the clinicians' professional training and discipline, d.) asking about their training in biases and heuristics, e.) asking about the thought process behind their approach to psychodiagnosis, or how it is that they arrived at their diagnosis f.) asking about participants' confidence level, g.) examining simultaneous versus sequential presentation of clinical data, h.) decreasing the amount of time in between the presentation of the vignettes at time one and time two, i.) extending the number of follow-ups to allow for a decrease in the confirmation bias over time, j.) having client-actors serve as the base of information rather than use vignettes, and finally k.) investigating whether one type of disorder in the vignette pulls for the confirmation bias more than others.

Finally, there are both strengths and weaknesses in the design, data collection strategy, and external validity in this study. Analogue studies are useful tools to investigate the process of psychodiagnosis. Whereas this is an analogue study, or a first step in investigating the effect of these confirmation bias instructions, there are not only several options for changes and additions to future analogue studies, but there are many avenues open for future research to more closely replicate real world scenarios as well. It

is important to address these concerns, as all mental health fields would benefit to work toward more accuracy during the psychodiagnostic process.

The confirmation bias is an important problem which needs immediate attention. The results from this study are an indication of just how powerful and ubiquitous the confirmation bias is. Based on the work of Neighbors (2003), Loftus (1998), McNally (2001), as well as the results of the present study, the confirmation bias is evidenced far more than is acceptable, and the ramifications of this bias are far reaching and potentially detrimental to both clinicians as well as clients. Given the problems stemming from the confirmation bias and the frequency with which this bias appears to be made in the process of psychodiagnosis, we have an urgent need to find a remedy to counteract this bias.

List of References

Anderson, R. (1970). The History of Witchcraft: A Review with Some Psychiatric Comments. *American Journal of Psychiatry*, *126*, 1727-1735.

Arkes, H. (1981). Impediments to Accurate Clinical Judgment and Possible Ways to Minimize Their Impact. *Journal of Consulting and Clinical Psychology*, *49*, 323-330.

Arkes, H. (1991). Costs and benefits of judgment errors: Implications for debiasing. *Psychological Bulletin*, *110*, 486-498.

Arkes, H., Christensen, C., Lai, C., & Blumer, C. (1987). Two methods of reducing overconfidence. *Organizational Behavior and Human Decision Processes*, *39*, 133-144.

Bissett, R. & Hayes, S. (1999). The Likely Success of Functional Analysis Tied to the DSM. *Behaviour Research and Therapy*, *37*, 379-383.

Coolidge, F. & Segal, D. (1998). Evolution of Personality Disorder Diagnosis in the *Diagnostic and Statistical Manual of Mental Disorders*. *Clinical Psychology Review*, *18*, 585-599.

Crews, F. & Bulkeley, K. Dialogue with a skeptic. Bulkeley, K. (Ed). (2001). *Dreams: A Reader on Religious, Cultural and Psychological Dimensions of Dreaming*. (pp. 361-376).

Dittman, V. (1991). Modern Psychiatric Classification in Research and Clinical Practice. *Archives Suisses de Neurologie et Psychiatrie*, *142*, 341-353.

Eva, K.W., (2003). The influence of the differentially processing evidence on diagnostic decision making. *Dissertation Abstracts International: Section B: The Sciences and Engineering*. *63(9-B)*, 4393.

Evans, J., Newstead, E., Allen, J., & Pollard, P. (1994). Debiasing by Instruction: The case of belief bias. *European Journal of Cognitive Psychology*, *6*, 263-285.

Fabrega, H. (1991). Culture and History in Psychiatric Diagnosis and Practice. *Cultural Psychiatry: International Perspectives*, *24*, 391-405.

First, M.B., Spitzer, R.L., Gibbon, M., & Williams, J.B. (1996). Structured Clinical Interview for DSM-IV Axis I Disorders-Patient Edition (SCID-I/P, Version 2.0). New York: Biometrics Research Department, New York State Psychiatric Institute.

- Fleming, G. W. T. H. (1933). The Revision of the Classification of Mental Disorders. *Journal of Mental Science*, 79, 753-757.
- Fleming, G.W.T.H. (1934). Definitions and Explanatory Notes on the Classification of Mental Disorders. *Journal of Mental Science*, 80, 409-410.
- Folletee, W. & Houts, A. (1996). Models of Scientific Progress and the Role of Theory in Taxonomy Development: A Case Study of the DSM. *Journal of Consulting and Clinical Psychology*, 64, 1120-1132.
- Frances, A. & Ross, R. (2001). *DSM-IV-TR Case Studies: A Clinical Guide to Differential Diagnosis*. American Psychiatric Association, Washington, D.C.
- Friedlander, M., Phillips, S. (1984). Preventing Anchoring Errors in Clinical Judgment. *Journal of Consulting and Clinical Psychology*, 52, 366-371.
- Friedlander, M., & Stockman, (1983). Anchoring and publicity effects in clinical judgment. *Journal of Clinical Psychology*, 39, 637-643.
- Fugelsang, J., Stein, C., Green, D., Dunbar, K., (2004). Theory and Data Interactions of the Scientific Mind: Evidence from the Molecular and the Cognitive Laboratory. *Canadian Journal of Experimental Psychology*. 58(2), 86-95.
- Galinsky, A., & Moskowitz, G. (2000). Perspective-taking: Decreasing stereotype expression, stereotype accessibility, and in-group favoritism. *Journal of Personality and Social Psychology*, 78, 708-724.
- Garb, H. (1998). *Studying the Clinician: Judgment research and Psychological Assessment*. American Psychological Association, Washington D.C.
- Garb, H. (1986). The Appropriateness of Confidence Ratings in Clinical Judgment. *Journal of Clinical Psychology*, 42, 190-197
- Garb, H. & Boyle, P. (1976). Understanding Why Some Clinicians Use Pseudoscientific Methods: Findings from Research on Clinical Judgment. In Lilienfeld, S., Lynn, S., Lohr, J. (Eds.), *Science and Pseudoscience in Clinical Psychology*. The Guilford Press, New York.
- Gold, L. (2002). Psychiatric Diagnoses and the Retrospective Assessment of Mental States. In D. Shuman, (Ed.), *Retrospective Assessment of Mental States in Litigation: Predicting the Past* (pp. 335-367). Washington, D.C. American Psychiatric Publishing.
- Goncalves, O., Machado P.P., Korman, Y., Angus, L. (2002). Assessing Psychotherapy: A Narrative Approach. In Beutler, L. & Malik, L. (Eds.), *Rethinking the*

DSM: A Psychological Perspective. American Psychological Association, Washington, D.C.

Gorman, M. & Gorman, M. (1984). A comparison of disconfirmatory, confirmatory and control strategies on Wason's 2-4-6 task. *The Quarterly Journal of Experimental Psychology*, 36A, 629-648.

Green, E., Flynn, M., & Loftus, E. (1982). Inducing Resistance to Misleading Information. *Journal of Verbal Learning and Verbal Behavior*, 21, 207-219.

Grob, G. (1991). Origins of DSM-I: A Study in Appearance and Reality. *The American Journal of Psychiatry*, 148, 421-431.

Guaron, E. & Dickinson, J. (1969). The Influence of Seeing the Patient First on Diagnostic Decision Making in Psychiatry. *American Journal of Psychiatry*, 126, 85-91.

Haverkamp, B. (1993). Confirmatory Bias in Hypothesis Testing for Client-Identified and Counselor Self-Generated Hypotheses. *Journal of Counseling Psychology*, 40, 303-315.

Herbert, J., Hope, D., & Bellack, A. (1992). Validity of the Distinction Between Generalized Social Phobia and Avoidant Personality Disorder. *Journal of Abnormal Psychology*, 101, 332-339.

Herbert, D., Nelson, R., & Herbert, J. (1988). Effects of Psychodiagnostic Labels, Depression Severity and Instructions on Assessment. *Professional Psychology: Research and Practice*, 19, 496-502.

