From the Editor...

Welcome to the wonderful world of undergraduate psychology at the University of Pennsylvania. This is the fifth issue of Perspectives, a journal created by students as a way of sharing viewpoints, methods of inquiry, findings, and topics of interest. Perspectives enables undergraduate psychology students at the University of Pennsylvania an opportunity to publish their research. Hopefully, this journal will continue to provide a medium that students will be able to utilize as they probe advances in the various fields of psychology.

Representative of the great diversity in psychology, this year’s Perspectives presents research from a wide range of subjects. This year offers us an especially exciting assortment of papers, for many of the articles cover topics relevant to Penn undergraduate students. Samantha Prestia, Jason Silverston, Katie Woods, and Lisa Zigarmi examine the relationship between the attractiveness and popularity of Penn sorority sisters. Encompassing a more general female population, Danesh Modi investigates the dispositional factors that influence Penn women’s attitudes towards abortion. Also relevant to female college students, Theresia Laksmana discusses eating disorders and debates the pros and cons of using the Internet as a therapeutic tool. One disease that affects a population beyond women and college students, is depression. Brian Scott Ehrlich and Derek M. Isaacowitz present a study on the levels of subjective well-being in various age groups, while Jessica Murakumi takes a look at depression in terms of gender differences. Complementing these behavioral studies, Nicole Altman explores the brain and its organization of languages. Lastly, taking a look into our future, Joyce Tang conducts a study on the agreement and disagreement between the viewpoints of parents and children on the emotional and behavioral problems of the child in question.

The editorial board would like to thank the board of the Undergraduate Psychology Society, the Department of Psychology, and the undergraduate chair and journal advisor, David Williams, for all their help and support this year.

This was a great year working on the journal with such a dynamic group of editors. On behalf of the editorial board, we sincerely hope that you enjoy this journal and that it will make you laugh, pique your curiosity, and be an incentive for you to investigate further.

- Flora Ahn, Editor-in-Chief
Contents

Editor’s Note

The Effects of Attractiveness on Popularity; an Observational Study of Social Interaction Among College Students

*Samantha Prestia, Jason Silverston, Katie Wood, Lisa Zigarmi*

Predicting Factors of Parent and Child Agreement on Ratings of Child Emotional and Behavioral Problems

*Joyce Tang*

Does Subjective Well-Being Increase with Age?

*Brian Scott Ehrlich and Derek M. Isaacowitz*

Gender and Depression: Explaining the Different Rates of Depression Between Men and Women

*Jessica Murakumi*

Eating Disorders and the Internet: The Therapeutic Possibilities

*Theresia Laksmana*

Factors Influencing Women’s Attitudes Towards Abortion at the University of Pennsylvania

*Danesh Modi*

Cortical Organization of Bilingualism

*Nicole Altman*
The Effects of Attractiveness on Popularity; an Observational Study of Social Interaction Among College Students

Samantha Prestia, Jason Silverston, Katie Wood, Lisa Zigarmi

Fifty college women were rated on attractiveness with the top ten and bottom ten being put into groups labeled subjects of high attractiveness and subjects of lower attractiveness. Both groups of subjects were observed by researchers at college bars and parties. The number and type of approaches each subject had with other males and females were recorded with the number of approaches determining a subject’s popularity. Women of high attractiveness were found to be more popular because they had a higher total number of approaches than subjects of lower attractiveness. The total number of approaches with just females was found to be significant, suggesting that males have a strong influence on who is popular.

INTRODUCTION

Appearance is not supposed to matter. Society would have us believe that in an ideal world, an individual’s physical appearance is a relatively insignificant factor in others’ perceptions of him/her. In a meritocracy we are conditioned to believe that an individual’s worth is based on his/her skills, abilities, and personal conduct. Yet, according to The American Heritage Dictionary (1994: 54), to attract is “to cause, to draw near, to adhere,” or more relevantly, “to arouse the interest, admiration or attention of.” Therefore, by definition, attractiveness is a very powerful quality. Furthermore, years of research in the psychological and social sciences have shown that looks count in human affairs. Studies have shown that people who are considered attractive fare better with parents and teachers, make more friends and more money, and have better sex with more people (Cowley, 1996). The interpersonal consequences of physical attractiveness have led thousands of people to spend millions of dollars on beauty products and even cosmetic surgery to improve their looks (Cowley, 1996). The full impact of attractiveness is still unclear, but studies have established that a sense of what is attractive is innate and consistent across age, race, and culture. Its impact is subtle but powerful.

Attractiveness

In a groundbreaking study, psychologist Judith Langlois (1989) concluded that infants share with adults a sense of what is attractive. Three and six month old babies were shown pairs of facial photographs that were previously rated as attractive and unattractive. Langlois found that infants gazed significantly longer at “attractive” faces than at “unattractive” faces. From this study, we can see that while it may be impossible to create a clear definition of what makes a face attractive, we somehow share an innate sense of it even before we are socialized. In a lab study by Berry (2000), children showed more positive affect when interacting with an attractive adult. Therefore, Berry concludes that “attractiveness, at least facial attractiveness, is both discriminated and preferred at a very young age” (Berry, 2000: 278). These studies show that humans share an innate idea of what we like to see, and attractiveness draws us to those who posses it. The inevitable question is, why do we prefer this “attractiveness” in the appearance of others? Two basic theories address this question: the evolutionary perspective and a social conditioning perspective.

The Evolutionary Perspective on Attractiveness

Evolutionary theories of attractiveness propose that our ancestors evolved preferences for features con-
considered “attractive” because of the reproductive advantages those features ultimately yielded. In particular they hypothesize that attractiveness is a signal of potential “reproductive success.” Attractive features are posited to be “honest advertisements” of reproductive status and genetic fitness (Berry & Miller, 2001). Individuals who valued attractiveness or were attractive themselves are hypothesized to out-reproduce those who do or are not. Therefore, such heritable preferences evolved over time within the population (Berry & Miller, 2001). Although this study will not explore facial symmetry, symmetry is a consistently documented indicator of attractiveness within animals and humans. Thornhill and Gangestad (1994), researchers at the University of Michigan, found evidence that facial symmetry is associated with actual health. In their analysis of diaries kept by one hundred students over a two-month period, they found that subjects with the least symmetrical faces had the most physical complaints. These problems ranged from congestion to insomnia. Student subjects with less symmetrical features (i.e. less attractive) also reported more anger and jealousy. Direct knowledge of a woman’s age and health is not always available, but qualities such as symmetry, associated with age and health are readily accessible and tend to correlate with physical attractiveness. Hence, it makes sense that attractiveness is favored in a population, since it is a phenotypic heuristic indicating health and reproductive potential (Cowley, 1996). “Thus, men’s preferences for these attractive features presumably evolved because greater reproductive success accrued to men who preferred and mated with attractive women than to men who preferred and mated with less attractive women.” (Buss, 1989).

Social Conditioning Perspective

The second theory which attempts to explain the importance of attractiveness lies within the vast body of research exploring the socio-cultural effects of appearance on behavior and beliefs. A study by Goldman and Lewis (1977) indicates a relationship between physical attractiveness and social skill. This study asked subjects to talk on the phone with three unknown people of the opposite sex. The subjects were to indicate how much they liked each of the three people with whom they talked and how socially skilled they found each to be. Goldman and Lewis found a "significant tendency for the more attractive subjects to be rated as more skillful in their telephone conversation" (128). This finding illustrates the correlation between physical attractiveness and social skills. This correlation may begin in childhood. Goldman and Lewis provide an explanation for this by postulating that "it seems possible that attractive children, who receive more favorable reactions from others, will be more comfortable in social settings and, through the operation of the positive expectations and reactions of others, develop better social skills than less attractive children. This can carry into adulthood, thus making them more popular" (126).

