2000

A Study of the Effect of Projective Drawings on the Hindsight Bias

Elizabeth A. Dufresne
University of Rhode Island

Follow this and additional works at: <https://digitalcommons.uri.edu/theses>

Recommended Citation

Dufresne, Elizabeth A., "A Study of the Effect of Projective Drawings on the Hindsight Bias" (2000). *Open Access Master's Theses*. Paper 1587.
<https://digitalcommons.uri.edu/theses/1587>

This Thesis is brought to you for free and open access by DigitalCommons@URI. It has been accepted for inclusion in Open Access Master's Theses by an authorized administrator of DigitalCommons@URI. For more information, please contact digitalcommons-group@uri.edu.

A STUDY OF THE EFFECT OF PROJECTIVE DRAWINGS ON
THE HINDSIGHT BIAS

BY

ELIZABETH A. DUFRESNE

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
MASTER OF ARTS
IN PSYCHOLOGY

UNIVERSITY OF RHODE ISLAND

2000

#45823534

HS4
D847
2000

Abstract

This study examined the effect of including a human figure drawing on the hindsight bias among undergraduate psychology students. Six groups were given a description of a 15 year-old boy who was experiencing school difficulties. Some of these groups also were given outcome information suggesting that the boy's difficulties were due either to social-emotional problems or to learning problems. In addition, some groups also were given a human figure drawing made by the boy. Participants then assigned likelihood probabilities to the two outcomes. Results of non-parametric analyses showed the hindsight bias for only one of the six groups. In contrast, results of parametric analyses showed a main effect for the outcome variable. Results are discussed in terms of previous research on the hindsight bias.

Acknowledgement

I could not have completed this thesis without the support of several people. I give special thanks to my major professor, Dr. Grant Willis, for his continued support and belief that I would successfully complete the project. I couldn't have done this without him. In addition, I am grateful to my other committee members, Dr. David Faust, Dr. Paul DeMesquita, and Dr. Suzie Kim, for their careful review of this work. Specifically, to Dr. David Faust, whose thoughtful feedback and advice will make me a better psychologist.

I must acknowledge my family and friends who have encouraged me during this process. Specifically, I am grateful to my good friend, Tim Heitzman, for the drawing of the man, and the endless conversations about the draw-a-person test. Most importantly, I thank my beautiful family, my husband, Russ, and my children, Sam, and Abby. They allowed me to keep things in perspective and encouraged me to continue with this work even when life interfered with my progress.

Table of Contents

Abstract	ii
Acknowledgements	iii
List of Tables	vi
List of Figures	vii
Introduction	1
DAP as a Measure of Cognitive Ability	2
DAP as a Measure of Social-Emotional Functioning	7
Hindsight Bias	19
Research Questions.....	27
Method	27
Participants	27
Case Materials	28
Results	33
Non-Parametric Analysis	33
Parametric Analysis	35
Discussion	36
Implications	41
Limitations	43
Conclusion	44
Tables	45
Figures	48

Appendices 54

 Appendix A - Consent Form 54

 Appendix B – Draw-a-Person 56

 Appendix C – Sample Statements from
 Pilot of Drawing 57

Bibliography 59

List of Tables

Table 1 Design of the Study 45

Table 2 Mean Probability Assigned to Each Outcome 46

Table 3 Analysis of Variance Summary Table for Comparisons of
Hindsight Bias and Projective Drawing Conditions 47

List of Figures

Figure 1	<u>Frequency of Percentages Assigned to Social- Emotional Problems - Foresight; Without Drawing</u>	48
Figure 2	<u>Frequency of Percentages Assigned to Social- Emotional Problems - Foresight; With Drawing</u>	49
Figure 3	<u>Frequency of Percentages Assigned to Social- Emotional Problems - Hindsight (learning); Without Drawing</u> .	50
Figure 4	<u>Frequency of Percentages Assigned to Social- Emotional Problems - Hindsight (s-e); Without Drawing</u>	51
Figure 5	<u>Frequency of Percentages Assigned to Social- Emotional Problems - Hindsight (learning); With Drawing</u>	52
Figure 6	<u>Frequency of Percentages Assigned to Social- Emotional Problems - Hindsight (s-e) With Drawing</u>	53

Introduction

A common argument for using human figure drawings as part of a psychological test battery is that, at best, they might provide some clinical information or at least help to establish rapport with a client, and, at worst, they might prove irrelevant (Fuller & Goh, 1983; Kennedy, Faust, Willis, & Piotrowski, 1994; Lubin, Larsen, & Matarazzo, 1994; Piotrowski, Sherry, & Keller, 1985; Piotrowski and Zalewski, 1993). Given the considerable research (Arkes, Wortmann, Saville, & Harkness, 1981; Chapman and Chapman, 1967; Fischhoff, 1975; Fischhoff and Beyth, 1975; Hawkins and Hastie, 1990) examining the effect of biases on the accuracy of clinical decisions, however, it is reasonable to wonder if including irrelevant data could potentially exacerbate bias. Because the hindsight bias is well documented, it could provide a useful methodology for assessing this hypothesis. Thus, the goal of the present study was to assess whether the magnitude of the hindsight bias would be affected by including a human figure drawing.

In the following review, the Draw-a-Person Test (DAP) first is discussed as a measure of cognitive ability and then of social-emotional functioning. Included is an examination of various techniques that have been developed to score the DAP. Next, a review of the research of the hindsight bias is presented. The original work of Fischhoff (1975) is described, as well as works that developed from his original paradigm. Finally, the two topics are integrated to form the research questions considered.

The DAP as a Measure of Cognitive Ability

Although study of the use of drawings to estimate intelligence extends back to the 1880s, the first standardized method was developed in 1926 by Florence Goodenough with her Draw-a-Man Test. Goodenough (1926) proposed that there was a relationship between children's concept development as seen in drawings and general intelligence. According to Goodenough, drawing, to a child, is language, and children draw what they know, rather than what they see. She observed that the developmental trends of drawings are remarkably constant among young children (ages 4 to 10 years). Finally, she observed that children with cognitive impairments produce drawings resembling those of younger children.

Based on these assumptions and observations, Goodenough developed an objective scoring system that was standardized using 4,000 drawings. Although the majority of her sample was from schools in New Jersey, several ethnic groups were represented from a variety of socio-economic backgrounds. The result of this study was a 51-point scale for measuring intelligence from human figure drawings.

Goodenough (1926) found inter-rater and test-retest reliabilities ranging from $r_s = .70$ to $.90$. Comparison of IQs estimated from the drawings with scores from the Stanford-Binet across separate age groups resulted in an average correlation of $r = .76$. In addition, prognosis for school success by teachers and scores from the drawings were compared and resulted in a correlation of $r = .60$. From this work, age norms were developed for children ages 4 to 10 years. It

was intended as a method of measuring nonverbal intelligence, particularly useful with non-English speaking and hearing impaired children.

In 1948, John N. Buck published a technique called the House-Tree-Person (H-T-P). He proposed that, in addition to measuring intelligence, drawings also could be used to measure personality factors. He postulated that the H-T-P measured cognitive function in a situation designed to activate non-intellective aspects of the personality that serve either to enhance or to diminish the efficiency of intellectual functioning. He further postulated that the method is a projective (or associative) one, portraying thoughts, feelings, and traits of the individual. To this end, he proposed that in order to interpret the drawings in terms of intellectual functioning, one also must interpret the projected personality aspects of the drawing.

In a revised edition of the manual, Buck (1966) described standardization studies that were done when developing the scale for measuring intelligence. The standardization sample of 120 participants was carefully selected and ranged from what today would be termed severely mentally retarded to intellectually gifted. Buck described the classification procedure in this way: "The ultimate criterion for inclusion in a level of intellectual function was the clinically demonstrated level of intellectual function and not a score on one or more standard intelligence tests" (Buck, 1966, p. 8). These drawings then were analyzed, and a scoring system was developed. The resulting quantitative scoring system was outlined in 40 pages in the manual, followed by another 100 pages describing qualitative interpretation, which already had been identified as a

requirement of interpretation. In terms of the concurrent validity of the method, Buck reported correlations ranging from $r_s = .40$ to $.70$ with scores from the Stanford-Binet and the Wechsler-Bellevue, with higher correlations restricted to lower IQ groups (mean IQ = 70). There was no explanation of how estimates were obtained; as noted, classifications were not based on standardized tests, but instead on clinical presentation. Overall, this method is difficult to use and lacks the empirical evidence to support its clinical and practical utility in a psychological test battery.

The Goodenough-Harris Draw-a-Person Test (Harris, 1963) was an attempt to extend the age norms into adolescence and to devise different forms of the scale to include a woman figure, as well as a drawing of the self as a possible method for studying the emerging self-concept. Attempts to develop adequate norms with adolescents were not successful, providing support for Goodenough's original assertion that after the age of 10 years, drawings are less predictive of intelligence.