Highhouse, S. & Bottrill, K. (1995). The Influence of Social (Mis)Information on Memory for Behavior in an Employment Interview. *Organizational Behavior and Human Decision Processes*, 62, 220-229.

Hill, C. & Ridley, C. (2001). Diagnostic Decision Making: Do Counselors Delay Final Judgments? *Journal of Counseling and Development*, 79.

Hirt, E. & Markman, K. (1995). Multiple Explanation: A Consider- an- Alternative Strategy for Debiasing Judgments. *Journal of Personality and Social Psychology*, 69, 1069-1086.

Holmes, C. (1967). A Somatic Interpretation of the Psychiatry of Benjamin Rush. *American Journal of Psychiatry*, 124, 133-139.

Houts, A. & Galante, M. (1985). The Impact of Evaluative Disposition and Subsequent Information on Clinical Impressions. *Journal of Clinical Psychology*, 3, 201-212.

- Huebner, S. (1990). The Generalizability of the Confirmation Bias Among School Psychologists. *School Psychology International*, 11, 281-286.
- Jonas, E., Schultz-Hardt, S., Frey, D., (2001). Confirmation bias in information seeking under simultaneous vs sequential information presentation. *Journal of Experimental Psychology*, 48(3), 239-247.
- Klerman, G. (1986). Historical Perspectives on Contemporary Schools of Psychopathology. In Millon, T. & Klerman, G. *Contemporary Directions in Psychopathology, Toward the DSM-IV*. The Guilford Press, New York.
- Knoff, W. (1970). A History of the Concept of Neurosis, with a Memoir of William Cullen. *American Journal of Psychiatry*, 127, 120-125.
- Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 107-118.
- Kramer, H., and Sprener, L.: *Malleus Maleficarum* (1486), trans by M. Summers. London: John Rodker 1928.
- Kroll, J. (1973). A Reappraisal of Psychiatry in the Middle Ages. *Archives of General Psychiatry*, 29, 276-283.
- Kurtz, R. & Garfield, S. (1978). Illusory Correlation: A Further Exploration of Chapman's Paradigm. *Journal of Consulting and Clinical Psychology*, 46, 1009-1015.
- Langer, E., & Abelson, R. (1974). A Patient By Any Other Name...Clinicians Groups Difference in Labeling Bias. *Journal of Consulting and Clinical Psychology*, 42, 4-9.
- Lehmann, H. (1977). Classification of Depressive States. *Canadian Psychiatric Association Journal*, 22, 381-390.
- Lilienfeld, S. & Marino, L. (1995). Mental Disorder as a Roschian Concept: A critique of Wakefield's "Harmful Dysfunction" Analysis. *Journal of Abnormal Psychology*, 104, 400-411.
- Lilienfeld, S., & Marino, L. (1999). Essentialism Revisited: Evolutionary Theory and the Concept of Mental Disorder. *Journal of Abnormal Psychology*, 108, 3.
- Loftus, E. (2003). Make-Believe Memories. *American Psychologist*, 58, 867-873.
- Lord, C., Lepper, M., & Preston, E. (1984). Considering the Opposite: A Corrective Strategy for Social Judgment. *Journal of Personality and Social Psychology*, 47, 1231-1243.

- Martin, J. (2001). Confirmation bias in the therapy session: The effects of expertise, external validity, instruction set, confidence and diagnostic accuracy. *Dissertation Abstracts International: Section B: The Sciences and Engineering*. (61) 7-B), 3851.
- Matarrazo, J. (1983). The Reliability of Psychiatric and Psychological Diagnosis. *Clinical Psychology Review*, 3, 103-145.
- Meehl, P. (1973). *Psychodiagnosis: Selected Papers*. University of Minnesota Press, Minneapolis.
- Mezzich, J., Kirmayer, L., Kleinman, A., Fabrega, H., Parron, D., Good, B., Lin, K., & Manson, S. (1999). The Place of Culture in DSM-IV. *The Journal of Nervous and Mental Disease*, 187, 457-464.
- Mumma, G., & Steven, W. (1991). Procedural debiasing of Primacy/Anchoring Effects in Clinical-Like Judgments. *Journal of Clinical Psychology*, 51, 841-853.
- Mynatt, C., Doherty, M., & Tweney, R. (1977). Confirmation Bias in a Simulated Research Environment: An experimental Study of Scientific Inference. *Quarterly Journal of Experimental Psychology*, 29, 85-95.
- Mynatt, C. Doherty, M., & Tweney, R. (1978). Consequences of Confirmation and Disconfirmation in a Simulated Research Environment. *Quarterly Journal of Experimental Psychology*, 30, 395-406.
- Neighbors, H., Trierweiler, S., Ford, B., Muroff, J. (2003). Racial Differences in DSM Diagnosis Using a Semi-Structured Instrument: The Importance of Clinical Judgment in the Diagnosis of African Americans. *Journal of Health and Social Behavior*, 44, 237-256.
- Newstead, S., Pollard, P., Evans, J., & Allen, J. (1992). The Source of Belief Bias in Syllogistic Reasoning. *Cognition*, 45, 257-284.
- Nickerson, R. (1998). Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology*, 2, 175-220.
- Oskamp, S. (1965). Overconfidence in Case-Study Judgments. *Journal of Consulting Psychology*, 29, 261-265.
- Renner, C., & Renner, M. (2001). But I thought I knew that: Using Confidence Estimation as a Debiasing Technique to Improve Classroom Performance. *Applied Cognitive Psychology*, 15, 23-32.
- Rogler, L. (1997). Making Sense of Historical Changes and Statistical Manual of Mental Disorders: Five Propositions. *Journal of Health and Social Behavior*, 38, 9-20.

- Rosen, G. (1968). *Madness in Society. Chapters in the Historical Sociology of Mental Illness*. The University of Chicago Press, Chicago.
- Sandifer, M., Hordern, A., & Green, L. (1970). The Psychiatric Interview: The Impact of the First Three Minutes. *American Journal of Psychiatry*, 126, 92-97.
- Scheinbaum, B. (1979). Psychiatric Diagnostic Error. *Schizophrenia Bulletin*, 5, 560-563.
- Schoeneman, T. (1977). The Role of Mental Illness in the European Witch Hunts of the Sixteenth and Seventeenth Centuries: An Assessment. *Journal of the History of the Behavioral Sciences*, 13, 337-351.
- Silverman, B. (1992). Modeling and Critiquing the Confirmation Bias in Human Reasoning. *IEEE Transactions on Systems, Man, and Cybernetics*, 22, 972-982.
- Slovic, P. & Fischhoff, B. (1977). On the psychology of experimental surprises. *Journal of Experimental Psychology*, 3, 544-551.
- Snyder, M., & Campbell, B. (1980). Testing Hypotheses About Other People: The Role of the Hypothesis. *Personality and Social Psychology Bulletin*, 6, 421-426.
- Snyder, M., & Swann, W. B. Jr. (1978). Hypothesis-testing Processes in Social Interaction. *Journal of Personality and Social Psychology*, 36, 1202-1212.
- Snyder, M., Tanke, E., & Berschied, E. (1977). Social Perception and Interpersonal Behavior: On the Self- Fulfilling Nature of Social Stereotypes. *Journal of Clinical Psychology*, 35, 656-666.
- Spengler, P., Strohmer, D., Dixon, D. & Shivy, V.(1995). A Scientist-Practitioner Model of Psychological Assessment: Implications for Training, Practice and Research. *The Counseling Psychologist*, 23, 506-534.
- Spitzer, R. (1983). Psychiatric Diagnosis: Are Clinicians Still Necessary? *Comprehensive Psychiatry*, 24, 399-411.
- Spitzer, R., Gibbon, M., Skodol, A., Williams, J., & First, M. (1994). *DSM-IV Case Book: A Learning Companion to the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition*. American Psychiatric Association, Washington, D.C.
- Spitzer, R., & Wilson, P. (1968). A Guide to The American Psychiatric Association's New Diagnostic Criteria. *The American Journal of Psychiatry*, 124, 41-51.
- Temerlin, M. (1968). Suggestion Effects in Psychiatric Diagnosis. *The Journal of Nervous and Mental Disease*, 147, 349-353.