Extending Goldman and Lewis’s study (1977), Berry and Miller (2001) found that a woman’s attractiveness influences the nature of initial opposite-sex interactions. In their study, 51 previously unacquainted opposite sex dyads were videotaped while participating in an initial six-minute interaction. Participants then individually described their feelings about the interaction and their interaction partner. Observers later viewed the interactions and evaluated their quality. Four male and four female judges rated the attractiveness of each of the participants. Attractive women were associated with higher quality interactions than were less attractive women, and attractive women enjoyed their interactions more than did unattractive women. Interpersonal consequences of attractiveness are greater for women than for men within this context of initial opposite sex interactions.

Popularity

With the great emphasis society places on attractiveness, there is no wonder that research psychologists have conducted numerous studies to try to understand its effects better. A common assumption surrounding attractiveness is that it is related to or in some way confounded with an increased social acceptance or popularity. Popularity, as defined by the American Heritage Dictionary, means to be "generally liked or admired." (1994: 492) In a study by Krantz (1987) the connection between attractiveness and popularity was explored using 24 female and 24 male kindergarten students. Subjects were asked to pick the facial pictures of two classmates of the same sex with whom "they would like to be friends" in the upcoming school year. It was shown that female kindergarten students indicated a desire for friendship with same sex classmates who were previously rated by adults as attractive. In a similar study, 59 preschool children were rated by adults on a physical attractiveness scale. The children were then presented with a board which was filled with pictures of their classmates. They were asked to pick two pictures of children they "especially
liked" (Vaughn & Langlois, 1983:562). It was concluded that "physical attractiveness and popularity are significantly related" (Vaughn & Langlois, 565). Clearly there is a relationship between facial attractiveness and popularity. In fact, one of the most thoroughly documented findings in social psychology is the "attractiveness-halo effect" (Berry, 2001) which predicts that attractive people receive more positive responses from others than do unattractive people. This positive response is complimented by the general perception that attractive people are in some way more successful in social and emotional areas (Goldman & Lewis, 1977). Research on impression formation has found that people attribute socially desirable characteristics to good-looking individuals (Feingold, 1990). According to the attribution theory, individuals attribute more socially desirable personality and social characteristics to attractive than to unattractive target persons (Lee, Adams, & Dobson, 1984). Attractive individuals are perceived to have a myriad of desirable personality (self-esteem, self-concept, emotional stability) and social (occupational success, social skills, higher education) characteristics. Vaughn and Langlois point out that "even preschool age children tend to rate attractive peers as friendlier, smarter, and less likely to start fights than unattractive peers (Dion, 1973)" (Vaughn & Langlois, 561).

Berscheid and Walster (1974), using a self-report popularity index, found physical attractiveness and popularity to be significantly correlated (.46 for women and .31 for men), indicating that physical attractiveness is more important for a woman's social experience than for a man's. In a related study by Berscheid and Walster (1974), a man with either an attractive or unattractive woman walked into a bar. The individuals at the bar were asked to state their "overall impression" of the man, to indicate how well they thought they would personally like him, and to rate him on a number of personality scales. The response to the man when he was accompanied by an attractive woman was compared with the impression he made when accompanied by an unattractive woman. The study found that when a man was seen with an attractive woman, he received the "most favorable overall" impression from others. When he was accompanied by unattractive women, he was viewed negatively (Berscheid & Walster, 1974). Males gain considerable prestige by associating with physically attractive females (Goldman & Lewis, 1977). It is not only better for a male to be associated with a beautiful woman than not, but also being associated with an unattractive girl tends to detract from the favorableness of the man's overall impression (Berscheid & Walster, 1974).

Since studies on this topic rarely use observational methods and since little exploration of older populations exists, the following study will employ observational methods on college-aged individuals in an attempt to contribute to the evolving body of research on the effects of attractiveness on popularity in naturally occurring social interactions.

We hypothesize that college-aged women of high attractiveness (H's) are more popular than college-aged women of lower attractiveness (L's). We specifically hypothesize that:

1. In social situations at bars or parties, women of high attractiveness will be approached by both male and female students more frequently than will women of lower attractiveness.
2. In social situations at bars or parties, women of high attractiveness will approach both male and female students more frequently than will women of lower attractiveness.
3. Males will approach as well as be approached by women of high attractiveness with more frequent physical behaviors than they will with women of lower attractiveness.
4. Male and female judges will not differ in their ratings of the facial attractiveness of female subjects.

METHOD

Participants

Twenty college-aged female sorority members were chosen as subjects. All were in their junior or senior year of study. The attractiveness of the subjects was determined by a survey of 20 men and 20 women from a neighboring university. Two-inch by two-inch color facial photographs of 50 members of the sorority were passed on blank, white index cards and handed to volunteer judges. The cards were in no particular order and were shuffled after each sorting. Each judge was given the following instructions: "Please place each individual picture in one box based on your opinion of their attractiveness. The three boxes have been labeled "low attractiveness," "average attractiveness," and "high attractiveness." Researchers assigned each attractiveness level a numerical value (High = 2, Average = 1, Lower = 0). The ten subjects in the High Attractiveness group were those with the highest combined score. The ten with the lowest combined score were labeled as having Lower Attractiveness from the 40 ratings.
Procedure

Each subject was observed for two ten-minute periods at different social gatherings. Observers were unaware of the subject’s attractiveness rating until observations were complete. All observations were made in the first hour of the event in an attempt to control for the effects of alcohol. College social gatherings are defined as local bars or parties frequented heavily by college students or Greek off-campus parties.

During the ten-minute observation intervals, the number of times a subject approached another individual was recorded as was the number of times the subject was approached by another. Observers recorded whether the subject approached a male or a female, and whether the subject was approached by a male or female. The types of approach behaviors recorded were: a) verbal communication, b) hugging, c) kissing, d) touching, e) nodding, and f) waving.

Each approach between the subject and another individual was recorded for content and direction using a focal observational style (Martin & Bateson, 1993: 84). A subject was chosen randomly from the pool of attractive and unattractive subjects by picking a name out of a hat. If the subject was not present, another subject was chosen in the same way. This procedure was followed until a subject was found at the event. The start of the ten-minute observation period was determined by picking a place in the room and waiting for the subject to walk past it. Two researchers observed a subject at the same time from different sides of the room in order to control for the subject’s movement throughout the bar or party and make sure that all approaches were recorded accurately.

Popularity was defined as the number of approaches a subject had with different individuals within the two ten-minute period. Subsequent approaches with one individual beyond the first encounter were not counted. Those subjects with totals above the median were defined as popular and subjects with totals below the median were defined as unpopular. Therefore, the more approaches a subject was involved in, the more popular she was considered to be.

The four research observers simultaneously observed one subject for two ten-minute periods and calculated inter-observer agreement by dividing the total number of observations minus the number of observations that were different, by the total number of observations and then multiplying this number by 100%. The inter-observer agreement was determined to be 92% before data collection began.

RESULTS

Since our data are nominal, cannot be assumed to be normally distributed, and contain a small number of subjects, non-parametric tests of significance have been used (Martin & Bateson, 1993). One such test is the Wilcoxon Rank Sums Test. This test is a more powerful form of the median test because it does not ignore the specific rank-order of subjects. The requirements for a standard median test are: 1) a comparison between two or more samples, 2) ordinal data, and 3) random sampling (Levin & Fox, 2000). Our data fully meet criteria one and two and adequately meet criteria three. In order to see any differences in the rank order of our subjects, a Wilcoxon Rank Sums Test has been used. It was used because it examines the rank ordering of all subjects to determine whether the ranked values for a variable are equally distributed throughout the two samples. However, for sub-hypothesis number four we performed a parametric logistic regression. For all tests a .05 p-value was used to reject the null hypotheses.