More recently, analyses were conducted using the Buck and Goodenough-Harris scoring systems (Abell, Heiberger, & Johnson, 1994). The researchers were interested in examining the relationship between these measures and scores on the Wechsler Adult Intelligence Scale-Revised (Wechsler, 1981). Their results showed moderate correlations between Performance and Full Scale IQs and estimated standard scores from the two scoring systems, with the Buck system performing slightly better. Further analysis of these relationships revealed significant differences between the

estimated scores and WAIS-R scores. Both the Buck and Goodenough systems significantly underestimated all three IQ scores of the WAIS-R (i.e., VIQ, PIQ, & FSIQ). Overall, the authors concluded that the question regarding the use of drawings as a measure of cognitive functioning in adults remains equivocal and warrants further research. This review did not refer to any of the previously mentioned concerns regarding the Buck scoring system.

To address some of the concerns previously discussed, Naglieri (1988) developed the Draw-a-Person: A Quantitative Scoring System (DAP:QSS). The purpose of the DAP:QSS was to provide a brief nonverbal measure of ability to be used either as part of a test battery or as a screening device. The scoring system used modern scoring criteria and was normed on a large stratified sample of 2,622 students across the United States ranging in age from 5 to 17 years.

Each individual in the standardization sample provided drawings of a man, a woman, and the self. Scores provided include both a composite standard score, as well as standard scores for single drawings based on a 64-item scale. Age-related changes showed the fastest increase between ages 5 and 9 years, and then increased less rapidly. Thus, norms were developed for each age range, with quarter-year intervals for 5 through 8-year-olds, and half-year intervals for 9 and 10-year-olds. Ages 11 to 17 years were collapsed into a single group because their means were quite similar. The internal consistency for the composite scale ranged from $r_s = .83$ to $.89$, whereas those for the individual drawings ranged from $r_s = .56$ to $.78$.

Test-retest reliabilities ranged from $r_s = .60$ to $.89$ across various age ranges for a four-week interval. These were calculated based on a small sample, so readers are cautioned to interpret them carefully. Both inter-rater and intra-rater reliability estimates were high, ranging from $r_s = .86$ to $.97$ for children in Grades 1 to 7. Older age ranges were not evaluated. The reported construct validity of the DAP:QSS was evaluated in two ways. First, the author suggested that there should be a developmental change in mean scores. They reported age-related increases in total raw scores for children ages 5 to 11 years. In order to establish concurrent validity, drawings were scored using the DAP:QSS and the Goodenough-Harris scoring system. Resulting correlations ranged from $r = .75$ to $.84$. For Grades Kindergarten through 3, correlations ranged from $r_s = .28$ to $.31$, and for Grades 4 through 12, ranged from $r_s = .19$ to $.27$.

Overall, the methodology used to develop the DAP:QSS was sound. It provides an estimate of non-verbal ability with good reliability. Caution should be used when using it with children over the age of 11, as its efficacy has not been established with adolescents. In addition, whereas the rigor with which it was developed is impressive, its overall concurrent validity remains questionable.

Kamphaus and Pleiss (1991) reviewed the DAP:QSS. They found that although this technique was adequately normed and showed good reliability, its concurrent validity with other comprehensive measures of children's intelligence was mediocre. Willis (1992) also provided a review of this method. He concluded that the DAP:QSS was a well-developed screening measure of general nonverbal ability. Willis cautioned potential users, however, that the DAP

was neither developed nor validated as a social-emotional assessment technique, and should not be used as such.

Over the past 75 years, there have been many techniques developed to measure intelligence with human figure drawings. These methods have demonstrated variable psychometric properties. In general, correlations are better for children of lower IQ levels, than average to above-average IQ levels. Review of various techniques leads to the conclusion that it is possible, using well-developed, adequately normed tests, to gather an estimate of non-verbal ability correlated with other standardized tests. These conclusions apply primarily to young children (ages 5-10 years), as the current body of research does not lend support for valid interpretation of drawings of older children and adults. Overall, reliability tends to be good with recent tests, whereas concurrent and construct validity have been shown to be less impressive.

The DAP as a Measure of Social-Emotional Functioning

In contrast to its use in cognitive assessment, the use of the DAP in social-emotional and personality assessment is based on the projective hypothesis (Frank, 1939). This hypothesis states that when an individual structures and organizes an ambiguous stimulus situation while creating a response, the given response will at least partially reflect some of the individual's personality traits. The more ambiguous the stimulus, the greater the sensitivity of the projection process. It assumes that the more ambiguous the stimulus, the less likely it will elicit defensive reactions by clients. Anastasi (1988) suggested that projective

test materials represent a mechanism by which individuals reveal their needs, anxieties, conflicts, and thought processes. Major features of projective techniques include the following: (a) they represent disguised testing procedures, (b) they represent a global approach to the appraisal of personality, and perhaps most importantly, according to proponents, (c) they reveal covert, latent, or unconscious aspects of personality (Anastasi, 1988).

Over time, Goodenough and others (Buck, 1948; Hammer, 1958; Machover, 1949) began to use the DAP to assess social-emotional and personality variables. Hammer (1958) proposed that figure drawings would allow individuals to draw pictures of their inner worlds, where beliefs and personality characteristics could be expressed, and strengths and weaknesses exposed. Handler (1996) proposed several advantages to Draw-a-Person techniques. For example, he stated that the DAP is an easy task for most individuals and that it usually elicits cooperation in the assessment process. He further suggested that because children with internalized disorders often do not demonstrate their difficulties overtly, the DAP could provide an assessment of their discomfort. In addition, it could prove useful with individuals who are inhibited and non-talkative, because it is a relatively nonverbal task. It also allows the clinician to observe the client's functioning during an unstructured situation, and when compared with performance on a structured task, the clinician may be able to determine the extent to which the client needs external structure in order to function. Finally, he stated that the DAP has few age and intelligence limitations, and can be used with individuals from various socio-economic and cultural backgrounds.

Projective drawings such as the DAP often are used to establish rapport during an interview or evaluation. Kennedy et al. (1994) surveyed school psychologists about their practice, and discovered that many clinicians state that they use projectives primarily as a method of generating hypotheses, with a smaller percentage suggesting they use them primarily to help establish rapport. Although these appear to be harmless uses of this technique, it is difficult to measure the impact these drawings may have on overall impressions and diagnostic decisions. For example, although well intended, clinicians may have limited awareness of the influence these drawings have on their diagnostic decisions. Aspel and Willis (1998) found that clinicians' appraisals of how information influences decisions often is inaccurate.

Although the purpose of projectives usually is disguised, research on the illusory correlation has demonstrated that people may have preconceived beliefs about the meaningfulness of certain signs. Illusory correlation refers to a false belief in the association between variables. It can reflect either a belief in an association when one is absent, an overestimation of the strength of the relationship, or a belief of a relationship in one direction when it is actually in the opposite direction. Chapman and Chapman (1967), in a classic study of the illusory correlation, found that both clinicians and non-trained college students made similar inferences about the meaningfulness of particular signs in human figure drawings. They speculated that these signs led to particular associations about personality characteristics, regardless of the lack of empirical support for the validity of those signs.

With regard to the use of the DAP as a diagnostic tool, illusory beliefs could exert a significant impact on attempts to mislead the clinician. Specifically, if the client has any intention or inclination to be less than truthful, it may not be detected in the DAP and in fact may lead the clinician to disregard other evidence that may suggest the client is attempting to fake. For example, if the client has beliefs about the meaningfulness of certain signs in drawings and the clinician has similar beliefs, then it would be relatively easy to draw a "pathological" drawing. In addition, given the projective hypothesis, the clinician would be confident that the drawing represents covert or latent characteristics of the individual, and may ignore objective indicators of faking.

Critique of the DAP as a projective test. Many methods of drawing interpretation have been developed; this discussion reviews some of the earliest work, and then considers how current research and study have used this information. In 1949, Karen Machover published her manual Personality Projection in the Drawing of the Human Figure. Its purpose was as a method of personality analysis based on the interpretation of human figure drawings. The premise outlined in the book was that the body is a vehicle for self-expression and that the drawing reflects the individual's self-perception. The picture represents the whole system of psychic values, specifically drives and needs. This theory was based on Machover's personal experience as well as those of colleagues she respected. Machover speculated that "all creative activity bears the specific stamp of conflict and needs pressing upon the individual who is creating" (p.4). Although clinical experience and observations are useful for

generating ideas, scientific study is required to demonstrate the relationship between these observations and reliable and valid interpretation.

In describing the method that should be employed when using her system, Machover cautioned that interpretation should be done only by advanced practice psychologists with extensive training in psychodynamic theory. In addition, she recommended that interpretation should not be made in isolation, but instead that it should be considered in light of the whole clinical picture of the individual. This recommendation suggests that interpretation is subjective, in that it varies depending on the individual. Naturally, this raises concern about the reliability and validity of the measure.