Trierweiler, S., Neighbors, H., Munday, C., Thompson, E., Binion, V., & Gomez, J. (2000). Clinical Attributions Associated with the Diagnosis of Schizophrenia in African American and Non-African American Patients. *Journal of Consulting and Clinical Psychology, 68*, 171-175.

Tversky, A. & Kahneman, D. (2004). Judgment Under Uncertainty: Heuristics and Biases. *Science, 185*, 1124-1131.

Van Overwalle, F., & Labiouse, C. (2004). A Recurrent Connectionist Model of Person Impression Formation. *Personality and Social Psychology Review, 8*, 28-61.

Wakefield, J.C. (1992). The Concept of a Mental Disorder: On the Boundary Between Biological Facts and Social Values. *American Psychologist, 47*, 373-388.

Whaly, A. L., & Geller, P. A. (In press, March 1997). Toward a Cognitive Process Model of Ethnic/ Racial Biases in Clinical Judgment. *Review of Clinical Psychology*.

Widiger, T., Frances, A., Pincus, H., Ross, R., First, M., Davis, W., & Kline, M. (1988). *DSM-IV Sourcebook, Volume 4*. American Psychiatric Association, Washington, D.C.

Wilson, M. (1993). DSM-III and the Transformation of American Psychiatry: A History. *American Journal of Psychiatry, 150*, 399-410.

Wilson, T., Centerbar, D., & Brekke, N. (2002). In T. Gilovich, & D. Griffin (Eds.), *Heuristics and Biases: The Psychology of Intuitive Judgment*. (pp. 185-200). New York, NY. US: Cambridge University Press.

Zilboorg & Henry, (1941). *A History of Medical Psychology*. W.W. Norton & Company, Inc., New York, N.Y.

Figure 1.
Research Design

		Baseline	1-Week Retest
Bias Sensitization Group	Diagnosis Should Change	Panic Disorder → or Schizophreniform Disorder→	Social Anxiety Disorder or Alcohol Induced Psychotic Disorder
	Diagnosis Should not Change	Schizophreniform Disorder→ or Panic Disorder→	Schizophreniform Disorder or Panic Disorder
No Bias Sensitization Group	Diagnosis Should Change	Panic Disorder→ or Schizophreniform Disorder→	Social Anxiety Disorder or Alcohol Induced Psychotic Disorder
	Diagnosis Should not Change	Panic Disorder→ or Schizophreniform Disorder→	Panic Disorder or Schizophreniform Disorder

Note: This figure represents the research design indicating the possible vignettes each clinician received by change condition and confirmation bias instruction condition.

Figure 2.
Accurate Diagnoses by Vignette

		Baseline		1 Week Retest
Set A	No Confirmation Bias Instructions	Ms. A	Schizophreniform	Schizophreniform
		Mr. B	Panic	Social Anxiety Disorder
Set B	Confirmation Bias Instructions	Ms. A	Schizophreniform	Alcohol Induced Psychosis
		Mr. B	Panic	Panic
Set C	Confirmation Bias Instructions	Ms. A	Schizophreniform	Schizophreniform
		Mr. B	Panic	Social Anxiety Disorder
Set D	No Confirmation Bias Instructions	Ms. A	Schizophreniform	Alcohol Induced Psychosis
		Mr. B	Panic	Panic

Note: This table represents the accurate diagnosis in each condition.