To test the general hypothesis that women of high attractiveness are more popular than women of lower attractiveness, we did a Wilcoxon Rank Sums Test to compare the number of subjects of high attractiveness who fall above the median number of total approaches for all subjects to the number of subjects of lower attractiveness who are above the median. We found that there is a chi square of 13.79 which was significant at a p-value of .0002. This shows that women of high attractiveness had a significantly higher frequency of total approaches than women of lower attractiveness [See Figure 1]. We also tested the same hypothesis for the popularity of the subjects by running the same test on total approaches involving only males and total approaches involving only females. We performed these last two tests to determine whether the sex of the individuals the subjects approached and were approached by had an effect. These two tests helped us determine whether males or females contributed equally to the popularity scores of the subjects. We found that for total approaches involving only males there was a chi square of 14.43, which was significant at a p-value of .0001. This showed that women of high attractiveness had a significantly higher frequency of total approaches involving only males than do women of lower attractiveness. We found that in total approaches involving only females there was not a significant chi square of 3.60 at a p-value of .0578. This shows that women of high attractiveness did not have a
significantly higher frequency of total approaches involving only females than women of lower attractiveness, meaning that women approach and are approached by each other with no significant attention to attractiveness.

To test the first sub-hypothesis, that women of high attractiveness will be approached by males and females with greater frequency than women of lower attractiveness, we used the Wilcoxon Rank Sums Test. There were three ways to test this, and the Wilcoxon Rank Sums Test allowed us to make all three comparisons. We tested to see where high and lower attractiveness subjects fell compared to the median for the total number of times other individuals approached them, where subjects fell compared to the median for the number of times just males approached them, and where subjects fell compared to the median for the number of times just females approached them. We found that for the total number of times subjects were approached by other individuals there was a chi square of 9.78, which is significant at a p-value of .0018. This shows that women of high attractiveness were approached at a significantly higher frequency than women of lower attractiveness [See Figure 2]. We found that for the total number of times subjects approached a male there was a chi square of 12.81, which was significant at a p-value of .0003. This showed that women of high attractiveness approached males at a significantly higher frequency than do women of lower attractiveness. We found that for the total number of times subjects approached a female there was a chi square of 3.97, which was significant at a p-value of .0462. This shows that women of high attractiveness were approached by females at a significantly higher frequency than women of lower attractiveness.

To test the second sub-hypothesis, that women of high attractiveness will approach both male and female students with greater frequency than will women of lower attractiveness, we used the Wilcoxon Rank Sums Test. There were three ways to test this, and the Wilcoxon Rank Sums Test was used for all three. We tested to see where subjects fell compared to the median for the total number of times the subject approached an individual, the median for the number of times the subject approached a male, and the median for the number of times the subject approached a female. The tests for each sex were done to determine if it had an effect on whom our subjects approached. We found that for the total number of times subjects approached an individual there was a chi square of 10.21, which was significant at a p-value of .0014. This showed that women of high attractiveness approached other individuals at a significantly higher frequency than women of lower attractiveness [See Figure 3]. We found that for the total number of times subjects approached a male there was a chi square of 7.84, which was significant at a p-value of .0051. This showed that women of high attractiveness approached males significantly more frequently than do women of lower attractiveness. We found that for the total number of times subjects approached a female there was a chi square of 5.80, which was significant at a p-value of .0160. This showed that women of high attractiveness approached females significantly more frequently than did women of lower attractiveness.
Figure 3: This graph shows that women of high attractiveness approach individuals significantly more than do women of low attractiveness. Note: each dot may represent more than one subject.

Figure 4: This graph shows that women of high attractiveness are involved in significantly more approaches with males that involve physical behavior than are women of low attractiveness. Note: each dot may represent more than one subject.

To test the third sub-hypothesis, that males will approach as well as be approached by women of high attractiveness with a higher frequency of physical behaviors than they will with women of lower attractiveness, a Wilcoxon Rank Sums Test was used to compare where both groups of women fell compared to the median for the number of physical behaviors (hugging, kissing and touching) with males. We found that the number of times subjects approached a male with a physical behavior or were approached by a male with a physical behavior there was a chi square of 14.23 which was significant at a p-value of .0002. This showed that women of high attractiveness approach males with physical behaviors and were approached by males with physical behaviors significantly more frequently than did women of lower attractiveness [See Figure 4].

To test the fourth sub-hypothesis, that male and female judges will not differ in their ratings of the attractiveness of female subjects we used a Logistic Regression Test. This test looked at whether gender made a difference in the ratings of attractiveness. An insignificant test supports our sub-hypothesis that males and females rate the attractiveness of subjects similarly. With a p-value of .1628, we accepted the null hypothesis that found that males and females do not differ in their ratings of the attractiveness of female subjects.

We acknowledge that when running our tests, we reanalyzed our data several times which may increase the chances of getting a significant result. After the initial test of total approaches, each subsequent test is not considered independent. Therefore, although nine of the ten Wilcoxon Rank Sums Tests were significant; the fact that nine of these tests were dependent on the tenth may have given us a higher number of significant results than if we had performed fewer analyses. See the Appendix for further data analysis.

DISCUSSION

Could it be true, that in today’s society beauty is still rewarded over any other attribute? It would seem that it is, at least it is in the sexually charged context of a college bar or party. In these settings, each individual present agrees on who is attractive and who is not. We found that facial attractiveness is rated relatively consistent by both male and female college students, as results show from a comparison of male and female judges’ ratings of the subject pool. Our analysis has revealed that college-aged women, rated high on attractiveness, are more popular overall than are women of lower attractiveness. Women of high attractiveness are not only involved in more approaches, but are approaching significantly more members of both sexes - reflecting greater popularity. Higher rates of socializing behavior (measured by number of approaches) suggests that attractive women may, in fact, be more socially skilled, capable of initiating more interactions and increasing their popularity. Greater social behaviors indicate skill, because it reflects a level of comfort within the environment and with others, which would confirm Goldman & Lewis’s (1997), as well as Berry & Miller’s (2001) findings. The difference in popularity
between women of high attractiveness and women of lower attractiveness is in their respective total interactions with males. We found that women of high attractiveness not only approached more males in general, but interestingly, they were approached by males significantly more than were women of lower attractiveness. Relative to the data we collected about total approaches with females, where women approach and are approached by each other with no significant attention to attractiveness, the higher frequency of male approaches with women of high attractiveness indicates that males have a larger influence on the popularity of women of high attractiveness than females.

The influence of males on the overall popularity of women is clear yet the rationale behind their influence is murky. One reason males may associate with women of higher attractiveness more than with women of lower attractiveness is that the status of a man may be inflated by the presence of an attractive woman, as found by Berscheid & Walster (1974). This study further found that the presence of an unattractive woman would lead to a negative evaluation of the male by onlookers. Thus, there is significant social motivation for males to associate themselves with highly attractive females since the advantages of her beauty may transfer to him. The "beautiful is good" stereotype indicates just how vast the positive traits associated with attractive people can be; good-looking people are judged by others to be more intelligent, successful, confident, assertive and happy. In addition to elevating a male’s status, the presence of an attractive female could simply be inherently rewarding for aesthetic reasons, the same way it is rewarding to view a beautiful work of art. Buss (1989) argues that this intrinsic value of a beautiful face can be seen as a woman’s contribution to a relationship. Men, on the other hand, are expected to offer worldly success, while women are rewarded with the fruits of that success in exchange for her beauty. In a study by Buss (1989), large numbers of men and women across cultures were asked to rank order attributes in order of importance when choosing a mate. Good looks were valued more by men while women valued good financial prospect, rendering women, predictably, "sex objects", and men "success objects." This would coincide with the evolutionary perspective which finds there to be a universal tendency for men to seek younger women (those who are most fertile) and women to desire older men (those most likely to have financial resources). Across cultures, Buss (1989) found that on average, men wanted to marry a woman 2.7 years younger, and women wanted to marry men who were 3.4 years older (Brehm, Kassin and Fein, 1999).