The manual does not refer to specific studies of the psychometric properties of the technique. Machover (1949) made reference to a comparison of drawing interpretations with Rorschach and handwriting analysis, stating that many of them were consistent. No explanation was offered about how this comparison was made, nor were any statistical analyses cited. In addition, Machover suggested that the method had been shown to be useful for prognostic purposes, yet she did not cite any studies evaluating this.

In the manual, Machover referred four times to the interpretation of pointy fingers as suggestive of overt aggression or paranoid repression of aggression, yet she failed to cite any validation of this sign beyond her own clinical experience. As will be shown in the discussion of other methods, pointy fingers is one of many signs that have been interpreted, yet the original source, as well as those thereafter, failed to offer evidence as to the validity of the sign.

Overall, the method suggested by Machover (1949) lacks the psychometric study required to demonstrate its efficacy. The manual does not mention any use of control groups when developing the technique. In fact, Machover suggested that the frequency of particular signs should not deter the clinician from interpreting them. As an example, she referred to the interpretation of conflict in the treatment of hands as indicating lack of confidence in achievement and social contacts. She stated that the fact that this is often seen in drawings suggests that there is a high level of competitiveness in society and that individuals who exhibit this sign should be interpreted as such. In other words, the base rate of a particular sign should not influence the clinician's interpretation of it. Clearly the lack of scientific validation, as well as inadequate control groups lead to serious concerns about the usefulness of this technique.

Urban (1963) produced another handbook for the interpretation of signs in drawings. It was reportedly a summary manual of the work of Buck (1948), Goodenough (1926), Hammer (1958), and Machover (1949). The author recommended that it be used only by those with background in dynamic personality theories to explore an individual's inner recesses, denied, unacceptable, repressed impulses. He stated, "Acute observers always have been able to detect emotional connotations in art work" (Urban, 1963, p. 1). Although the manual did not offer any information about specific studies of validity and reliability, it did advocate the interpretation of specific signs as indicative of pathology. For example, pointed fingers are said to be suggestive of aggression, and crossed eyes as warnings of severe mental illness (Urban, 1963, p. 61).

Jolles (1964) published A Catalog for the Qualitative Interpretation of the House-Tree-Person. The manual cites Buck (1948), Hammer (1958), and Machover (1949) as references. It offers specific interpretation of signs in pictures without referencing sources specifically. The manual did not refer to any psychometric study of the validity of the signs. Similar to the previously mentioned manual, the work suggests that large, spike-like hands are indicative of aggression and hostility. Without mention of specific study of this, the reader assumes that this sign (along with many others) has been translated from the original work of Machover (1949), Buck (1948), and Hammer (1958).

Goldstein and Rawn (1957) studied interpretive signs of aggression in human figure drawings. They hypothesized that experimentally induced feelings of aggression would elicit changes in figure drawings consistent with those signs suggested by Buck (1948) and Machover (1949) as indicators of aggression. These signs included line pressure, figure placement, and seven specific details generally interpreted as representing aggression (including spiked fingers, slash-line mouth, detailed teeth, clenched fists, nostril emphasis, squared shoulders, and toes in a non-nude figure). Their sample consisted of 39 male and female attendants in a state mental hospital. No other descriptors of participants were mentioned.

Participants were ranked based on measurement of line pressure from their first drawing and, based on these rankings, alternately assigned to the experimental or control groups. In order to induce feelings of aggression, members of the experimental group were required to wait before drawing their

second picture, and then were told they would have to work more hours with no salary increase. No standardized measure of aggression was given at this time, but was inferred based on "the multitude of spontaneous complaints" (p. 170).

Drawings before and after the aggression induction were compared, as well as the drawings of both of the control groups. They found that line pressure and figure size did not support the hypothesis as valid interpretive signs of aggression. They further reported that the seven sign indicators, when considered globally, did discriminate between the control and experimental groups, as well as before and after the aggression induction for the experimental group. They cautioned that none of the signs individually was a strong predictor, only the global ratings. Finally, they had two judges (trained psychologists) provide subjective impressions of aggression. These impressions proved to be poor predictors of members of the experimental group. In addition, these judges reported that they relied on line pressure and figure placement as indicators of aggression.

The results of this study must be interpreted cautiously for a variety of reasons. Of primary concern is the participant pool. Specific subject variables were not identified, including even basic descriptors such as age and gender distribution. The method by which aggression was measured also presents concern, because its validity was not assessed. Due to these kinds of methodological issues, at best these results offer equivocal support for the interpretation of drawings as indicative of aggression.

More recently, Naglieri, McNeish, and Bardos (1991) developed the Draw-a-Person: Screening Procedure for Emotional Disturbance (DAP: SPED). In contrast with previously mentioned methods that relied on the projective hypothesis, the objectives of this method were to make interpretations based on scientifically validated norms. The authors stated that it was intended as a scoring system made of items that easily and objectively can be scored. They also intended to develop a nationally normed system with demonstrated ability to differentiate between normal and disturbed populations. The reported use of the method was as a screening procedure to identify individuals for whom further evaluation is warranted. The authors cautioned that screening was the intended use and that it should not be used for diagnostic purposes.

The resulting scoring system was a 55-item scale including items taken from an extensive literature review, that occurred infrequently among normal individuals, and demonstrated good psychometric properties. Scoring rules are clear and objective. Global scores are obtained and compared to norms, using T scores. According to the manual, scores less than 55 suggest further evaluation is not warranted, scores of 55-64 suggest that further evaluation is indicated, and scores above 65 suggest that further evaluation is strongly indicated. One issue that should be considered regarding these cutoffs is the fact that using T scores, a score of over 55 (assuming a normal distribution) would represent over 30% of the population. The manual did not report the base rate in the population of children needing evaluation for emotional problems, but clearly 30% is unusually high.

The DAP:SPED was standardized with 2,260 individuals aged 5 to 17 years. Intrarater (over a one month time period) and interrater reliabilities were $r_s = .83$ and $.84$, respectively. Test-retest reliability was $r = .67$ over a one-week time interval. The validity of the system was assessed by four separate studies.

The first study compared mean scores of 81 students in special education (with emotional and behavioral disorders) with students from the standardization sample. Mean T-scores were 55.3 for the special education group and 49.5 for the standardization group. These scores were reliably different from each other, and the authors concluded that the DAP:SPED validly discriminated between these two groups.

The second study used a similar methodology, with the clinical sample drawn from a class for students with serious emotional disturbances. Again, results showed mean T-scores of 57 for the clinical sample and 49 for the standardization group, a reliable difference. The third study used similar methodology, with the clinical sample being drawn from students in special education classes for students with emotional disturbances. Comparison of mean T-scores for the clinical (54.8) and standardization (49.7) groups yielded a reliable difference. Again, the conclusion was drawn that the DAP:SPED was a good discriminator between the groups.

The fourth and final study was similar to the others. The clinical sample was a group of 54 students with serious emotional disturbances who were enrolled in a day school program. Mean T-scores for the clinical (56.6) and standardization (49.9) groups once again were reliably different, offering support

for the DAP:SPED's discriminatory power. This study was reviewed both in the manual and separately (Naglieri & Pfeiffer, 1992). Whereas the results reported in the manual are consistent with those in the article, the authors did not mention the fact that when group membership was assessed based on a T score cutoff of 55, 78% of the standardization group, yet only 48% of the clinical group were correctly identified.

Overall, the DAP:SPED is a carefully designed, well normed scoring system. It demonstrates good reliability, much improvement over its predecessors. Yet questions regarding its validity remain. For example, the authors caution many times in the manual that it is a screening procedure only, and should not be used for diagnostic purposes. They state clearly that it can help to determine what children may need further evaluation. Yet, review of the four studies reported in the manual, that used children with serious emotional disturbances in self-contained classrooms, demonstrated that these groups just barely reached the cutoff score of 55. Therefore, these studies suggest a minimal effect size for the clinical group. The clinical implications of the use of this method include the possibility that individuals who are in need of further evaluation may not score in the clinically significant range based on this scoring procedure and therefore will not receive further evaluation. The risk of false negative results is substantial, and clinicians should be aware of this when using this method. Conversely, the DAP:SPED is better at correctly identifying individuals who are not in need of further evaluation, with over 70% of participants in the standardization group scoring in the non-clinical range.

Riordan and Verdel (1991) have suggested that in addition to the DAP's use with a standardized scoring method, it also provides clinically useful information without a scoring system. They reviewed the use of the Draw-a-Person to predict sexual abuse in children. They did not refer to any specific scoring method or procedure to be used for interpretation. They reported specific signs in children's drawings that suggest that sexual abuse has occurred (e.g., eyes closed or without pupils, nose overemphasized, elongated neck). Yet, they offered only qualitative evidence and did not cite empirical support for these assertions. In addition, they advocated that the DAP could be used by non-trained individuals as an assessment measure of definite indicators of sexual abuse. Given the lack of empirical support for these assertions, this report should be interpreted cautiously.