Table 1.
Gender

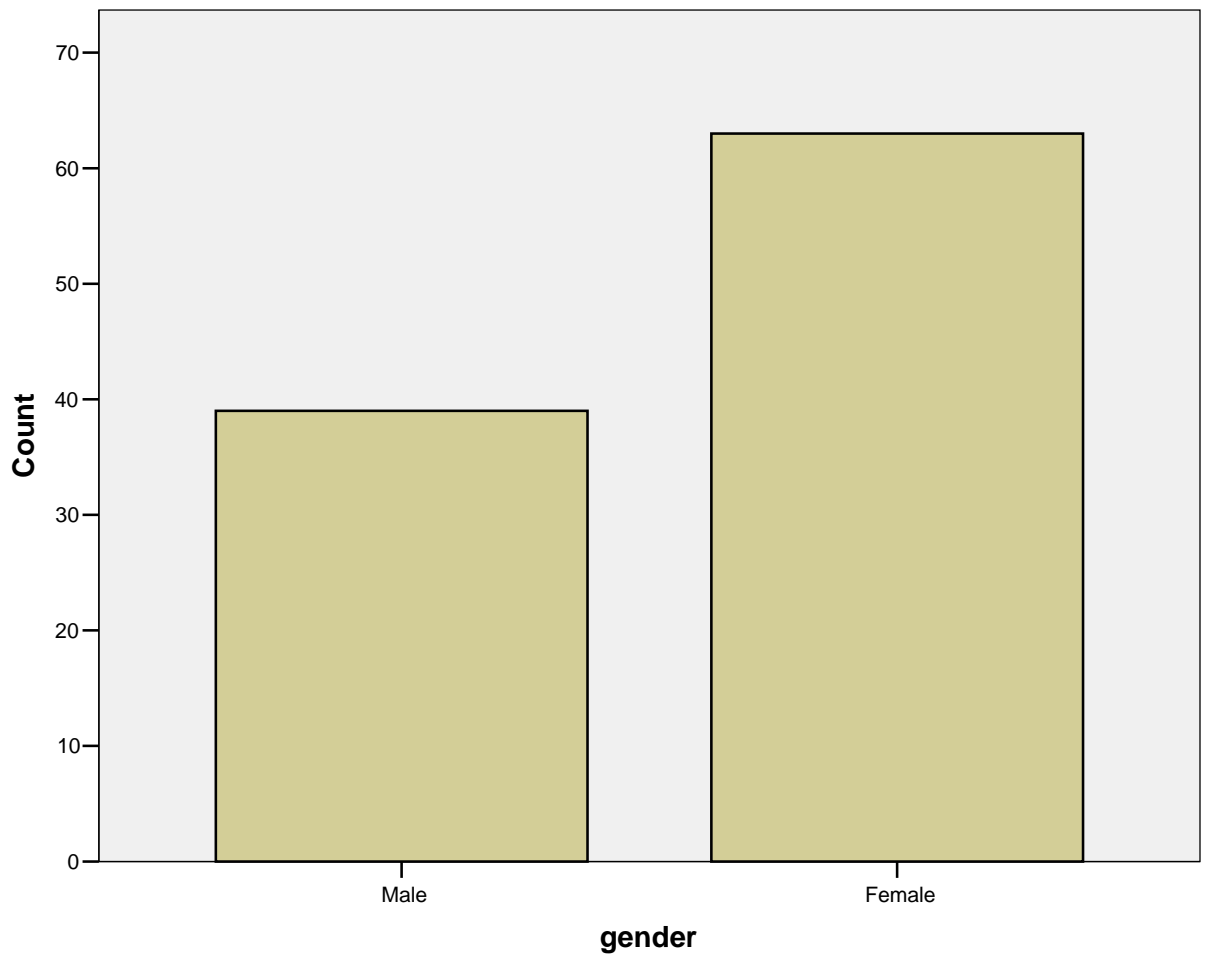


Table 2.
Age

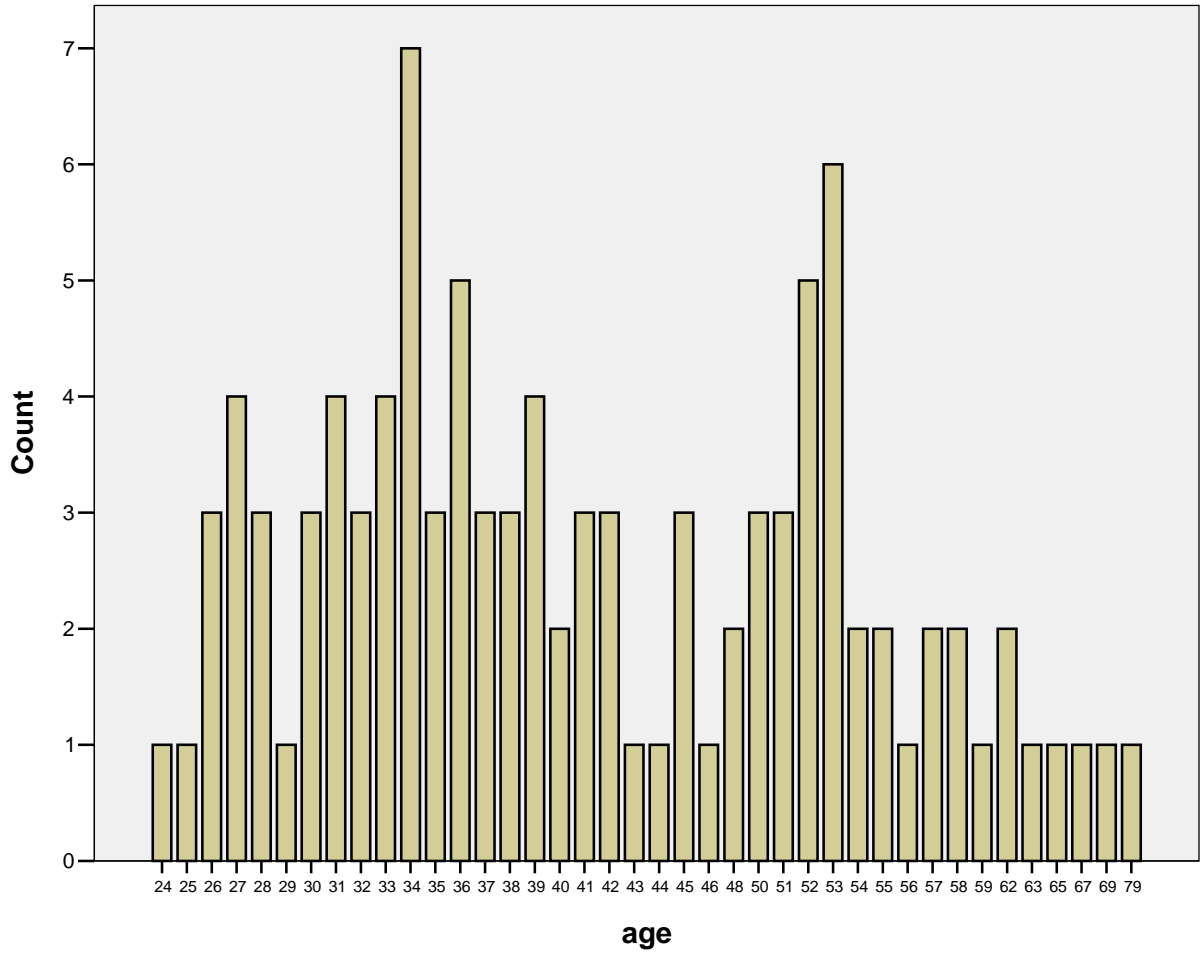


Table 3.
Years in Practice

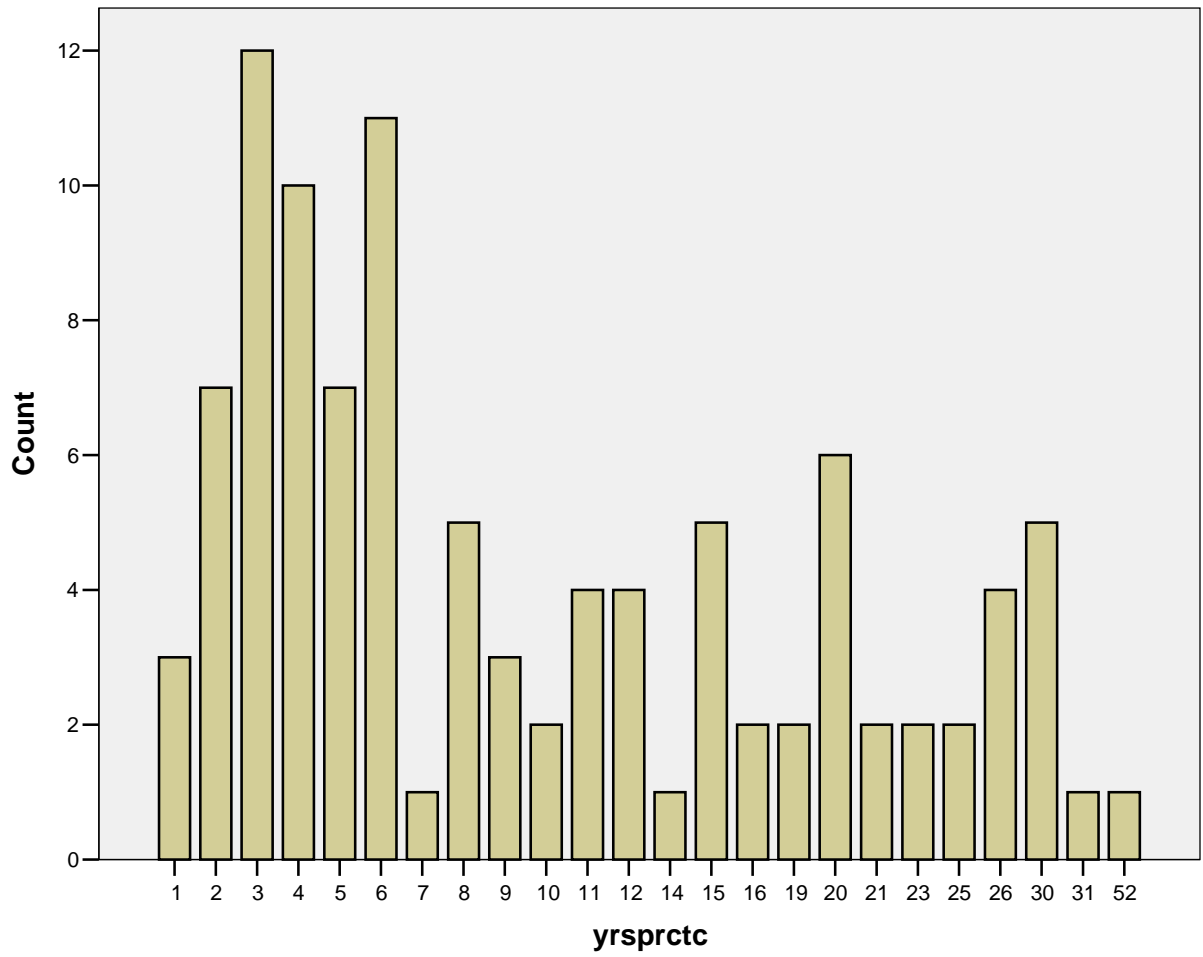


Table 4.
Orientation

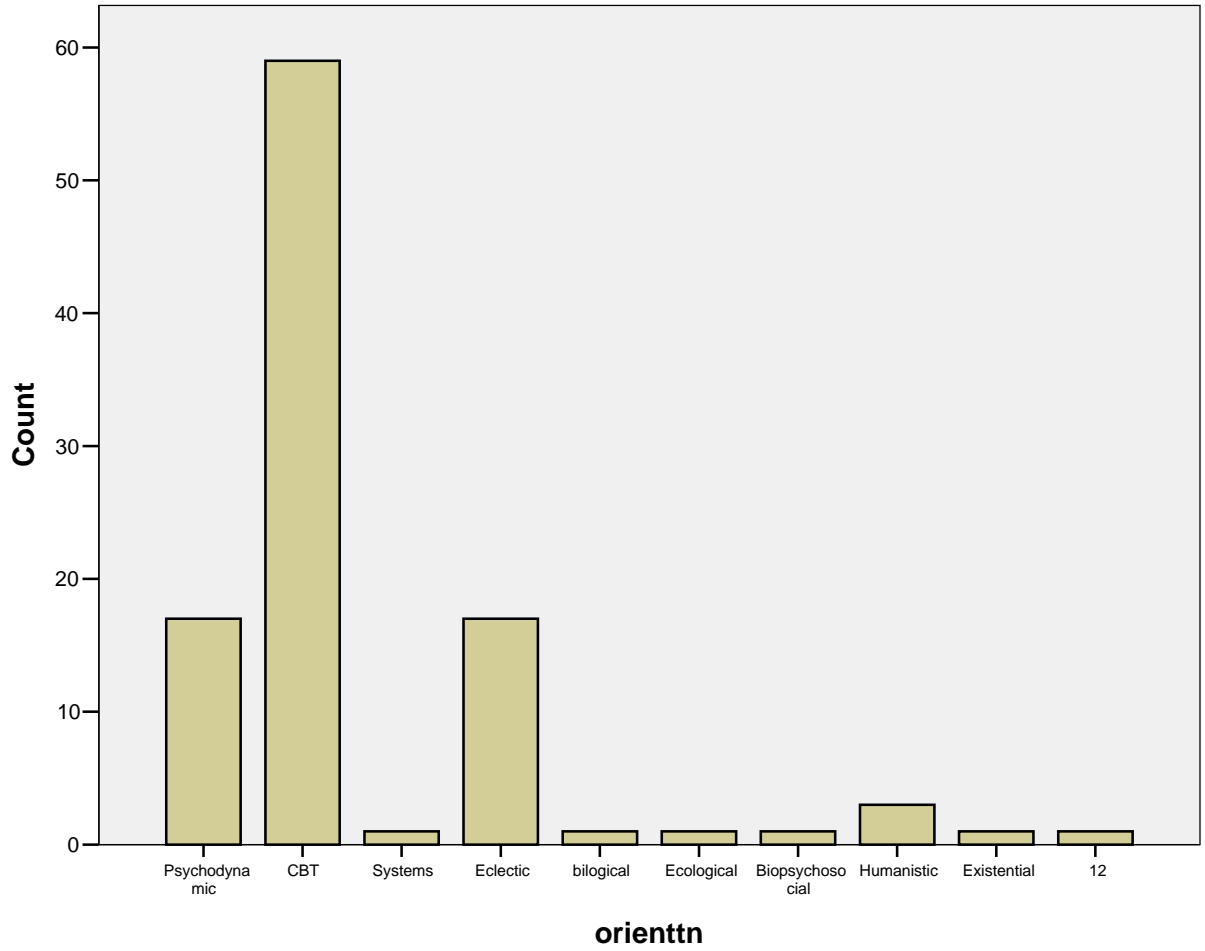


Table 5a.
Percentage of Change in Each Condition

	No Change in Diagnosis	Change in Diagnosis
No Confirmation Bias Instructions	3/49 Changed (6.1%)	33/49 Changed *(67%)
Confirmation Bias Instructions	3/53 Changed (5.7%)	41/53 Changed ** (77)

Note: This table represents the percentage of time that clinicians changed their diagnosis from time one to time two in each condition regardless of whether or not they *should* have changed their diagnosis based on the information presented.

* This cell represents the difference between 100% and 67% (33%) or, the percentage of time the clinicians evidenced the confirmation bias without confirmation bias instructions.

** The difference between 100% and 77% (23%) represents the percentage of time these clinicians continued to evidence the confirmation bias even with the confirmation bias instructions.

Table 5b.
Percentage of Change in Each Condition by Vignette Type

	No Change in Diagnosis	Change in Diagnosis
No Confirmation Bias Instructions	Psychotic Vignette 0/24 Changed (0%)	Psychotic Vignette 21/25 Changed *(84%)
	Anxiety Vignette 3/25 Changed (12%)	Anxiety Vignette 12/24 Changed *(50%)
Confirmation Bias Instructions	Psychotic Vignette 0/28 Changed (0%)	Psychotic Vignette 22/25 Changed ** (88%)
	Anxiety Vignette 3/25 Changed (12%)	Anxiety Vignette 19/28 Changed ** (68%)

Note: This table represents the percentage of time that clinicians changed their diagnosis from time one to time two in each condition regardless of whether or not they *should* have changed their diagnosis based on the information presented.

* The difference of 100% and 84% is 16% and the difference between 100% and 50% is 50%, or the percentage of time clinicians evidenced the confirmation bias without confirmation bias instructions.

** The difference between 100% and 88% is 12% and the difference between 100% and 68% is 32%, or the percentage of clinicians who made the confirmation bias even with confirmation bias instructions.

Table 6.
Accuracy in Diagnoses Across Condition

	No Change in Diagnosis	Change in Diagnosis
No Confirmation Bias Instructions	46/49 Accurate (94%)	33/49 Accurate *(67%)
Confirmation Bias Instructions	50/53 Accurate (94%)	41/53 Accurate *(77)

Note: * It appears as though the confirmation bias instructions resulted in a trend for participants to perform more accurately when a change in diagnosis was required; however this effect did not reach statistical significance.

Table 7.
Accuracy in Diagnoses Across Condition by Vignette Type

	No Change in Diagnosis	Change in Diagnosis
No Confirmation Bias Instructions	Psychotic Vignette 24/24 Correct (100%)	Psychotic Vignette 21/25 Correct (84%)
	Anxiety Vignette 22/25 Correct (88%)	Anxiety Vignette 12/24 Correct (50%)
Confirmation Bias Instructions	Psychotic Vignette 28/28 Correct (100%)	Psychotic Vignette 22/25 Correct (88%)
	Anxiety Vignette 22/25 Correct (88%)	Anxiety Vignette 19/28 Correct (68%)

Note: This table represents the percentage of accuracy in each condition specific to vignette type.