The present study indicates how males influence the popularity of highly attractive women. What this study has not explored is how the social setting itself sets up a dynamic that influences the popularity of these highly attractive women. It is unclear whether popularity is stable over contexts. Women of high attractiveness may be frequenting the college bars or parties where this study was conducted more often than women of lower attractiveness. If they do spend more time at these bars and college parties, the number of approaches they participate in may reflect an inflated popularity. By law of probability, the more time an individual spends at a bar or party, the more likely she is to meet those who also frequent bars and parties often. In addition, women of high attractiveness may be visiting these venues more because they may get lots of attention there. Women of lower attractiveness, however, may be more popular in other contexts such as class, athletic fields, dining halls or coffee shops. In addition, the higher level of male approaches that women of high attractiveness garner in college bars or parties may be due to the effects of alcohol lessening men’s inhibitions and making women of high attractiveness more approachable. Further research dealing with the effects of context on popularity would be useful in determining the effects of these confounding variables.

The context of the present study was that of college bars or off campus Greek parties. A widely held belief about these situations, which the researchers of this study will endorse, is that bars and parties are charged with sexual energy. Young men and women with alcohol and the freedom to act upon their impulses create an environment which can be laden with sexual undertones. This can account for findings of high levels of approaches made with physical behaviors. Women of high attractiveness approach males with physical behaviors (i.e. hugs, touches, kisses, etc.) and are approached by males with physical behaviors at a significantly higher frequency than are women of lower attractiveness. Thus, the popularity of a woman of high attractiveness could be significantly influenced by sexual motives due to the context. Approaches made or received with touches, kisses, or hugs, could be an indication of her sexual desires or appeal rather than her popularity. Research on the sexual nature of popularity would be interesting to pursue in order to determine if popularity has a sexual component.

This study attempted to shed light on one facet
of the relationship between attractiveness and popular-
ity. It is clear that attractive women are more popular in college bars and off campus Greek parties in that they are approached and approach others significantly more than women of lower attractiveness. In order to fully understand the nature of popularity and the impli-
cations of attractiveness on it, several more questions need to be answered. As mentioned above, the stabil-
ity of popularity over context is yet to be determined. It has been suggested by the social conditioning perspec-
tive that those who are more attractive through out their lifetimes receive greater positive attention from others making them more socially comfortable, more socially skilled, and more popular. What is unclear is whether attractiveness is stable over time and whether these attractive people have actually been able to develop these social skills.

On another note, this study did not attend to the appearance of the individuals who were approached by the subjects or who approached the subjects. An inter-
esting area of further research would be to look at the attractiveness of these individuals. Does the matching hypothesis hold up? Do people tend to approach people who are equivalent in their physical attractiveness? Does that mean that popularity is relative among levels of attractiveness? It is possible that attractive people are popular only among other attractive people. Finally, since this study examined a rather homogenous sample consisting of only 20- to 22-year-old University of Pennsylvania students who are members of one sorority, further research on a more representative pop-
ulation would be more generalizable.

APPENDIX

Figure 5: This graph shows that women of high attractiveness approach males significantly more than do women of low attractiveness.

Figure 6: This graph shows that women of high attractiveness approach females significantly more than do women of low attractiveness.

Figure 7: This graph shows that females approach women of high attractiveness significantly more than they approach women of low attractiveness.
Figure 8: This graph shows that women of high attractiveness are involved in significantly more approaches with males than are women of low attractiveness.

Figure 9: This graph shows that women of high attractiveness are involved in significantly more approaches with females than are women of low attractiveness.

Appendix Note: each dot may represent more than one subject.

REFERENCES


The purpose of this study was to determine the level of agreement between parent and child perceptions of the child’s emotional and behavioral problems and to discover what particular factors predicted this agreement. Our sample contained 139 subjects in an outpatient clinic setting whose ages ranged from 11 to 18 years. Parents completed the Child Behavioral Checklist and children completed the Youth Self Report. Agreement was measured by mean differences between the CBCL and YSR scores of subjects. Parents consistently reported a greater severity of problems than the children on all subscales. Analyses of variance determined significant effects for type of problem, age, and child functioning. Higher agreement was found for Externalizing problems. As age increased, children’s agreement with their parents converged on the Externalizing scale. As functioning in the child decreased, agreement was found to decrease for the Total, Internalizing, and Externalizing scales.

INTRODUCTION

The level of agreement between parent and child perceptions of the severity of the child’s emotional and behavioral problems has been an area of examination in the past. These various studies aimed, for the most part, at determining reporter validity of the child's problems. Because the observations of parents and various other sources are used in the clinical setting to evaluate the child's problems, which then aid in establishing a treatment, discrepancies between parent and child perceptions hold many implications. Parents are the most frequent observers of children's behavior in the majority of situations, while children themselves may be the most accurate in rating their own first-hand experiences. However, parents may only be able to accurately observe situational, external, and obvious behaviors of the child, while the child's maturity and emotional problems may impair the validity of his or her own observations. For the purposes of efficient and valid treatment effectiveness studies, researchers and clinicians are interested in determining who is the most accurate reporter of the child's symptoms. However, a focused study on agreement is also essential in discerning how perceptions vary according to particular variables instead of simply focusing on the validity of the reporter. In cases of high agreement and in cases of high discrepancy it is important to extract what factors are influential or determinant of these cases. Specific factors of the situation, such as characteristics of the parent or child, may be the key to determining who is the best reporter in a particular situation. In considering these factors, this study looks to variables that would be affected by such considerations. The main question this study attempts to answer is what is the level of agreement between parent and child perceptions of the child's emotional and behavioral problems and what are the significant factors that are related to this agreement.

There have been studies in the past addressing the issue of parent and child agreement of emotional and behavioral problems, but most of the findings have been weak and inconsistent. The most conclusive finding concerns the setting of the subject. In clinical studies, parents have consistently reported greater scores on the Child Behavioral Checklist (CBCL) than their children's scores on the Youth Self Report (YSR) indicating that parents perceive a greater severity of problems than do children. However, in nonclinical settings, children were consistently found to report higher scores than their parents. Only one study, by Kolko & Kazdin (1993) found nonclinical patients to report a lesser
severity of problems than their parents.

Studies have also frequently looked at the type of problem. Agreement in most studies tends to be higher for the Externalizing scale, as well as more significant. Items from the Aggressive Behavior and Delinquent Behavior subscales make up the Externalizing scale while items from the Withdrawn, Somatic Complaints, and Anxious/Depressed subscales comprise the Internalizing scale. These subscales are generally characterized as symptoms that are under-controlled or overcontrolled (Achenbach, McConaughy, & Howell, 1987). Most studies have concluded there to be a higher agreement for Externalizing problems than for Internalizing problems because of the nature of their observability and interpretability. But there are also contradictory findings. In two separate studies conducted by Handwerk, Larzelere, Soper, & Friman (1999) and Verhulst & van der Ende (1992), it was found that there was a greater agreement between parent and child on Internalizing scores. The first sample examined clinical patients in different settings and the latter study utilized the general population.

Gender has also been an inconsistent variable. Generally if a sex difference has been found to be an effect, girls tend to have greater agreement with parents than boys. These were the results in the studies of Kolko & Kazdin (1993) and Verhulst & Van der Ende (1992). However, this discovery is not absolute. Often times there are no significant differences between gender groups as in the studies of Handwerk, Larzelere, Soper, & Friman (1999) and Achenbach, McConaughy, & Howell (1987). Findings concerning age seem to be fairly consistent. In one study (Achenbach, McConaughy, & Howell, 1987), younger children, categorized between the ages of 6 and 11, were reported to have significantly more agreement with parents on the Internalizing scale. This pattern is also reflected in another study (Handwerk, Larzelere, Soper, & Friman, 1999) in which the discrepancy was larger for older youth (ages 15 to 18) than for younger youth (ages 11 to 14) on the Internalizing scale. In the study by Verhulst & van der Ende (1992), there was a non-significant finding for decreased agreement as age increased, with higher scores on the YSR than the CBCL. There have also been a few studies that have not uncovered any significant age differences, but for the most part, increasing age tends to correlate with a greater parent and child discrepancy.