Feyh and Holmes (1994) attempted to replicate early studies of the predictive power of particular signs on the DAP with children with conduct disorders. They used eight indicators of aggression identified by Koppitz (1966) and Machover (1949). Their results failed to replicate earlier reports, with no differences in the frequency of particular signs between children with or without conduct-disorders.

Although these techniques continue to be used frequently by clinicians in psychiatric clinical settings and in schools (Fuller & Goh, 1983; Kennedy et al., 1994; Lubin et al., 1984; Piotrowski et al., 1985; Piotrowski & Zalewski, 1993), questions about their reliability and validity persist. Those methods reviewed have been cited in more recent manuals for the interpretation of human figure drawings

(Ogdon, 1990). It is interesting to note that the original works suggested that further studies need to be conducted, yet the previously mentioned recent manual primarily referenced these sources.

Given the multitude of research questioning the usefulness of the DAP as a measure of social-emotional functioning, and yet the tendency for clinicians and non-trained individuals to interpret it based on intuitive beliefs, its inclusion in a psychological test battery raises serious concerns. Specifically, the risk of diagnostic errors is significant. In addition, it is the responsibility of professionals (e.g., school psychologists) to caution other team members about the risks of including the DAP in an assessment battery.

Diagnostic accuracy can be influenced by many different variables. For example, cognitive biases have been demonstrated to exert a negative effect on accuracy. One well-researched cognitive bias is the hindsight bias. The following discussion considers this and its potential effect on decision making.

Hindsight Bias

The hindsight bias is a phenomenon that has strong empirical support (Hawkins & Hastie, 1990). The term refers to the tendency to overestimate probability estimates for a given outcome once the outcome is known (Hawkins & Hastie, 1990). Individuals who are provided with outcome information and then asked to estimate how likely they would have estimated an occurrence tend to overestimate the likelihood that they would have applied to the given outcome. In contrast, individuals without outcome information tend to assign lower

probabilities to the same outcome. Fischhoff (1977) described this as a “knew-it-all-along effect” (p. 349) .

Study of the hindsight bias has reflected the original paradigm established by Fischhoff (1975). In his original study, he used four obscure events, two historical events, and two clinical-psychology cases. For each of these between-groups experiments he provided a brief description of the event followed either by a list of possible outcomes (foresight) or a sentence presenting the actual outcome (hindsight). Participants then were instructed to assign probabilities of likelihood of occurrence to each outcome as if they did not have outcome information.

Fischhoff (1975) used the nonparametric sign test for his analysis. Results showed that individuals with outcome information overestimated the probabilities they would have assigned if they had not had outcome information. Their beliefs about the accuracy of their decisions were, therefore, influenced by the presence of outcome information. Fischhoff proposed the term “creeping determinism” (p. 288) to describe a process by which outcome information immediately and automatically is integrated into a person’s knowledge about the events preceding the outcome. People, without realizing it, incorporate the outcome into a plausible explanation about why the end result had to occur. The effect of this, as represented in probability estimates, is to perceive the outcome as inevitable. It seems as though people seem unable to recapture their original estimates of likelihood of outcomes once they have outcome knowledge.

Following these results, Fischhoff and Beyth (1975) examined the effect of the hindsight bias on current news events. They varied the previous studies by using a within-groups design, in which participants assigned foresight probabilities. After the outcome was known, participants were asked to remember their original probabilities, as if they did not have outcome information. The instructions for this experiment explicitly stated "give the same probabilities which you gave then (two weeks ago)," (p. 5) thereby addressing the possibility that participants misunderstood what they were supposed to do.

Again, these researchers found a significant hindsight bias. Participants rated events they believed had occurred as more likely in hindsight than they had in foresight, and rated events they believed had not occurred as less likely in hindsight than they had in foresight. Careful examination of the results showed that 75% of the participants remembered having assigned higher probabilities than they actually had to events that they believed had happened. Moreover, 57% of participants reported lower probabilities for events they believed had not happened.

Arkes et al., (1981) attempted to replicate the work of Fischhoff with physicians making medical diagnoses. They used a case description of a frequently encountered medical problem, followed by four possible diagnoses. The physicians were required to assign likelihood estimates to each outcome, equaling 100. There were four hindsight groups, each one told that one of the four diagnoses was the true outcome. Likelihood estimates were examined using nonparametric analyses (sign test). Results demonstrated that the physicians

demonstrated the hindsight bias. Yet the bias was restricted to the two diagnoses assigned the lowest likelihood estimates in foresight. These results were consistent with previous research suggesting that the hindsight bias is strongest for events initially judged to be least plausible (Fischhoff, 1977; Wood, 1978).

Synodinos (1986) conducted a between-groups study of this bias using a gubernatorial race in Hawaii. The researcher was interested in several issues. First, he hypothesized that participants would be more confident and assign more accurate probability estimates of percentage of votes in hindsight than in foresight. In addition, he hypothesized that, in hindsight, people with greater political involvement would be motivated to distort their answers in the direction of the "knew-it-all-along" effect more than people with lower political involvement. This hypothesis relates to the role of self-esteem as a motivational factor in the bias.

Participants were required to fill out a questionnaire assessing their degree of interest, level of knowledge, and perceived importance of the election. They then were asked to assign a percentage of statewide votes that they believed each of three candidates would receive (or had received, in hindsight) in the election (estimates must equal 100% for all three candidates). Finally, the participants were asked to indicate their degree of confidence in their predictions. The researcher used parametric statistics to analyze the results. Although the comparisons were in the predicted direction, they did not consistently reach significance. Synodinos suggested that the participants' accuracy in their pre-

election (foresight) estimates allowed little room for distortion. Participants indicated a significantly higher level of confidence in hindsight than in foresight. The researcher suggested that subjective measures of confidence may be more sensitive measures of the hindsight effect, and suggested calling it a "sure-all-along" effect (p. 116).

Pennington (1981) questioned whether the effects found in laboratory studies would generalize to real-life, current, news events. He used a British firefighter's strike as the news event to be studied. As with the previous studies, Pennington provided participants with information about possible outcomes (e.g., terms of the settlement and length of strike), and elicited probability estimates in foresight and at two times in hindsight. Results supported Fischhoff's work in that the participants demonstrated the typical hindsight pattern of greater probability associated with the true outcome in hindsight compared with foresight. Also included in this study was a condition in which participants were required to generate their own possible outcomes in foresight. In this case, the typical hindsight bias was not demonstrated, although results suggested a trend in that direction. This is an early example of a technique used to decrease the bias. Finally, Pennington found that when he provided more detailed descriptions (several hundred words) to participants at the time the judgments were made, the hindsight effect was stronger than when a brief (150 words) summary of the strike events was presented.

Although the hindsight bias has been observed fairly consistently, these studies also suggest that there may be techniques to control or minimize the

effect. This early research provided a good framework for future studies and identified potential factors that could account for the hindsight bias, such as temporal setting (past or future), cognitive processes, and the possibility that when participants are required to generate spontaneous outcomes, the hindsight bias may diminish (higher level of processing of information). These factors stimulated further research (e.g., Arkes et al., 1988; Connolly & Bukszar, 1990; Creyer & Ross, 1993; Fischhoff, 1976; Pohl & Hell, 1996; Schkade & Kilbourne, 1991; Sharpe & Adair, 1993), and the latter concept has been useful in the consideration of debiasing techniques (e.g., Arkes, Faust, Guilmette, & Hart, 1988).

Arkes et al. (1988) evaluated a technique for eliminating the hindsight bias. They used the classic hindsight paradigm, using foresight and hindsight groups, and requested estimations of likelihood of diagnoses. They added an additional condition whereby some participants in each group were required to generate reasons, based on the case description, about why they assigned the probabilities they did. Results suggested that when required to provide reasons for decisions, participants' susceptibilities to the hindsight bias decreased. This study provides support for the assertion that there are techniques that can diminish the effects of the hindsight bias.

In a review of this literature, Hawkins and Hastie (1990) concluded that the basic effect of the hindsight bias, higher retrospective probabilities associated with reported outcomes, is supported across a variety of tasks, and varying time intervals (minutes to weeks) between initial judgments, outcome, and second

judgments. They suggest that the cumulative research of this phenomenon indicate, in most cases, that the hindsight bias can be observed. In addition, the implications for its effect on decision making can be substantial, because it can hinder the detection of errors, inflate errors, bias future decisions, and lead to unduly influenced second opinions.

In an effort to condense the wide body of research on the hindsight bias, Christensen-Szalanski and Willham (1991) conducted a meta-analysis of 122 studies. They stated that because the prevalence of the bias has been well established, they were interested in identifying personal or task characteristics that may moderate the level to which the bias is present, and in evaluating the practical significance of the bias. The variable analyzed was estimations of effect size, as this would demonstrate the degree to which the phenomenon was present in the population (Cohen, 1977, pp. 9-10). The actual test statistics were transformed to effect sizes, either as Cohen's d or the Pearson product moment correlation coefficient (r).