Appendix A: Listserv Administrator Email

Dear Listserv Administrator:

We are interested in collecting data from licensed mental health professionals who are currently involved in the practice of psychotherapy. With your permission, we would like to distribute information about our study through your listserv. We are interested in the general clinical impressions of mental health professionals. The study will involve reading several clinical case vignettes and recording clinical impressions for each vignette. Participants will read two clinical vignettes initially, and then one week later, they will have the option to log on to our website, www.clinicalimpressionresearch.com, to complete the study. At that time they will read two additional vignettes and record several follow up impressions. *Participation in the study will take approximately 5-10 minutes.* At the end of the study, participants will have the option of viewing several points to consider in the process of psychotherapy. They will also have the option of viewing the study results at our website:

(<http://www.psychology.drexel.edu/anxietyresearch/pages/>) once the study has been completed.

If you are willing, we would be grateful if you would send the following message to the members of your listserv (see below). This message contains a link for interested participants to access directions for how to participate in the study. If you do send the message below, please cc Meagan Carleton at mec38@drexel.edu on the email that you send to your listserv.

If you have any questions, please do not hesitate to contact our project coordinator, Meagan Carleton, M.S., at (215) 762-3327, or by email at mec38@drexel.edu.

Sincerely,

James Herbert, Ph.D.
Associate Professor
Director of Clinical Training
Drexel University

Appendix B: Recruitment Letter

Dear colleague,

We are currently conducting a study of the clinical impressions of licensed mental health professionals currently engaged in the practice of psychotherapy, and we would be grateful if you would consider participating in this project. The study involves reading several clinical case vignettes and recording clinical impressions for each vignette. Participation involves reading two clinical vignettes initially, and then one week later, signing onto our website to complete the second part of the study. At that time you will read two additional vignettes and record several follow up impressions. Participation in the study will take only approximately 5-10 minutes, and all responses are anonymous.

At the end of the study, you will have the option of viewing several points to consider, in the process of psychodiagnosis. You will also have the option of viewing the study results at our lab website, (<http://www.psychology.drexel.edu/anxietyresearch/pages/>) once the study has been completed. If you have any questions, please do not reply to this email. Rather, please contact our project coordinator, Meagan Carleton, M.S., at (215) 762-3327, or by email at mec38@drexel.edu. If you are willing to participate, please simply click on the link below to read more information about the study and to participate. Thank you for your consideration.