One particular study, conducted by Kolko & Kazdin (1993), found the effect of parent functioning and family stress to be of the most salient predictors for parent-child agreement. Family stress was measured by the Children's Life Inventory Events (CLIE), in which parents documented potentially stressful life events of a child. Parent functioning was assessed with the Hopkins Symptom Checklist (SCL-90), which evaluates the degree of discomfort caused by symptoms of several emotional dimensions. Parent function was found to be the most salient predictor of parent and child agreement, beyond all other variables studied, such as age, gender, clinical status, type of problem, and demographical variables.

Taking from and building upon these past studies, our particular study examines age, gender, and type of problem. The child's level of functioning and parental distress level were additional variables examined that were unique to this study. The child's level of functioning may influence agreement because a child who is not functioning well may not be an accurate judge of his or her own emotional and behavioral state, or may heighten a parent's awareness of the child's emotional and behavioral state. Decreased functioning may also be reflective or determinate of problematic emotional and behavioral states. It was predicted that a higher level of parental distress would correlate to a lower level of agreement between parent and child because greater stress on the parent's part would probably create conflicts in the relationship, leading to contrasting perspectives. Hypotheses for age, gender, type of problem, and setting were in accordance to the majority of findings in previous studies. Agreement would probably decrease with age because the adolescent child is perhaps more rebellious or independent of his or her parent's views. Gender would most likely not produce any significant differences between groups, but if there were a significant finding, it would probably show that girls have a higher agreement with parents than boys. This is due to the general sentiment that girls are more willing to share their emotional and behavioral problems because they are more expressive. Externalizing problems would have a greater agreement because of their observability, and since this study was conducted in a clinical setting, parents were expected to report a greater severity of problems than their children would.

**METHOD**

**Participants**

This study stems from a pediatric mental health
outcomes initiative (PMHOI) recently implemented in the Division of Child and Adolescent Psychiatry at Stanford University. The sample of subjects included in this study consisted of parents and their children who sought treatment through the outpatient Child and Adolescent Psychiatry clinic during a one-year period. All subjects with a complete Child Behavioral Checklist (CBCL) and Youth Self Report (YSR) were included in the study. The resultant sample included 139 children. Seventy-four subjects were male and 65 were female. The ages of these subjects ranged from 11 to 18 years, with a mean age of 13.8 years. The age range of these subjects was limited by the age restrictions of the YSR. Income was coded into bracketed categories, and the average income level was within the $75,000 to $100,000 per year range. The majority of subject's salaries fell into the $100,000 to $125,000 per year range. Seventy-eight point four percent of children came from 2-parent families. Seventy-five point seven percent of children were Caucasian, eight point eight percent Asian, and seven point four percent were Hispanic. Thirty-one point one percent of parents were college graduates and twenty-nine point four percent had a graduate degree. Only ten point nine percent of parents had an education level below that of a high school degree.

Measures

Parent perceptions of behavioral problems were assessed using the Child Behavioral Checklist (CBCL), and the child's opinion of their own problems was assessed with the Youth Self Report (YSR). These forms are standardized rating scales of emotional and behavioral problems of the child, and assess corresponding emotional and behavioral factors. The CBCL is intended for children aged 4 to 18 while the YSR is only administered to children ages 11 to 18, at which they can reliably read the form. There are eight "syndrome" scales generated from the items of the CBCL and YSR which include: Withdrawn, Somatic Complaints, Anxious/Depressed, Social Problems, Thought Problems, Attention Problems, Delinquent Behavior, and Aggressive Behavior. Of these, the Withdrawn, Somatic Complaints, and Anxious/Depressed subscales comprise the Internalizing problems scale, whereas Delinquent and Aggressive Behaviors subscales comprise the Externalizing scale. Both questionnaires yield a Total scale that includes all subscales. For both measures, raw scores were converted to standardized T scores.

The Columbia Impairment Scale, Revised (CIS) was filled out by parents to determine the child's level of functioning. The questions on this questionnaire address a child's behavioral functioning across several domains, such as sibling relationships, school behavior, involvement in activities, and agreeability. There are thirteen questions which are rated on a scale of 0 to 4.

The Parenting Stress Index (PSI) was used to determine the level of stress both parents experience in relation to their child. Both parents filled out this form. It contains 36 questions that are scored on a 0 to 5 point scale. Responses yield five scale scores: Defensive Responding, Parental Distress, Parent-Child Dysfunctional Interaction, Difficult Child, and Total score. Each subscale refers to a domain representing the source of a parent's stress. Defensive Responding is a validity scale indicating that a respondent with a score equal to or less than ten is overtly defensive.

Procedure

Data was collected as a part of the PMHOI. Parents and children were mailed a packet containing demographic, medical history, and emotional/behavioral assessment forms prior to their appointment. Data from the current and previous year was taken from the database of the PMHOI study and analyzed in this study. The only form completed by both parents was the PSI. Only one CBCL and CIS questionnaire was given to each family, and most of the time the mother completed these forms. Although completion was voluntary, almost all packets were mailed back by the parents. Cases in which parents scored their children as overtly defensive on the Defensive Responding scale of the PSI were excluded from analyses. For the purposes of these analyses each set of parents' scale scores on the PSI were averaged to create a single set of parental stress scores per household. An average of the raw scores of both parents were taken if both filled out the form. In our analysis only the mean total score was used. Statistical analysis of all data was performed by multiple regressions, analyses of variances, and multiple analyses of variances through computer software.

RESULTS

The descriptive statistics provided a summary of the general shape of our data (Table 1). The age of subjects ranged from 11 to 18 years, with a mean age of 13.77 (SD = 2.72). Parents reported greater severity of emotional and behavioral problems on the CBCL than
their children did on the YSR for all three scales examined (Total, Internalizing, and Externalizing). Scores seemed to be relatively balanced on both ends and generally had a mean centering the clinically borderline score of T>60. The CIS scores were mostly on the lower end of the functioning scale. The maximum score reported was 48 out of 72 with a mean score of 22.33, indicating that not many subjects were in the severely impaired functioning range. The same was true of the PSI Mean Total scores, which had a maximum score of 136.5 out of a possible 180, with a mean of 61.53.

Mean scores for the CBCL and YSR subscale showed that parents consistently rated higher problem behavior scores than children (Table 2). Scores for the Internalizing problems were higher for both the CBCL and the YSR than for the Externalizing scores. Mean scores for each subscale of the CBCL ranged from 60.93 (Aggressive) to 65.09 (Withdrawn and Anxious/Depressed). Mean scores for each subscale of the YSR ranged from 57.04 (Aggressive) to 60.09 (Anxious/Depressed). The CBCL scores were all generally in the borderline clinical range or slightly above borderline, while all the YSR scores were slightly below borderline.

Agreement was determined by mean differences, which were calculated by subtracting the YSR scores from the CBCL scores. Paired sample t-tests were used to determine if differences between the CBCL and the YSR scores were of any significance. All differences between parent and child were reported to be significant at p < .000, except for the Delinquent scale, which had a p < .001 (see Table 2). Discrepancies ranged from 3.89 points (Aggressive) to 6.85 points (Withdrawn) on the subscales. The mean difference for the Externalizing scale was only 4.56 while the mean difference for the Internalizing scale was 7.65. Difference scores on the major broadband scales were used as the dependent variables for all remaining analyses.

Because gender was a categorical variable, a one-way ANOVA was used to test if discrepancies of scores between children and their parents were related to gender. Gender did not turn out to be significantly associated with the CBCL-YSR differences on any of the broadband scales.

Separate multiple regressions were then run for each broadband scale against all variables to uncover which particular variables would reveal themselves as influential predictors of differences between the CBCL and the YSR scores (Tables 3-5).