The study considered several moderator variables. The first variable was whether or not the outcome occurred. The researchers referred to Fischhoff and Beyth's (1975) observation that the hindsight bias seemed to be more pronounced when people were told that an event had occurred than when told that it had not occurred. They hypothesized that this was due to people's cognitive difficulty in processing negative information. The second moderator variable considered was the participants familiarity (experience or expertise) with the topic being studied. Two other moderator variables (problem difficulty and

event novelty) were discussed but not directly evaluated, because data were unavailable to examine them.

Results of this analysis found the average effect size of all 122 studies was $r = .17$, with a 95% confidence interval of $r = .14$ to $.20$. Therefore, they found support for the phenomenon of the hindsight bias, although its observed effect size was small. They further discovered that the size of the effect was correlated with both the participants' familiarity with the task and whether the outcome information stated that the event did or did not occur. Specifically, they discovered that the more familiar the participants were with the task, the smaller the effect size. In addition, in hindsight, people were less likely to reduce their likelihood estimates when told that an event did not occur, than they were to increase their likelihood estimates when told that an event did occur.

Post-hoc analyses uncovered an interesting phenomenon. They found that studies that used an unfamiliar task-event occur model were nearly all paper and pencil tasks using college undergraduates (in contrast with more professionals used in familiar task cases). Analysis of these undergraduate studies suggested that question format might be a moderating variable of the hindsight bias. Case-history problems generated an average effect size similar to the familiar task effect size (small effect), whereas almanac questions generated an average effect size that was nearly three times larger than any other observed in the meta-analysis. The present study used undergraduate students and the case-history model; therefore these observations are relevant here.

Research Questions

The purpose of this study was to see if the inclusion of projective information would increase the magnitude of the hindsight bias. A premise of the study was that projective techniques (e.g., DAP) are measures of limited clinical utility, and, in fact, actually may serve to decrease the overall accuracy of diagnostic decisions. The primary questions addressed were:

1. Will the probabilities assigned to reasons for difficulty be similar, regardless of whether outcome information is provided (replication of hindsight bias research).
2. To what extent is the hindsight bias influenced by the inclusion of projective data? I hypothesized that features that are inherent in projective techniques, such as the ambiguous nature of stimuli and subjective scoring and interpretation, would combine with the process of creeping determinism to make the outcomes with the drawing seem inevitable to participants, resulting in higher likelihood estimates given by participants who received hindsight information and a drawing.

Method

Participants

A total of 180 undergraduate students at the University of Rhode Island volunteered to participate in this study from a number of different choices designed to provide them with experience in psychological research. They received psychology course credit in exchange for their participation. Sign-up

sheets were designed so students who had participated in another study on the hindsight bias being conducted during the same semester were excluded. Participants were matched for gender, and randomly assigned to one of six groups, for a total of 30 per group. Appendix A shows the consent form for research the participants were required to sign.

The mean age of the sample was 19.5 years, with a standard deviation of 2.5 years. Within each condition, the age distribution was similar. The distribution of gender was 28.9% male, and 71.1% female. Participants' race was distributed as follows: 84.1% Caucasian, 5.7% African American, 4.0% Hispanic, and 6.2% Other; of the 180 participants, four did not provide information regarding race. Most participants (80.6%) received credit for an introductory psychology class, whereas 2.8% were from a sophomore-level course, and 16.6% were from a junior or senior-level psychology course. These variables were distributed similarly across the six different conditions.

Case Materials

The following was a case description constructed to portray a 15 year-old boy who was experiencing school related difficulties. It provides information that often is included in psychoeducational reports, but in a style appropriate for understanding by non-professionals:

We would greatly appreciate your taking a few minutes to read the following case description, which is part of a study on identifying reasons for school-related problems. It is the kind of

case that often is referred for evaluation to psychologists who work in school settings. We would like you to decide what probability you would have assigned to each of two possible reasons, had you been the school psychologist evaluating the case.

A teacher expressed concern about a fifteen-year old boy who has been earning failing grades in school and who appears isolated and withdrawn. One exception to this observation, however, is that he recently had two uncharacteristic arguments with peers that resulted in teacher intervention.

A review of his background information indicates that he has had trouble both in the past and more recently. Learning to read was difficult for him, and he required extra help in both the first and second grades. He has never been a good reader, and rarely reads for pleasure. His grades reflect these difficulties. He is generally a C student, although his reading and language arts grades have occasionally been Ds, with one F. In recent years, he has struggled with both his English and Spanish classes. In contrast, he consistently has earned Bs in Math, and he describes this subject as interesting and relatively easy for him. In addition, he always has been described as a nervous boy, getting uptight before exams and oral presentations.

He is reported to be an avid basketball fan, both as a player and observer. Over the past several months, however, his interest in basketball has diminished. His teachers and friends describe him as gloomy lately, and state that he does not seem to be interested in things he normally enjoys.

He recently was referred to the school psychologist who conducted an evaluation of these school-related difficulties. If you had been the school psychologist evaluating this case, what is the probability you would assign to each of the following two possible reasons for the school-related difficulties as the primary one? Assign an estimate to each, making sure they add to 100%.

Social-emotional problem _____%

(e.g., moodiness, difficulty relating to others,

problems with expressing feelings)

Learning problem _____%

(e.g., slow learner, problems with attention or

memory, difficulty with thinking or reasoning)

In order to determine if the reasons for difficulty identified subsequent to the case description were approximately equally likely, the case was piloted with 15 graduate students in psychology (8 school-psychology and 7 clinical-psychology students). The mean probability assigned to social-emotional problems as the primary difficulty was 53.9% (SD = 16.7%, range = 25% to 75%),

and learning problems was 46.1% ($SD = 16.7\%$, range = 25% to 75%). A t -test for dependent samples showed that the reasons generated were generally perceived as equally likely ($t < 1$).

A drawing of a person also was developed (see Appendix B). Included in the drawing were signs often interpreted as suggesting either social-emotional problems or learning problems (Machover, 1949; Ogdon, 1990). For example, pointy fingers frequently are interpreted as signs of aggression, and have been suggested to indicate severe mental disturbance (Buck, 1948; Machover, 1949; Urban, 1963). Shoulders omitted often are seen as indicating feelings of inferiority or depression, and crossed eyes have been interpreted as indicative of learning or cognitive difficulties. When evaluated using the DAP: Quantitative Scoring System (Naglieri, 1988), the drawing represents a measure of nonverbal cognitive development that is within normal limits (Standard Score = 107; $M = 100$, $SD = 15$). In addition, when scored using the DAP: SPED (Naglieri, McNeish, & Bardos, 1991) the drawing is not considered to be clinically significant (T-score = 40; $M = 50$, $SD = 10$). Thus, it is reasonable to infer that the drawing does not clearly reflect empirically supported interpretation as either developmentally or psychosocially pathological.

Procedure

The design for this study is illustrated in Table 1. Participants were matched for gender, then 30 were assigned to each of six conditions: (F1) a foresight condition with no drawing; (F2) a foresight condition with a drawing; (H1)

a hindsight condition for learning problems with no drawing; (H2) a hindsight condition for social-emotional problems with no drawing; (H3) a hindsight condition for learning problems with a drawing; and (H4) a hindsight drawing for social-emotional problems with a drawing. The 30 participants in the foresight condition were given the case description previously described.

Each of the two hindsight groups had an extra sentence inserted between the two sentences in the first paragraph of instructions to the foresight group. The two sentences (one for each of the hindsight groups) were as follows: *"This is a description of a boy who is experiencing difficulties primarily due to emotional problems,"* or *"This is a description of a boy who is experiencing difficulties primarily due to learning problems."*

Foresight-drawing group participants received the same descriptive information as the foresight participants, except that a drawing of a person was inserted after the final paragraph, preceded by the statement: *"During the evaluation, he made the following drawing."* The two hindsight-drawing groups differed from the hindsight ones in a similar way.

This study included two independent variables and one dependent variable. The independent variables were: (a) drawing, and (b) outcome. The dependent variable was the probability assigned for each of the possible reasons for difficulty, ranging from 0 to 100%.

Results

Figures 1 through 6 show the frequencies of probabilities for social emotional problems and learning problems by each group. Table 2 shows the mean probability assigned to each possible reason by each group. Also included in the Table (in parenthesis) is the number of participants whose probability for each reason exceeded the corresponding foresight estimate. I used two different statistical methods to analyze these data: one non-parametric and the other parametric. Because the probabilities for each outcome in the present study were ipsitive for each participant (i.e., $\Sigma p = 100\%$), these dependent measures clearly were related. Thus, I used the non-parametric (distribution-free) procedures employed in previous hindsight bias studies (e.g., Arkes et al., 1981; Fischhoff, 1975). As an alternative analysis, I isolated one level of the dependent variable (i.e., p values assigned to social-emotional problems) and analyzed these measures using parametric statistics (analysis of variance).