Sincerely,

James Herbert, Ph.D.
Associate Professor
Director of Clinical Training
Drexel University

<http://www.clinicalimpressionresearch.com>

Appendix C: Page One of Data Collection Website
www.clinicalimpressionresearch.com

Clinical Impressions of Practicing Psychotherapists

Principle Investigator: James D. Herbert, Ph.D.

Co-Investigator: Meagan Carleton, M.S.

Department of Psychology,

Drexel University

(215) 762-1692

Study Purpose: The primary purpose of this study is to better understand the process of making clinical decisions in the mental health field.

Study Procedures: Participation in this study will take approximately 5-10 minutes. We are interested in your impressions of two clinical cases, “Ms. A, and Mr. B.” You will be asked to read two case vignettes at two time points (today and approximately one week from today) and complete a short (1 page) form after reading each vignette. Please only participate if you are willing to take part at these two different time points. The first set of vignettes are linked to this form (click on “I agree to participate” at the bottom of this page to read the case vignettes.) You will then be asked to enter a five digit identification number and this will serve as your user identification number throughout the duration of the study. This will be the only means of identification, and therefore, all responses will remain confidential and anonymous. Upon reading each vignette and completing the short forms, simply press the “submit” button, and your responses will be submitted to a secure database. Immediately upon receipt of your response, James D. Herbert, Ph.D., or a research assistant will look up your response and will only use your selected five digit user identification number to identify your information. This website and our server are secure sites. However, if you are concerned about submitting your data via the internet, you may print the vignettes and the forms and mail in a paper copy. The mailing address is available at the end of this form. Surveys received by mail will be treated with the same protocol and confidentiality as surveys received via the internet. Approximately one

week after you complete the initial questionnaires, please log on to our website again and enter in your five digit identification number. Once you enter your five digit user identification number, you will be linked to a secure database that will give you access to a) the initial vignettes you read, b) your initial clinical impressions of those vignettes and, c) additional clinical information in the form of two additional vignettes for each “patient.” That is, there will be two additional vignettes that will provide additional information on each of your “patients.” Finally, you will be asked to complete two final forms asking for your clinical impressions of these vignettes.

Risks: Despite using password protected, secure databases and servers to keep all data confidential and anonymous, there is always a slight possibility that data may be revealed through transmission over the internet.

Confidentiality: All information obtained is strictly confidential and anonymous. Only the Principle Investigator (James D. Herbert, Ph.D.) and research assistants directly working on the project will have access to the data. All data will be kept in a locked cabinet at Drexel University. Electronic data will be stored on a computer in the Department of Psychology at Drexel University and will be password protected. All information will be reported in group form and will remain strictly anonymous.

Compensation: There is no monetary remuneration for participation. However, upon completion of the study you will have the opportunity to view the general findings, which will be posted to our lab website

(<http://www.psychology.drexel.edu/anxietyresearch/pages/>). In addition, immediately upon completion, you will have the opportunity to review several points to consider when forming clinical impressions.

Contact: If you have any questions or desire further information with respect to the study, please contact Meagan Carleton, M.S. by phone at (215) 762-3327 or by email at mec38@drexel.edu.

Agreement to participate: I am aware that I have the right to ask questions and receive answers to any questions related to this study. I am aware that I have the right to refuse to participate or may withdraw from the study at any time. I also realize that all information is strictly confidential. I understand that there is a possibility that despite the best efforts of the researchers to ensure that all data is kept completely confidential and anonymous,

there is a possibility that data may be revealed while being transmitted over the internet. I am aware that I have a second option to mail in the questionnaire if I prefer. I am aware that, if the forms are completed, it will be assumed that I have agreed to participate. I am also aware that by reading this cover page and by continuing on with the survey I am not waiving any of my legal rights. I understand that if I have any questions or concerns regarding my treatment or rights as a research participant, I may contact the Office of Research at Drexel University by phone at (215) 895-5849.

Please print this page and keep for your records and reference.

I have read and understand the contents on this form and I agree to participate.

Please randomly select one, and only one, of the following Sets: A,B,C or D. It doesn't matter which one you select.

- [Set A](#)
- [Set B](#)
- [Set C](#)
- [Set D](#)

If you're logging in to complete the second part of the study at Week Two, please [Click here](#).

James D. Herbert, Ph.D. or Meagan Carleton, M.S.
Department of Psychology
Drexel University
245 N 15th Street, MS 988
Philadelphia, PA 19102-1192

Appendix D: General Instructions Provided to Both Instructional Groups at Time 1

Thank you for your interest in participating!

We are interested in your clinical impressions of two clinical cases, “Ms. A and Mr. B.” Please read the information provided in the following case vignettes. Then complete the attached forms inquiring about your general clinical impressions. Please read these vignettes and record your clinical impressions as if you were receiving this information during the early phase of assessment of a client. As we are interested in your own general clinical impressions, please do not consult with the DSM or other materials in making your judgments.

Prior to beginning the study, please chose a five digit number and please make sure to write it down. This is the only way that we will be able to identify your data, as this number will serve as your user identification number. This number will also serve as your password when you log onto our website one week from today.

Appendix E: General Instructions Provided to Both Instructional Groups at Time 2

Principle Investigator: James D. Herbert, Ph.D.

Co-Investigator: Meagan Carleton, M.S.

Department of Psychology,

Drexel University

(215) 762-1692

Dear Clinicians,

Thank you for taking a few moments to complete the clinical exercise last week. As mentioned in the materials that you received last week, we are interested in clinicians' general clinical impressions of two clinical cases, Ms. A. and Mr. B.

Please take a few moments to read the following case vignettes, which are an extension of the vignettes you read last week. That is, you can think of these new vignettes as information that you might gather in a follow up session from last week. After reading each vignette, please complete the clinical impression rating forms. If you would like to refer back to the first vignettes you read, you may do so. Once you click the "Submit" button below, you will be able to access the vignettes you read last week, as well as your initial clinical impressions. If you would like, please feel free to revisit these materials to refresh your memory.

This is the final part of this exercise. Please understand that this is not a test of your clinical abilities. We are only interested in your general clinical impressions and, therefore, we ask that you do not consult with any materials such as the DSM. If you have any questions, please feel free to contact Meagan Carleton, M.S., project coordinator, at (215) 762-3327, or via email at mec38@drexel.edu. Again, thank you for your time.

Sincerely,

James D. Herbert, Ph.D.

Appendix F: Bias Sensitization Instructions

These instructions are provided at time 2 only, and only for the bias sensitization group (sets B and C); based on the instructions provided by Friedlander and Phillips, 1984):

Before completing the clinical exercise, please read the following information regarding the confirmation bias.

Previous research has found that when people complete tasks like these, they often display several biases, one of which is known as the confirmation bias. That is, they tend to stick closely to their initial opinion even if additional evidence is inconsistent or even contradictory to the initial information. In sum, people often do not adjust enough from their initial judgments when important new information is provided.

One example of the confirmation bias occurs when a therapist, like yourself, assigns a diagnosis to a client after the first session based on all information available at that time. At a later session, the therapist finds out some important information that should lead them to change the diagnosis. If the therapist is subject to the confirmation bias, they may not change the diagnosis. Rather, he or she may either ignore the new information, or even reinterpret it to fit with the initial diagnosis. This type of error has been found in a number of studies.

One reason that this bias might occur is that people like to remain consistent. This tendency toward being consistent may keep you from revising your diagnosis. When you complete the next step in this study, please do everything you can to avoid this bias. To avoid this bias, consider whether the new information provided warrants a change from your previous diagnosis. When you complete the next step in this study, please make an attempt to avoid this bias. To avoid this bias, consider whether the new information provided warrants a change from your previous diagnosis.