The results of these regressions consistently showed a huge significance for CIS as a predictive variable with all p-values equal to or lesser than .001. It was significant across all broadband scales, and therefore, for all types of problems. Age became significant at a level of .05 in the External Difference Regression model with a p-value of .038, but it remained insignificant for the other models. No other variables were found to be statistically significant.

A multiple regression was then run with only age and CIS as predicting variables to exclude the noise of irrelevant variables and determine the weight of influence of each variable. Two different models were run: one with age alone, and the other with both age and CIS. This was performed to determine whether age could be a significant factor by itself.

Functioning was responsible for the significance of Model 2 on all scales. Also, age was significant and added predictive value only on the Externalizing scale, in accordance with the original multiple regression. Model 2 produced the greatest F-values indicating the strongest models on the Externalizing, Total, and Internalizing scales respectively. Model 1 had the next greatest F-values on the Externalizing scale. No other regressions were significant. The beta value for age on the externalizing model was -.189, while the beta values for CIS were .365, .285, and .385 for Total, Internalizing, and Externalizing, respectively.

### Table 1. Descriptive Data of Continuous Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>13</td>
<td>18</td>
<td>13.77</td>
<td>2.72</td>
</tr>
<tr>
<td>CBCL Total</td>
<td>40</td>
<td>89</td>
<td>65.40</td>
<td>8.88</td>
</tr>
<tr>
<td>YSR Total</td>
<td>30</td>
<td>81</td>
<td>57.69</td>
<td>10.66</td>
</tr>
<tr>
<td>CBCL Internalizing</td>
<td>32</td>
<td>87</td>
<td>65.21</td>
<td>10.71</td>
</tr>
<tr>
<td>YSR Internalizing</td>
<td>26</td>
<td>85</td>
<td>57.62</td>
<td>11.61</td>
</tr>
<tr>
<td>CBCL Externalizing</td>
<td>33</td>
<td>85</td>
<td>59.72</td>
<td>11.22</td>
</tr>
<tr>
<td>YSR Externalizing</td>
<td>30</td>
<td>87</td>
<td>55.16</td>
<td>11.17</td>
</tr>
<tr>
<td>CIS</td>
<td>3</td>
<td>48</td>
<td>22.33</td>
<td>9.36</td>
</tr>
<tr>
<td>PSI Mean Total</td>
<td>0</td>
<td>136.5</td>
<td>61.53</td>
<td>37.00</td>
</tr>
</tbody>
</table>
### Table 2. Mean Differences Between CBCL and YSR by Subscale

<table>
<thead>
<tr>
<th>Scale</th>
<th>Mean Difference</th>
<th>Std. Deviation</th>
<th>T</th>
<th>Sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Withdrawn</td>
<td>6.8511</td>
<td>11.8025</td>
<td>6.893</td>
<td>.000</td>
</tr>
<tr>
<td>Somatic</td>
<td>3.1844</td>
<td>10.2933</td>
<td>3.674</td>
<td>.000</td>
</tr>
<tr>
<td>Anxious/Depressed</td>
<td>4.9929</td>
<td>10.6009</td>
<td>5.593</td>
<td>.000</td>
</tr>
<tr>
<td>Social Problems</td>
<td>3.4539</td>
<td>10.0779</td>
<td>4.070</td>
<td>.000</td>
</tr>
<tr>
<td>Thought Problems</td>
<td>6.2340</td>
<td>10.2836</td>
<td>7.198</td>
<td>.000</td>
</tr>
<tr>
<td>Attention Problems</td>
<td>6.4752</td>
<td>11.1941</td>
<td>6.869</td>
<td>.000</td>
</tr>
<tr>
<td>Delinquent</td>
<td>2.4752</td>
<td>8.6426</td>
<td>3.401</td>
<td>.001</td>
</tr>
<tr>
<td>Aggressive</td>
<td>3.8865</td>
<td>9.7980</td>
<td>4.710</td>
<td>.000</td>
</tr>
<tr>
<td>Total</td>
<td>7.7660</td>
<td>11.1692</td>
<td>8.256</td>
<td>.000</td>
</tr>
<tr>
<td>Internalizing</td>
<td>7.6454</td>
<td>12.5574</td>
<td>7.230</td>
<td>.000</td>
</tr>
<tr>
<td>Externalizing</td>
<td>4.5571</td>
<td>11.9147</td>
<td>4.526</td>
<td>.000</td>
</tr>
</tbody>
</table>

### Table 3. Coefficients for Total Difference Regression

<table>
<thead>
<tr>
<th>Variable</th>
<th>Beta</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>-1.712</td>
<td>.372</td>
</tr>
<tr>
<td>Age</td>
<td>-.525</td>
<td>.138</td>
</tr>
<tr>
<td>CIS</td>
<td>.440</td>
<td>.000**</td>
</tr>
<tr>
<td>Mean Total</td>
<td>-6.79E-03</td>
<td>.794</td>
</tr>
</tbody>
</table>

R Square value for model: .161

### Table 4. Coefficients for Internal Difference Regression

<table>
<thead>
<tr>
<th>Variable</th>
<th>Beta</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>-.911</td>
<td>.682</td>
</tr>
<tr>
<td>Age</td>
<td>-.348</td>
<td>.395</td>
</tr>
<tr>
<td>CIS</td>
<td>.400</td>
<td>.001**</td>
</tr>
<tr>
<td>Mean Total</td>
<td>-1.61E-02</td>
<td>.595</td>
</tr>
</tbody>
</table>

R Square value for model: .094

### Table 5. Coefficients for External Difference Regression

<table>
<thead>
<tr>
<th>Variable</th>
<th>Beta</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>-.172</td>
<td>.933</td>
</tr>
<tr>
<td>Age</td>
<td>-.782</td>
<td>.038*</td>
</tr>
<tr>
<td>CIS</td>
<td>.495</td>
<td>.000**</td>
</tr>
<tr>
<td>Mean Total</td>
<td>-1.08E-03</td>
<td>.969</td>
</tr>
</tbody>
</table>

R Square value for model: .181

### Table 6. Model Summary for Total, Internalizing, and Externalizing Differences Regressions

<table>
<thead>
<tr>
<th>Scale</th>
<th>Model</th>
<th>R</th>
<th>R Square</th>
<th>Std. Error</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Difference</td>
<td>1</td>
<td>.163</td>
<td>.026</td>
<td>11.0328</td>
<td>3.727</td>
<td>.056</td>
</tr>
<tr>
<td>Total Difference</td>
<td>2</td>
<td>.399</td>
<td>.159</td>
<td>10.2897</td>
<td>12.894</td>
<td>.000**</td>
</tr>
<tr>
<td>Internalizing</td>
<td>1</td>
<td>.100</td>
<td>.010</td>
<td>12.4083</td>
<td>1.397</td>
<td>.239</td>
</tr>
<tr>
<td>Internalizing</td>
<td>2</td>
<td>.302</td>
<td>.091</td>
<td>11.9332</td>
<td>6.818</td>
<td>.002**</td>
</tr>
<tr>
<td>Externalizing</td>
<td>1</td>
<td>.191</td>
<td>.029</td>
<td>11.7893</td>
<td>5.136</td>
<td>.025*</td>
</tr>
<tr>
<td>Externalizing</td>
<td>2</td>
<td>.430</td>
<td>.173</td>
<td>10.8841</td>
<td>15.295</td>
<td>.000**</td>
</tr>
</tbody>
</table>

Model 1 Predictors: Age
Model 2 Predictors: Age, CIS

** Correlation is significant at the 0.01 level (2-tailed)
* Correlation is significant at the 0.05 level (2-tailed)
DISCUSSION

Our clinical settings produced greater scores of emotional and behavioral problems reported from the parent's perspective. This was consistent with studies of CBCL and YSR report comparisons of the past. All differences between subscale scores were found to be significant. This discrepancy may be interpreted in a number of ways. For example, parents may be reporting more severe levels of emotional distress and behavioral problems in order to validate seeking out treatment. Children may also be denying the presence of problems because they see their behavior as justified or because they may be unaware of the problems. Having been introduced into a clinical setting, sometimes unwillingly, may also be causing the child to deny particular problems. Also, children may not see their behavior through the perspective of being defined as a problem, whereas parents may be more prone to label or readily assume that their child's behavior may be characterized into a preexisting socially defined problem. Another possibility is that the YSR is worded slightly offensively to adolescents and is therefore more likely to produce a defensive response to its questions (e.g. "I don't feel guilty after doing something I shouldn't," "I threaten to hurt people").