Non-parametric Analysis

I used the sign test to compare the number of participants who assigned higher probabilities in hindsight than the probabilities assigned in foresight. The sign test is a procedure that computes the differences between two variables for all cases and classifies the difference as either positive, negative, or tied. If the two variables are distributed similarly, then the number of positive and negative differences will not differ significantly (Glass & Hopkins, 1996). My results demonstrated that, overall, 73 out of 120 (i.e., 61%) hindsight participants

assigned higher probabilities to the known-to-have- occurred outcome than the corresponding foresight estimate ($z = 2.28$; $p = .011$).

I also conducted separate sign tests for the following hindsight vs. foresight comparisons: (a) no drawing, and (b) drawing. In the no drawing condition, 37 out of 60 (i.e., 62%) of the hindsight participants assigned higher probabilities to the known-to- have-occurred outcome than in the corresponding foresight group ($z = 1.68$; $p = .047$); whereas, in the drawing condition 36 out of 60 (i.e., 60%) participants assigned higher probabilities ($z = 1.42$; $p = .078$).

Finally, I also compared hindsight versus foresight probabilities within each drawing by outcome condition, that is: (a) drawing, social-emotional; (b) drawing, learning; (c) no drawing, social-emotional; and (d) no drawing, learning. In the drawing, social-emotional group, 22 of 30 (i.e., 73%) of the hindsight participants assigned higher probability estimates to social-emotional problems than the corresponding foresight group ($z = 2.41$; $p = .008$). In the drawing, learning group, 14 of 30 (i.e., 47%) of the hindsight participants assigned higher probabilities than the foresight group ($z = .44$; $p = .330$). In the no drawing, social-emotional group, 18 of 30 (i.e., 60%) of the hindsight participants gave higher estimates than the corresponding foresight group ($z = .91$; $p = .181$). In the no drawing, learning group, 19 of 30 (i.e., 63%) of the hindsight participants assigned higher probabilities than the corresponding foresight estimate ($z = 1.28$; $p = .100$). My results demonstrate that the only individual group that demonstrated the bias (at $p < .05$) was the hindsight for social-emotional problems with the drawing group.

Parametric Analysis

I also analyzed the data with a two by three between groups analysis of variance test (ANOVA). The design was $A \times B$, where A = drawing (two levels: either drawing or no drawing), and B = outcome (three levels: no outcome given, social-emotional problem, or learning problem). To address the previously mentioned concern regarding the dependence of the two probability estimates provided by each participant, the probability estimate assigned to social-emotional problems (versus learning) arbitrarily was chosen as the single dependent variable. The ANOVA showed a main effect for the outcome variable, $F(2, 174) = 5.273$, $p = .006$, but not for the drawing variable, $F(1, 174) = 3.364$, $p = .068$, or for the interaction between these two variables, $F < 1$. The ANOVA summary table appears in Table 2. As a follow-up analysis, I used the Tukey Honestly Significant Difference Test (Tukey HSD). This analysis compared each of the three outcome groups. I found that the only comparison that yielded a significant result was the hindsight condition for learning problems compared with the hindsight condition for social-emotional problems ($p = .003$).

Finally, I examined the results of the ANOVA to determine effect size of the significant result as well as the power of the non-significant results. The method for calculating effect size was omega-squared, measuring the proportion of variance accounted for (Keppel, 1982). For the outcome variable ω^2 was .057, a small effect. I examined the power of the non-significant result to examine the likelihood of Type II errors, that is how likely it was that this design did not detect an effect of the drawing or interaction, when one in fact was present. The power

of the drawing condition was moderate ($\beta = .554$). This implies that if an effect of the drawing was present, it was 55% likely that my study was not powerful enough to detect it. The power of the interaction was quite low ($\beta = .923$). In terms of the likelihood that I would have found a significant interaction if it were present, my chances were minimal, or it was 92% likely that I would not detect an effect if it were present. Therefore, the interpretation of the non-significant result is potentially misleading. Further study will need to consider this issue in order to decrease the likelihood of a Type II error.

Discussion

These results provide support for the first research question, regarding the replication of previous research on the hindsight bias. As noted, many different methods have been used to measure the hindsight bias. Both parametric and non-parametric statistics have been used. Consistent with Arkes et al. (1981) and Fischhoff (1975), my study demonstrated the hindsight bias using the non-parametric sign test. The bias was limited to one specific hindsight condition, the hindsight social-emotional with drawing group.

Arkes et al. (1981) and Fischhoff and Beyth (1975) also found the bias to be present in only some of their groups. They concluded that the hindsight bias tends to be stronger for those events initially judged to be least plausible (low base-rate events). In the present study, each outcome was deemed equally likely given the data provided (approximately 50/50 as indicated by the pilot study and the foresight, no drawing estimates). Therefore, other explanations should be

considered. Pennington (1981) suggested that more data tend to increase the hindsight bias. This explanation is not supported by the present study. In fact, the opposite was observed in terms of the drawing vs. no drawing groups. The no-drawing group demonstrated the bias, whereas the drawing condition overall did not.

If the observations of Arkes et al. (1981) and Fischhoff (1975) are generalizable, that least plausible outcomes tend to demonstrate the bias, whereas higher base-rate outcomes are less likely, then the logical question may be: Why did the drawing social-emotional group demonstrate the bias at all? Perhaps there was something inherent in the kind of information given. If the DAP tends to be interpreted in terms of emotional functioning more frequently, then the hindsight information may have become more salient to participants than it was in the other conditions. The projective hypothesis, advocating subjective interpretation of ambiguous stimuli, combined with the concept of creeping determinism, may offer a plausible explanation. To reiterate, creeping determinism refers to the observation that reporting an event's occurrence increases its perceived inevitability. If participants in my study, in fact, were susceptible to the effects of this cognitive process, and the ambiguity of the drawing was persuasive enough to encourage bias toward social-emotional problems, then the bias would be expected as observed. This question could be evaluated further in subsequent research.

Characteristics of the sample population also may provide insight into the observed results. Specifically, there may be differences in the way professionals

and nonprofessionals analyze case histories. Although undergraduate students have been used in previous research of the hindsight bias, it has been shown that they are more likely to demonstrate the bias when considering historical or almanac data than case-history data (Christensen-Szalanski & Willham, 1991; Fischhoff, 1977).

The second research question I was interested in asked to what extent the hindsight bias was affected by inclusion of the DAP. The results of the ANOVA suggest that the outcome information had a significant overall effect on assigned probabilities. Specifically, the combined hindsight learning groups (with and without drawing) differed reliably from the combined hindsight social-emotional groups (with and without drawing). Each of these means were in the expected direction, with the hindsight social-emotional group assigning probabilities to social-emotional problems higher than 50 and the hindsight learning group assigning probabilities lower than 50.

The drawing condition overall did not demonstrate a reliable effect on likelihood estimates. Yet, when the drawing was presented with social-emotional outcome information, estimates were significantly higher than when the drawing was presented with learning outcome information. This suggests that the combined effect of the type of outcome information (social-emotional problems) and the drawing exerted an effect on the way the case information was interpreted. In other words, simply including the drawing with the outcome information for social-emotional problems persuaded participants to interpret the

case information differently. An interesting follow-up to these observations would be to examine the incremental validity effect of the drawing.

In addition, it may also be useful to examine more closely the differential effect of the DAP on cognitive versus social-emotional information. For example, a majority of the literature on the DAP refers to its use as a measure of social-emotional functioning. The Chapmans' (1967) study demonstrated that the presence of drawings elicits stereotypes of social-emotional functioning by both trained clinicians and untrained college students. For the purpose of this study, the drawing that was used was given to a group of students in an undergraduate psychology class. They were asked to provide qualitative descriptors of the drawing. All of the students made statements about the person who drew the picture, even though that was not the instruction given. In addition, 98% of the statements referred to the social-emotional functioning of the person who drew the picture, whereas only a few referred to cognitive functioning. A sample of these statements appears in Appendix C. This is an interesting observation, because the DAP is a better predictor of non-verbal cognitive functioning than of social-emotional functioning. The influence of the picture on the way other information is interpreted should be considered by practicing clinicians who claim that the worst effect will be no effect at all. This question also should be considered further in subsequent research.

The observed effect size of the outcome variable was consistent with results reported by Christensen-Szalanski and Willham's (1991) meta-analysis of 122 studies of the hindsight bias. They found an overall small effect size for the

hindsight bias. Within that analysis, Christensen-Szalanski and Willham (1991) examined several moderator variables and found that the hindsight effect was correlated with both the participants' familiarity with the task and whether the outcome information stated that the event did or did not occur. They concluded that the more familiar the participants were with the task, the smaller the effect of the hindsight bias. In addition, they observed that people were less likely to reduce their probability estimates retrospectively when told that an event did not occur, than they were to increase their likelihood estimates retrospectively when told that an event did occur.