Appendix G: Case Vignettes

Schizophreniform Disorder: Part I

Ms. A: Session 1

Ms. A. is a 26 year old divorced mother of two. Her children are 2 and 4 years old. She reports suffering from hallucinations and delusions for the past 12 weeks. She lost her job three weeks ago soon after she called in sick to work everyday for two weeks. At that time, her boss called and told her that if she did not come to work, she would lose her job. She returned to work the following day; however, her behaviors were notably bizarre. She told her co-workers that she knew that they had been scheming against her and were trying to get her fired. She refused to use the phone at her desk as she was convinced that her boss was recording all her conversations and would relay this information to the FBI in an attempt to have her imprisoned. Her co-workers soon started to avoid her because of her bizarre behavior. For example, she repeated phrases that she had heard, including nonsense phrases. In addition, she tried to convince another co-worker to quit her job because “they were after her too.” She was fired shortly thereafter because of her disruptive and bizarre behavior. Shortly after she was fired, she started hearing voices telling her that she was going to be killed for causing other people to commit crimes. She also has very low energy and feels confused a great deal of the time. Currently, her ex-husband is trying to gain custody of their children, as he observed her leaving both children in the house alone while she went out for several hours. Ms. A. reports no memory of this and feels that her ex-husband is taking part in the plot to try to

place her in jail. She recently confided in a friend that her ex-husband and her ex-boss have been making videotapes of her wherever she goes and is giving them to the FBI.

Schizophreniform Disorder: Part II

Ms. A: Session 2

Two days ago Ms. A.'s sister brought her to the emergency room shortly after they spoke over the phone. When Mrs. A. spoke with her sister on the phone, she informed her about her ex-husband and her boss making videotapes of her. Her sister was alarmed after hearing what Ms. A. was saying, and suggested they go to get some help together.

Once in the emergency room, Ms. A. told her sister and the crisis team about her upstairs neighbors, on whom she has been fixated for the past 12 weeks. She reported that during the time that she was home sick from work, the neighbors started harassing her. When her sister probed a bit more, she reported that they had been "accessing" her thoughts and repeating them to her. She reported that she was not as disturbed by people accessing her thoughts as she was about not being able to control the process. She reported that she has been so upset by the neighbors harassing her that she has gone up to bang on their door and yell at them several times a day. She reported that she has also considered hiring someone to "teach them a lesson," although she is concerned for the safety of her neighbors' son, who has nothing to do with the harassment. Ms. A. reported that she believes that most people are able to access others' thoughts; however, she believes that she just recently developed this ability herself, as she was a "late bloomer." When asked about her first experience with others accessing her thoughts, she reported that she was in her kitchen making dinner approximately 3 months ago, and she heard the voices of

people in the street reciting the entire dinner menu. When asked why she thought that this was a real experience and not only in her mind, she reported that when she looked out the window, she actually saw people in the street looking up at her as they repeated the menu.

Alcohol Induced Psychotic Disorder Part II

Ms. A: Session 2

Ms. A. reported feeling especially distressed this past week due to the possibility that she may lose custody of her children. She reported that she may go to the police to report her husband and his friends for videotaping her every move. When asked about her history, Ms. A reported that two years ago, she began drinking alcohol to “calm her nerves.” As her tolerance increased, she gradually increased her drinking from 1-2 drinks per day to 4-6 drinks per day. She tried to stop drinking on two occasions and was successful; however, she returned to drinking on both occasions. Whereas she has been drinking for the past two years, Ms. A reported a dramatic increase in her drinking that started approximately 12 weeks ago.

This past month, Ms. A. went to court in an attempt to keep custody of her children. The week before the custody hearing, she was extremely nervous about going to court and decided to try to relax using her usual routine, which included drinking alcohol throughout the day. However, due to the especially stressful upcoming event, she drank even more than usual during the days before the court date. She later realized that she should not go to court smelling as though she had been drinking, however, and she decided to not to drink the day before. She entered the courtroom again feeling confused and started repeating phrases at random. She then believed that she was able to control

the judge's decision through her rate of breathing. She began to breathe long slow breaths through her nose and believed that if she continued this throughout the entire day, she would leave with custody of her two children. When she returned home after the hearing, she was very upset about the possibility of losing her children and started to drink. She drank excessively throughout most of the next day and the next night, and later took her children to a hotel because she believed that her ex-husband had set up a video-camera in her apartment. She paid in cash for the hotel and parked at the back of the hotel so that her ex-husband would not be able to trace her whereabouts.

Panic Disorder with Agoraphobia: Part I

Mr. B: Session 1

Mr. B. is a 23 year old unemployed accountant who has become increasingly incapacitated by severe physical symptoms during which he suddenly feels as though he may be having a heart attack. He reports an increase in heart rate, sweating, shaking and nausea. The attacks started during his mid-teens, however they have increased in frequency during the past three years. When the attacks first started, he feared he might be dying. However, throughout the past year he believes that he may be going crazy, as there is no known physical cause to the attacks. He reports a moderate level of anxiety between attacks, as he fears that he may have another attack at any time. He has become so preoccupied with somatic feelings that he can no longer tolerate going out in public unless he is accompanied by his mother. In addition, when he has experienced these symptoms in public, he has felt as though he had humiliated himself and has since tried to avoid all public situations. He has experienced these symptoms for the past six years, although he feels that his symptoms have become increasingly worse over the past year.

Lately, he has not left his home for over four months and will only leave in case of an emergency. He recently went to his primary care physician who prescribed Ativan and although he has been taking it as prescribed, he continues to refuse to leave his home.

Panic Disorder with Agoraphobia: Part II

Mr. B: Session 2

Mr. B is the only child of older parents, who expected to be childless in their late thirties when he was born. His parents were loving and caring, although they often did not have the energy to take him out to play or have other children over to play. Since that time, he has always felt more comfortable with adults than with peers and preferred to stay at home rather than go out with friends, even when he was in high-school. When he was a teenager, he never told anyone about the attacks because he feared that he may “find out the worst.” That is, he feared that he may be dying. Instead, he kept the attacks to himself, hoping they would pass very quickly. This was difficult to do, especially if he had an attack while at school or out with friends. On the rare occasion that his friends or a teacher noticed that something was wrong, he would report that he was not feeling well and that he just needed to get some air. The attacks always passed within ten minutes and he was able to return to what he was doing. Currently, Mr. B. feels as though he is defective and inferior and although he would like to get help from a psychologist, he is fearful that if he leaves his home he will have another attack. Mr. B. reported that he has recently had these attacks while at home as well, especially when he is thinking of situations that he may encounter if he leaves his house. This has been problematic for him, as he has tried to prepare himself for leaving his home by imagining a trip outside of his home from beginning to end.

Mr. B. reported that the Ativan has helped to decrease his symptoms after an attack starts, although he reports that he does not want to be dependent on the Ativan, having to carry it with him at all times.

Social Anxiety Disorder Part II:

Mr. B: Session 2

As a child, Mr. B was always shy and did not want to be left alone with babysitters. During his childhood, Mr. B. developed mild school refusal and was never willing to try summer camp. During adolescence, he was interested in dating, but he was usually too shy to initiate his own relationships with women, although at times he went out on dates that his mother set up for him. Mr. B.'s symptoms have become increasingly worse throughout his early twenties. On occasion, he has tried to establish friendships and romantic relationships. However, each effort has ended in failure and humiliation because Mr. B. becomes anxious, ruminates that his is doing the wrong thing, and has an anxiety attack. He then finally gives up and returns to the "family routine," which he feels is safe. In the past year he has also avoided phone calls for fear that he may say something wrong and have an attack while on the phone. In addition, over the past year Mr. B. has started having attacks while he is at home, just thinking about social situations that he may encounter if he were to leave his house. His panic attacks only occur when he is anticipating or is actually engaged in some kind of interpersonal situation. He recognizes that he needs to start working soon, as his parents cannot support him forever. However, just thinking about being evaluated during a job interview brings on attacks as well.