The mean differences found between the CBCL and the YSR measures on all subscales were found to be significant. The Delinquent subscale was found to have the highest correlation and the least mean difference, indicating that parents and children are in greater agreement on this subscale. Possible explanations could include that children are admittedly more capable of identifying delinquent problems and also that parents are able to identify them because they are easily observable. Convergence of scores may also be due to the fact that opinions concerning the boundaries of delinquent behavior are rigidly defined during childhood and adolescent years in classrooms and by parents themselves. Parents and figures of authority are continually enforcing the social constructs of acceptable behavior upon the children. This may lead to clearly defined and easily labeled behavior for both parent and child. The largest mean difference occurred on the Withdrawn subscale in which parents reported a greater severity of problems than their children did. The Withdrawn symptoms are less obvious and overt than Delinquent symptoms, although still easily observable. The discrepancy observed here might be due to parents over-reporting problems while the child fails to perceive a problem. It may also be due to children under-reporting their emotional and behavioral state because they are unfamiliar with it and unable to recognize it. On the broadband scales, the differences on the Externalizing scale were almost half those of the Internalizing scale. This has been a well-documented result of various studies in which parents and children are consistently in greater agreement on Externalizing problems. This is most likely because overall Externalizing problems are more overt and evident, and thus more noticeable by the parent, as is the case with Delinquent symptoms. Internalizing symptoms are also harder to detect, and this decreases the parent's ability to judge their child's emotional state.

There have been weak findings in the past concerning gender differences in the reporting of problems. Some studies have found greater parent-child agreement with girls, while other studies have failed to find any significant differences based on gender. Our study also failed to find any significant effects of gender on parent-child agreement of emotional and behavioral problems. The hypothesized reason behind gender differences is that girls are more open to discuss their emotional and behavioral issues with their parents. However, our data suggests that gender, and perhaps female communication behavior, does not predict differing levels of agreement between parents and children.

The study done by Kolko & Kazdin (1993), which found heightened family stress to be a predictor of parent-child agreement, was not reflected in our study. In that particular study parental stress was assessed using the Hopkins Symptom Checklist, which evaluates overall parental dysfunction, and the Children's Life Events Inventory, which documents potentially stressful life events to a child and family. In our analysis, the mean PSI total was not found to be significant for any of the broadband scales and also had weak correlations. The difference in measures between the Kolko & Kazdin study and ours is that the PSI focuses more upon the parents' level of satisfaction in their parental role while the Hopkins Symptom Checklist deals more with the degree of negative emotional symptoms of the parent. This is important to note because it suggests that how parents view themselves in their role relative to their child does not have an impact on the scores reported in the CBCL or the YSR, and that a parent's emotional state is more of a significant influencer.

The remaining variables which were found to be significant predictors of CBCL and YSR agreement...
were age and functioning of the child. From the multiple regressions it was apparent that type of problem was also significant in determining parent-child agreement. There was much greater agreement on the Externalizing scale than the Internalizing scale, as was found in the majority of previous studies, perhaps suggesting that Externalizing symptoms are easier to judge from a parent's perspective and easier to identify from the child's perspective. Age was found to be significantly predictive by itself on the Externalizing scale only. It was never found to be significant on the Internalizing scale, but it did add predictive value to the Total scale. On the Externalizing scale, age had a slope of -.189 in the final model (Model 2), which signifies that for every increase in year of age there is a .189 decrease in the mean CBCL and YSR T score difference. So as children grow older, their perceptions of emotional and behavioral problems begin to converge more with their parent's perceptions, although at a relatively slow rate.

These findings are in opposition to the previous studies that have found that as children get older, they tend to have greater discrepancies with their parents concerning their reported emotional and behavioral problems. This irregular finding may be specific to our sample population and demographics. Previous studies have found that higher socioeconomic communities produce higher levels of agreement between parent-child ratings that were not necessarily dependent upon age. The majority of our subjects reported an annual household income of $100,000 to $125,000. A high percentage of the parents in our sample also had college or graduate degrees, and the majority of the children were part of a two-parent household. It is possible that parents in our sample are accurate and valid evaluators of their child's behavior because they are well educated in these areas, and as children grow older and become more educated in these areas as well, their perceptions begin to converge with their parents'. The fact that such an overwhelming majority of children are from non-divorced families also indicates family stability, which may further encourage emotional stability and communication. Keeping in mind that age is only significant in the Externalizing and Total difference regressions, another hypothesis to account for the convergence of agreement as children get older is that their Externalizing problems have been repeatedly identified and addressed because they appear to be more intrusive than other problems. These problems receive more attention and therefore children are admittedly more aware of the behaviors. Internalizing differences may not be affected by age because of the fact that its symptoms are difficult to observe and are also discussed less publicly because of particular social stigmas.

The child's functioning score turned out to be the greatest predictor of CBCL and YSR agreement. It was positively correlated and extremely significant for all types of problems. The greater the CIS score, the more severely impaired the functioning, and the greater the discrepancy between the CBCL and the YSR. However, like the CBCL, the CIS is also reflective of the parent's perception of their child's behavior. But the important distinction between the CBCL and the CIS is that the CIS aims at determining the child's overall level of functioning whereas the CBCL targets a child's emotional and behavioral problems. The relation of the CIS to the CBCL-YSR difference indicates that as a child's level of functioning decreases, as evidenced by a higher score, there will be greater discrepancies between the parent and child scores on the CBCL and the YSR. The fact that discrepancy is affected by a characteristic of the child and their functioning, and was unaffected by parental stress suggests that children may not be the best judge of their emotional and behavioral state and that they may be more easily influenced by other emotional and psychological factors. A child's impaired functioning may decrease his or her ability to accurately judge his or her own behavior, thereby creating a larger discrepancy as functioning decreases. Impaired functioning may lead to distorted and inaccurate views of the child's own behavior, causing him or her to rate the problems as less severe than they actually are.

Parent-child agreement was mainly found to be affected by the child's level of functioning, type of problem, and, on the External scale, by the child's age as well. The fact that agreement was indeed found to be dependent upon particular variables indicates that CBCL and YSR agreement is not merely a repetition of information but that it in fact reflects certain features of the child's age, functioning, and type of problem. The examination of sources of agreement in this study allows one to understand the source of discrepancy. The conclusion that this study allows us to make is that parents are most valuable reporters for younger children and lower functioning children than for older and higher functioning children. From our findings it appears that agreement is mostly dependent upon characteristics of the child. This is highly probably because children are more variable in their understanding of their environment, whereas adults are by far and large, less susceptible to extraneous influences affecting their
judgment. An important factor to keep in mind is the specificity and uniqueness of our sample population, which consists of a particularly high socio-economic group with most subjects coming from a well-educated, two-parent household. The majority of the subject's functioning was also mainly clustered around the less severe end of the spectrum, as was apparent from the CIS and mean PSI total descriptives. Our findings would need to be replicated in other samples with differing demographics in order to generalize our results.