Christensen-Szalanski and Willham (1991) also examined subgroups of their sample. A large number of studies were conducted using undergraduates. These studies tended to demonstrate the bias more than any other group (i.e., the average effect size demonstrated a medium effect). Within this group, they found that students who were given case-history problems (such as in the present study) generated a small effect size, whereas almanac type questions generated an average effect size nearly three times larger than any other that was observed. Therefore, my results are consistent with those found in the meta-analysis.

A potentially confounding influence in this design is that the amount of information in the case-description-plus-drawing conditions is greater than the case-description-only conditions. As noted, previous research has suggested that as amount of information increases, the magnitude of the hindsight bias is likely to increase (Pennington, 1981). Pennington's study compared longer

versus shorter verbal descriptions, whereas this study examined the inclusion of additional information in the form of a drawing. Within the drawing condition, the bias was observed only in the group that was given social-emotional outcome information, suggesting that another explanation should be considered. Due to the differential effects that were observed, isolating potential reasons for such differences (e.g., illusory beliefs based on projective information versus differential amounts of information) should be addressed in subsequent research.

Given the fact that the observed hindsight bias was in the minimal range, one must consider the cost of this cognitive error in daily functioning when deciding whether or not to attempt to decrease its effect. Although the results of this study are consistent with previous research, it would be of both theoretical and practical interest to see how practicing psychologists would respond in a similar situation. In addition, the issue of the potential differential effects of the DAP depending on the type of information (case versus almanac) included with it should be evaluated further.

Implications

My study provided support for the hindsight bias consistent with previous research. When non-parametric statistics and effect size analyses were used, the bias was observed. The combined strength of these methods lead to a fair degree of confidence in asserting that the bias is, in fact, present. In contrast, parametric tests, that compared the actual observed means of the foresight and hindsight groups, did not demonstrate the bias. The concern regarding the relatedness of the probability estimates is a valid one. With my use of parametric

statistics I attempted to diminish the effect (by using only the probabilities assigned to social-emotional functioning).

I have identified an issue here that has not been addressed in previous study of the hindsight bias. Specifically, given the differential results I observed when using parametric versus non-parametric statistics, study of the statistics of other hindsight bias research is warranted. Perhaps the bias is not as robust as previously believed. Alternatively, the construct of hindsight bias may be understood differently based on varying methodology. When non-parametric methods are used, one compares the number of participants who assigned higher probabilities than the mean foresight estimate, whereas the parametric methods compare actual means.

Another important result of my study was the observation that the DAP demonstrates a differential effect based on the kind of information with which it is presented. Given the vast amount of information provided during a psychoeducational evaluation, there remains a substantial risk that clinicians will be susceptible to the same cognitive errors demonstrated by this college population. As noted previously, clearly the worst effect is more significant than no effect. In fact, these kinds of beliefs actually can increase the likelihood that errors in diagnosis will occur.

Finally, it is important to consider the applied usefulness of the results I have presented. Although some equivocal results were observed, it is clear that individuals are susceptible to cognitive biases like the one I have reviewed. In addition, depending on analytic technique, projective methods (specifically the

DAP), exert an observable influence on how individuals interpret other information. Given these observations, and the implications for faulty decision making and diagnosis, these issues should be examined using practicing clinicians.

Limitations

There are limitations to this study that should be considered when making implications based on the results. In this study I considered one dependent variable, thereby potentially restricting the observation of the hindsight bias. As noted, Synodinos (1986) used both probabilities and degree of confidence as dependent variables. Synodinos reported that confidence might be a more sensitive measure of the hindsight effect, because it directly addresses participants' perceived accuracy. Although Fischhoff (1975) has clearly defined the hindsight phenomenon, perhaps further consideration of the construct itself should be considered.

It is important to use caution when making inferences about a particular population. Specifically, although my study demonstrates susceptibility to the hindsight bias and the negative impact of the DAP, the generalizability of these observations is questionable. It remains unclear how practicing clinicians would consider the same information used here. Naturally, in a real evaluation situation, the clinician would have access to more information than is included here. Consequently, although these results are interesting, their relevance to actual diagnostic situations remains questionable.

Conclusion

This study considered the effect of the inclusion of a human figure drawing on the hindsight bias. A review of the development of the Draw-a-Person test suggests that its clinical usefulness and psychometric properties are limited. Contrary to previously expressed belief about the benign effect of inclusion of the DAP, it was demonstrated that the DAP can exert an observable influence on the way in which other information is interpreted. In addition, this study examined the construct of the hindsight bias and proposed that the different methods of analyses that have been used may account for the variable observation of the bias.

Table 1

Design of the Study

	No Outcome Given	Social-Emotional-Problems Outcome Given	Learning-Problems Outcome Given
No Drawing Given	% assigned to S-E problems % assigned to learning problems	% assigned to S-E problems % assigned to learning problems	% assigned to S-E problems % assigned to learning problems
Drawing Given	% assigned to S-E problems % assigned to learning problems	% assigned to S-E problems % assigned to learning problems	% assigned to S-E problems % assigned to learning problems

Table 2

Mean Probability Assigned to Each Outcome

		Outcome Evaluated		
		Outcome	Social-Emotional	Learning
Group	N	Provided	Problem (S-E)	Problem (L)
No Drawing				
Foresight	30	None	51.83	48.17
Hindsight	30	S-E	55.60 (18)	44.40
Hindsight	30	L	44.50	55.50 (19)
Drawing				
Foresight	30	None	55.40	44.60
Hindsight	30	S-E	63.67 (22)	36.33
Hindsight	30	L	50.00	50.00 (14)

Note. The numbers in parentheses indicate the number of participants whose probability estimate for that particular outcome exceeds the corresponding foresight estimate.

Table 3

Analysis of Variance Summary Table for Comparisons of Hindsight Bias and
Projective Drawing Conditions

Sum of	Mean				
	Square	df	Square	F	Sig.
Drawing	1467.756	1	1467.756	3.364	.068
Outcome	4601.633	2	2300.817	5.273	.006
Drawing X Outcome	152.878	2	76.439	.175	.839
Error	75920.73	174	436.326		
Total	97348.00	180			

Figure 1

Foresight; Without drawing

Figure 2

Foresight; With drawing

Figure 3

Hindsight (learning); Without drawing

Percentage assigned to social-emotional problems.

 $\bar{M} = 44.5, SD = 17.68, N = 30$

Figure 4

Hindsight (social-emotional); Without drawing

Percentage assigned to social-emotional problems.

 $\underline{M} = 55.6, \underline{SD} = 21.88, \underline{N} = 30$

Figure 5

Hindsight (learning); With drawing

Figure 6

Hindsight (social-emotional); With drawing

Appendix A

The University of Rhode Island

Department of Psychology

A Study of Decision Making Strategies Among College Students

CONSENT FORM FOR RESEARCH

You have been asked to take part in a research project described below. The researcher will explain the project to you in detail. You should feel free to ask questions. If you have more questions later, Elizabeth Dufresne (783-3909), the person mainly responsible for the study, will discuss them with you. You must be at least 18 years old to be in this research project.

You have been asked to take part in the study that evaluates how people identify reasons for problems. If you decide to take part in this study here is what will happen: You will be given a brief description about an individual. You will be asked to answer a few questions about this description. In addition, you will be asked to fill out a questionnaire with various questions about you (age, year in college).

We do not expect that there are any risks or discomforts associated with this study. If, upon completion of the study, you feel any discomfort, please contact the person mainly responsible for the study.

Participation in this study will fulfill partial requirements for your Introductory Psychology course. Appropriate documentation of your participation will be made upon completion.

Your part in this study is confidential. None of the information will identify you by name. All records will be stored in a separate place from the consent forms so that confidentiality is maintained.

The decision to participate in this research is up to you. You do not have to participate and you can refuse to answer any question.

Participation in this study is not expected to be harmful or injurious to you. However, if this study causes you any injury, you should call Elizabeth Dufresne (783-3909) or W. Grant Willis, Ph.D., Faculty Advisor (874-4245).

If you are not satisfied with the way this study is performed, you may discuss your complaints with Elizabeth Dufresne. In addition, you may contact the office of the Vice Provost for Graduate Studies, Research and Outreach, 70 Lower College Road, Suite 2, University of Rhode Island, Kingston, Rhode Island, telephone: (410) 874-2635.

You have read the Consent Form. Your questions have been answered. Your signature on this form means that you understand the information and you agree to participate in this study.

Signature of Participant

Signature of Researcher

Appendix B

Draw-a-Person

Appendix C

Sample Statements From Pilot of Drawing

"The character looks stressed, which probably represents how he (the artist) feels."

"He feels inferior (on big piece of paper, drew small person)."

"He is unhappy and confused." "Very attentive. Looking forward, big ears."

"The eyes made me think that he is confused and overwhelmed, stressed."