Appendix H: Clinical Impression Record Form

Please take a few moments to complete the following:

I. Clinical Global Impression: Severity Rating

Clinical Rating: Please circle one

1. Normal, not at all ill
2. Borderline, mentally ill
3. Mildly ill
4. Moderately ill
5. Markedly ill
6. Severely ill
7. Among the most extremely ill patients

II. Diagnosis- 1st three axes (Please assign only 1 diagnosis per axis, i.e., the primary diagnosis)

I. _____

II. _____

III. _____

III. Medication recommendations- if any?

IV. Any further recommendations (aside from individual therapy)?

V. Potential obstacles to treatment (if any)?

CONTINUE

Appendix I: Demographic Form

Thank you for completing the previous exercise. Please take a moment to complete the following brief demographic questionnaire.

1.) Age: _____

2.) Gender: M F

3.) Please list your theoretical orientation (i.e., cognitive-behavioral, psychodynamic, etc.)

4.) Number of years practicing psychotherapy: _____

Appendix J: Confirmation Bias Information

Clinical Impressions of Practicing Psychotherapists

Principle Investigator: James D. Herbert, Ph.D., (215) 762-1692, jh49@drexel.edu

Co-Investigator: Meagan Carleton, M.S., (215) 762-3327, mec38@drexel.edu

Points to Consider In the Process of Psychodiagnosis

Research has demonstrated that there are several factors to consider that may impact the accuracy of psychodiagnosis. Some of these include a) unsystematic interviewing and data collection methods, b) lack of use of a multi-modal assessment, c) overconfidence in one's judgements and d) problems associated with reliance on memory. Some of these issues are relatively straightforward to address, such as using semi-structured clinical interviews to systematize data collection, using multi-modal assessment strategies, and taking extensive notes. However, other factors that may be more difficult to correct are the natural information processing heuristics that may bias the diagnostic process. One such bias in particular that has been shown to impact psychodiagnosis is the confirmation bias, which is known as the seeking or interpreting of evidence in ways that are partial to existing beliefs, expectations, or hypotheses. In other words, without necessarily being aware of it, people often search for evidence that confirms existing beliefs or hypotheses, while discounting disconfirming evidence. This bias raises obvious problems in situations in which judgment is crucial, such as psychodiagnosis. In order to be more certain that a hypothesis is accurate, confirming evidence alone does not provide the evidence needed. Even if disconfirming evidence is not found, of course, we cannot be certain that the hypothesis is "true." Nevertheless, we can be more confident in our conclusion than if only confirming evidence is examined.

Not only has the confirmation bias been shown to be operative in everyday interactions with students (e.g., Synder, Tanke, & Berscheid, 1977), it has also been shown to occur when clinicians are involved in the practice of psychodiagnosis. For example, Temerlin (1968) found that the mere presence of a label affected the way that an experimental group assigned a diagnosis to their patients, suggesting that this group did not search for disconfirming evidence. In addition, Haverkamp, (1993) found that counseling trainees tested hypotheses in a confirmatory manner 64% of the time and only tested hypotheses in a disconfirmatory manner 15% of the time. These studies are consistent with a much larger body of research that suggests that unless specifically trained to look for disconfirming evidence, people (including clinicians) are likely to fall prey to the confirmation bias.

One factor that may contribute to the apparently high frequency of the confirmation bias in the clinical setting is the speed with which some clinicians have been shown to make a diagnosis. For example, Gauron and Dickinson (1969) found that psychiatrists often made a diagnosis within 30 seconds of observing an interview with a client. This alone does not necessarily suggest a problem; however, Elstein, Shulman and Sprafka (1978) found that the most accurate clinicians were those who made a diagnosis later relative to those who made diagnostic decisions quickly.

Some research suggests that the confirmation bias can be resisted by subjects in a laboratory task (Green, Flynn, & Loftus, 1982; Highhouse & Bottrill, 1995; Tweeney, Doherty, Worner, Pliske, Mynatt, Gross & Arkkelin, 1980). Based on this research, our study seeks to investigate whether clinicians evidence a confirmation bias in an analogue diagnostic assessment task. That is, we are interested in whether clinicians will alter their

diagnostic judgements based on new data, i.e., in the vignettes provided at time point two, when the latter provide data that could disconfirm the initial diagnosis. In addition, we are interested in whether explicit instructions regarding the confirmation bias will mitigate its effects. Finally, we are interested in whether clinicians will alter an initial diagnosis when they are warned about the confirmation bias even when the information provided does not indicate that a change in diagnosis is warranted. In other words, we are interested if an unnecessary correction of the diagnoses will occur after being warned about the confirmation bias.

Although we do not yet know the results of the study, we would like to offer several suggestions to consider, based on the available literature in this area, on avoiding the confirmation bias in psychodiagnosis.

The first point to consider is to make a habit of routinely looking for disconfirming evidence. A second point is not to make a diagnosis too quickly on the basis of limited data. Refraining from making diagnostic judgments following a single assessment session, if possible, can increase diagnostic accuracy. A third point is to keep in mind is that case conceptualizations and diagnoses should always be viewed as “works in progress,” subject to revision, as new information is available. A fourth point is to consider all of the symptoms that the client actually endorses. Arkes and Harkness (1980) found that even making a diagnosis can change the way that a clinician remembers symptoms. For example, a clinician may think that a patient, who has been assigned a diagnosis of a major depressive episode, actually reported particular symptoms such as decreased appetite or decreased energy, even when they did not do so. The clinician may erroneously believe that the patient endorsed these symptoms because they are consistent

with the diagnosis. A useful strategy for addressing this tendency is to take careful notes. Finally, it is unwise to believe that more experience necessarily protects against the confirmation bias. In fact, some research suggests that trainees are just as accurate as experienced clinicians (Hammond, Hirsch & Todd, 1964). In addition, overconfidence that sometimes accompanies experience may actually lead to making these errors more frequently.

Please feel free to contact the investigators if you have any questions. In addition, the results of the study will be posted to our website at

<http://www.psychology.drexel.edu/anxietyresearch/pages/>.

Thank you very much for your participation!

Vita

MEAGAN CARLETON PARMLEY, Ph.D.

Education

Doctor of Philosophy: Clinical Psychology, Drexel University (formerly MCP Hahnemann University), Philadelphia, PA, October 2006, GPA 3.83/4.0

Master of Science: Clinical Psychology, Drexel University, Philadelphia, PA, May 2003, GPA: 3.89/4.0

Bachelor of Arts: Psychology, with a Minor in Sociology, Clark University, Worcester, MA, August 1996-May 2000, GPA: 3.55

Clinical Experience

Post Doctoral Fellow, Cognitive Behavioral Institute, LLC, Albuquerque, NM. October, 2006- present

Southwest Consortium Pre-doctoral Psychology Internship, Albuquerque, NM August, 2005- August 2006

Counselor, Drexel University Student Counseling Center, Center City Hahnemann Campus, Philadelphia, PA, June 2003- August 2005

Program Therapist, Anxiety Treatment and Research Program, Department of Psychology, Drexel University (formerly MCP Hahnemann University), Philadelphia, PA. August 2001- August 2005

Staff Therapist, Weight and Eating Disorders Program, University of Pennsylvania, Philadelphia, PA, April 2002- September 2003

Therapist, Depression in Cardiac Patients, MCP Hahnemann University Hospital, Philadelphia, PA, November, 2002-June, 2003

Research Experience

Project Coordinator, Anxiety Research and Treatment Program, Drexel University, Philadelphia, PA. June, 2004- August, 2005

Data Manager, Social Anxiety Treatment Program, Drexel University, Philadelphia, PA, May, 2003-August, 2004

Research Associate, Social Anxiety Treatment Program, Department of Clinical and Health Psychology, MCP Hahnemann University/ Drexel University, Philadelphia, PA, August 2001- August, 2005

Research Associate, Student Counseling Center, Center City Hahnemann Campus, Philadelphia, PA, January, 2004 – July, 2005

Research Assistant, Providence Family Study, Brown University, Providence, RI, August, 2000- May, 2001

Research Assistant, Families Through Time Study, Clark University, Worcester, MA, April 1998- May, 2000

Research Assistant, Family Play and Development during the Toddler Years, Clark University, Worcester, MA, November, 1996 – September, 1998