"He does not conform to standards and ideals-the number of fingers on the boy is not the standard 5."

"The picture seemed sad so the boy might also be upset about something."

"The hands look monstrous, perhaps this boy has some anger or violence."

"Everything on the left side of the body is drawn bigger, perhaps this boy has some kind of body dysmorphic disorder."

"This boy has a low self-esteem - the hands are unique and the boy in the picture is not smiling."

"Mood of boy gloomy when he drew the picture."

"The boy is mentally disabled." "He has delusional problems."

"He is unhappy when he drew this picture."

"The boy probably feels everyone has a monster inside of them."

"He is underdeveloped." "Often when things are left out of a picture such as the neck and wrists that can be signs of a lower developmental level."

"It is by a boy who is mentally retarded or autistic (some lack of intelligence)."

"He is not that intelligent because body parts are not accurate."

Bibliography

Abell, S. C., Heiberger, A. M., & Johnson, J. E. (1994). Cognitive evaluations of young adults by means of human figure drawings: An empirical investigation of two methods. *Journal of Clinical Psychology, 50* (6), 900-905.

American Psychological Association (1992). Ethical principals of psychologists and code of conduct. *American Psychologist, 47*, 1597-1611.

Anastasi, A. (1988). *Psychological testing*. New York: Macmillan Publishing Company.

Arkes, H. R., Faust, D., Guilmette, T. J., & Hart, K. (1988). Eliminating the hindsight bias. *Journal of Applied Psychology, 73*, 305-307.

Arkes, H. R., Wortmann, R. L., Saville, P. D., & Harkness, A. R. (1981). Hindsight bias among physicians weighing the likelihood of diagnoses. *Journal of Applied Psychology, 66* (2), 252-254.

Aspel, A. D., Willis, W. G., & Faust D. (1998). School psychologists' diagnostic decision-making processes: Objective-subjective discrepancies. *Journal of School Psychology, 36*, 137-149.

Buck, J. N. (1948). The H-T-P technique, a qualitative and quantitative scoring manual. *Journal of Clinical Psychology, 4*, 317-396.

Buck, J. N. (1966). *The House-Tree-Person technique: Revised Manual*. Beverly Hills: Western Psychological Services.

Chapman L.J., & Chapman, J. P. (1967). Genesis of popular but erroneous psychodiagnostic observations. *Journal of Abnormal Psychology, 72* (3), 193-204.

Christensen-Szalanski, J. J. & Willham, C. F. (1991). The hindsight bias: A meta-analysis. *Organizational, Behavioral and Human Decision Processes*, 48, 147-168.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (rev. ed.). New York: Academic Press.

Connolly, T., & Bukszar, E. W. (1990). Hindsight bias: Self-flattery or cognitive error? *Journal of Behavioral Decision Making*, 3, 205-211.

Creyer, E., & Ross, W. T. (1993). Hindsight bias and interferences in choice: The mediating effect of cognitive effort. *Organizational Behavior and Human Decision Processes*, 55, 61-77.

Feyth, J. M., & Holmes, C. B. (1994). Use of the draw-a-person with conduct disordered children. *Perceptual and Motor Skills*, 78, 1353-1354.

Fischhoff, B. (1975). Hindsight $\neq$ foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, 1, 288-299.

Fischhoff, B. (1976). The effect of temporal setting on likelihood estimates. *Organizational Behavior and Human Performance*, 15, 180-194.

Fischhoff, B. (1977). Perceived informativeness of facts. *Journal of Experimental Psychology: Human Perception and Performance*, 3, 349-358.

Fischhoff, B., & Beyth, R. (1975). "I knew it would happen": Remembered probabilities of once-future things. *Organizational Behavior and Human Performance*, 13, 1-16.

Frank, L. K. (1939). Projective methods for the study of personality. *The Journal of Psychology*, 8, 389-413.

Fuller, G. B., & Goh, D. S. (1983). Current practices in the assessment of personality and behavior by school psychologists. *School Psychology Review*, 12, 244-249.

Glass, G. V. & Hopkins, K. D. (1996). *Statistical methods in education and psychology* (3rd ed.). Allyn and Bacon: Boston.

Goldstein, A. P. & Rawn, M. L. (1957). The validity of interpretive signs of aggression in the drawing of the human figure. *Journal of Clinical Psychology*, 13, 169-171.

Goodenough, F. L. (1926). *Measurement of intelligence by drawings*. New York: Harcourt, Brace & World.

Hammer, E. (Ed.). (1958). *The clinical application of projective drawings*. Springfield, IL: Charles C. Thomas.

Handler, L. (1996). The clinical use of drawings: Draw-a-Person, House-Tree-Person, and Kinetic Family Drawings. In C. S. Newmark (Ed.), *Major Psychological Assessment Instruments* (2nd ed.) (pp206-293). Needham Heights, MA: Allyn & Bacon

Harris D. B. (1963). *Children's drawings as measures of intellectual maturity: A revision and extension of the Goodenough Draw-a-Man test*. Harcourt, Brace & World, Inc.: New York.

Hawkins, S. A. & Hastie, R., (1990). Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin*, 107, 311-327.

Jolles, I. (1964). *A catalogue for the qualitative interpretation of the H-T-P* (Rev. ed.). Beverly Hills: Western Psychological Services.

Kamphaus, R. W., & Pleiss, K. L. (1991). Draw-a-Person techniques: Tests in search of a construct. *Journal of school Psychology, 29* (4), 395-401.

Kennedy, M. L., Faust, D., Willis, W. G., & Piotrowski, C. (1994). Social-emotional assessment practices in school psychology. *Journal of Psychoeducational Assessment, 12*, 228-240.

Keppel, G. (1982). *Design and analysis: A researcher's handbook*. (2nd ed.). Prentice-Hall, Inc.: New Jersey.

Koppitz, E. M. (1966). Emotional indicators on human figure drawings of shy and aggressive behavior. *Journal of Clinical Psychology, 22*, 466-469.

Lubin, B., Larsen, R. M., & Matarazzo J. D. (1984). Psychological test usage in the United States: 1935-1969. *American Psychologist, 39*, 451-454.

Machover, K. (1949). *Personality projection in the drawing of human figure*. Springfield, IL: Thomas.

Naglieri, J. A. (1988). *Draw a person: A quantitative scoring system*. New York: Psychological Corporation.

Naglieri, J. A., McNeish, T. J., & Bardos, A. N. (1991). *Draw a Person: Screening Procedure for Emotional Disturbance*. Austin, TX: Pro-Ed.

Naglieri, J. A., & Pfeiffer, S. I. (1992). Performance of disruptive behavior disordered and normal samples on the Draw a Person: Screening Procedure for Emotional Disturbance. *Psychological Assessment, 4* (2), 156-159.

Ogdon, D. P. (1990). *Psychodiagnostics and Personality Assessment: A Handbook* (2nd ed.). Los Angeles, CA: Western Psychological Services.

Pennington, D. C. (1981). The British firemen's strike of 1977/78: An investigation of judgments in foresight and hindsight. *British Journal of Social Psychology*, 20, 89-96.

Piotrowski, C., Sherry, D., & Keller, J. W. (1985). Psychodiagnostic test usage: A survey of the Society for Personality Assessment. *Journal of Personality Assessment*, 49, 115-119.

Piotrowski, C., & Zalewski, C. (1993). Training in psychodiagnostic testing in APA-approved Psy.D. and Ph.D. clinical psychology programs. *Journal of Personality Assessment*, 61 (2), 394-405.

Pohl, R. F. & Hell, W. (1996). No reduction of hindsight bias after complete information and repeated testing. *Organizational Behavior and Human Decision Processes*, 67, 49-58.

Riordan, R. J., & Verdel, A. C. (1991). Evidence of sexual abuse in children's art products. *The School Counselor*, 39, 116-121.

Schkade, D. A. & Kilbourne, L. M. (1991). Expectation-outcome consistency and hindsight bias. *Organizational Behavior and Human Decision Processes*, 49, 105-123.

Schweighofer, A., & Coles, E. M. (1994). Note on the definition and ethics of projective tests. *Perceptual and Motor Skills*, 79, 51-54.

Sharpe, D. & Adair, J. G. (1993). Reversibility of the hindsight bias: Manipulation of experimental demands. *Organizational Behavior and Human Decision Processes*, 56, 233-245.

Urban, W. H. (1963). *The Draw-a-Person catalogue for interpretive analysis*. Los Angeles: Western Psychological Services.

Willis, W. G. (1992). Review of the Draw-a-Person: A Quantitative Scoring System. In J. J. Kramer, & J. C. Conoley (Eds.), *The Eleventh Mental Measurements Yearbook* (pp. 289-290). Lincoln, Nebraska: University of Nebraska Press.

Wood, G. (1978). The knew-it-all-along effect. *Journal of Experimental Psychology: Human Perception and Performance*, 4, 345-353.