Although the predictive variables of agreement in this study were mostly dependent upon the characteristics of the child, this does not necessarily conclude that parents are the most valid and accurate reporters of their child's behavior. It is evident that agreement is dependent upon various variables, and this may affect who is the most valid predictor of the child's emotional and behavioral problems depending on characteristics of the child and reporter. What is revealed is that parents are good predictors for a particular type of child within a particular type of population. One of the limitations of this study was that there was insufficient third party data (such as from the patient's clinician) to provide a clinical perspective of the child's emotional and behavioral problems. The PMHOI has already begun to increase its database for clinician and teacher reports, which can be used for later studies to determine what situations or what characteristics of the child would enable them to be the most useful observer and reporter of a child's problems. The discovery of factors influencing agreement provides insight into what behaviors or features an effective treatment should attempt to target, and this focus would lead to a more efficient and informative treatment evaluation process.

REFERENCES


Does Subjective Well-Being Increase with Age?

Brian Scott Ehrlich and Derek M. Isaacowitz

Individuals vary in their levels of Subjective Well-Being (SWB). SWB is a measure of how good an individual feels about his/her life at a moment in time. Early research predicted that SWB was influenced by a host of sociodemographic variables that explained individual differences in SWB (Diener, Suh, Lucas & Smith, 1999). SWB is now considered to consist of three primary components: people’s emotional responses (both positive and negative affect), domain satisfactions, and global judgments of life satisfaction (Diener, Suh, Lucas & Smith, 1999). Presently it is thought that older individuals have higher levels of SWB up to a certain age (Isaacowitz & Smith, 1999). The current study will use emotional responses and life satisfaction as the lens for investigating the differences in levels of SWB across age groups.

INTRODUCTION

Age and Emotional Response

Mroczek and Kolarz (1998) investigated affect among a nationally representative sample of 2,727 people aged 25 to 74, conducted by the McArthur Foundation Research Network on Successful Midlife Development. They administered the Midlife Development Inventory (MIDI affect scales) in an initial phone survey and a follow-up survey. They concluded that there were higher levels of positive affect in old age. Similarly, they found that older adults experienced lower levels of negative affect than younger adults. Isaacowitz & Smith (1999) examined the claim of higher SWB with age by analyzing data in the Berlin Aging Study (BASE). By extending the research of Mroczek & Kolarz into old age, this research found lower levels of positive and negative affect among the "oldest-old" age group (participants aged 70-100 years old). The research attributes this decline in advanced old age to health related problems and illnesses that preclude typical functioning, causing these older people to be less able to perform everyday tasks and activities. Isaacowitz & Smith (1999) posited that lower levels of negative affect do not equate to increased SWB. In fact, socio-emotional selectivity theory contends that while people may become better emotional regulators as they age, they do not necessarily have higher positive affect (Cartensen, Isaacowitz, & Charles, 1999). While there seems to be higher positive affect in adults as evidenced in the Mroczek & Kolarz article, the oldest-old age group showed lower levels in positive affect. According to Isaacowitz & Smith (1999), avoidance of negative affect can be a result of better emotional regulation.

Depression is perhaps the most extreme form of negative affect. The literature on the relationship between major depression and depressive symptoms reveals important conclusions about its age-related differences. Depression rates have been studied across age groups. The literature in this field can be divided into two groups. The first group of literature focuses on major depression as a unit of analysis while the second group focuses on depressive symptoms.

One study of major depression, Regier et al. (1988), measured the prevalence of mental disorders in the general population using the Diagnostic Interview Schedule (Regier et al., 1988). The highest rates for major depressive episodes were diagnosed in the two youngest age groups (18-24 yrs old, 25-44 yrs old) with prevalence rates of 2.2% and 3.0% respectively. The oldest age group (65 years and older) had the lowest prevalence rate for major depressive disorders of 0.7%. It may be that this age difference can be attributed to the stigma associated with such disorders and the potential reluctance and/or possible memory retrieval-related problems that older participants might face. Since older people may experience decreased functioning, it may be that the low prevalence rate is simply a function of older people forgetting to report symptoms.
in the interview process. Still, there is robust data to suggest that the previous trend of increases or stability in positive affect applies to that of the negative affect present in mental disorders.

Concerning depressive symptoms among adolescents and adults, it has been found that adolescents show high levels of depressive symptoms, and older people show low levels of depressive symptoms (Nolen-Hoeksema, 1988). Nolen-Hoeksema posits that the age differences may be attributed to the frequency of parental divorce and parental depression that have affected childhood in more recent cohorts. Conversely, elderly individuals do not experience many uncontrollable or aversive events and thus do not have a greater risk of depression. The literature is consistent with higher levels of less negative emotional responses in older adults.

Emotional experience is included in the component of emotional response of SWB. Carstensen, Pasupathi, Mayr, & Nesselroade (2000) explored age differences in emotional experience in adults 18 to 94 years old. Participants were electronically beeped and asked to report the degree to which s/he experienced a spectrum of emotions across a one-week period. While this research did not find age differences in the frequency of positive emotional experiences, it did find an age difference in the frequency of negative emotional experiences. Negative emotional experiences seemed to decline up until age 60 and then ceased to decline, showing a resurgent, albeit non-significant, upward trend in the frequency of negative emotional experiences. This research provides another piece of converging evidence that negative affect seems to stop declining at best, and is perhaps increasing in advanced old age.

Age and Life Satisfaction

Life satisfaction refers to the cognitive-judgmental aspects of SWB. While there is less literature on life satisfaction than on emotional response, Diener et al (1999) provide a summary of several studies on the age differences of life satisfaction. According to this summary, life satisfaction seems to stay the same, if not increase with age. This finding countered earlier conventional wisdom that older people were less satisfied because they were unhappy with their unfulfilled lives as they reached the uselessness of old age. The increase in life satisfaction with age may be attributed to a trend in greater involvement in satisfying areas of life among older cohorts. Nonetheless, there seems to be a slight increase in life satisfaction from age 20 to age 80 with negative affect held constant. Considering that life satisfaction stays the same or increases in old age, Diener, et al. suggests that people become better at adapting to their conditions as they get older (1999).

One piece of literature examines the levels of life satisfaction in a sample of 1,000 Canadians ranging in age from 15 to 95 (Horley & Lavery, 1995). Participants completed a number of SWB measures that included a version of the Affect Balance Scale, an 11-point life satisfaction rating, and an 11-point quality of life rating. Mean levels of life satisfaction increased in the 65-74 year old age group, but appeared to taper off in the 75 and older age group, with the oldest-old age group ceasing to exhibit the increase in higher levels of life satisfaction. One hundred and thirty-six participants completed these same SWB measures at the seven-year follow-up in the longitudinal component. Researchers found that younger people tended to report lower levels of life satisfaction over time. These two studies converge to show that adults in the older adult age group had higher levels of life satisfaction.

Summary of Past Findings

It seems that higher levels of Subjective Well Being are found in older people up to a certain age. Health related problems that occur at "advanced old age" might contribute to an upswing in negative affect and/or a lack of positive affect in the oldest-old cohorts, as well as the presence of minor depressive disorders. There is overwhelming cross-sectional evidence that suggests that the levels of positive affect and life satisfaction are at least consistent over time, if not getting higher as individuals age. In the oldest-old age group, there seems to be a resurgence in emotional responses of negative affect. Overall, we would expect to find increases in Subjective Well Being into older age up until advanced old age (Isaacowitz & Smith, 1999).

Hypothesis

The current study evaluated life satisfaction and positive and negative affect in young adults, (aged 18-25), middle-aged adults (37-59), and older people (60 and over). Using four measures of Subjective Well Being, this study assessed the levels of SWB among the age groups in the sample. It is hypothesized that we will find higher positive affect and lower levels of negative affect and depressive symptoms into old age. It is expected that life satisfaction will show similar high levels, with older adults showing higher life satisfaction.
METHOD

Participants

Two hundred and eighty research participants, ranging in age from 18 to 93 years old were recruited for this study. The sample was divided into three age/cohort groups (young people, aged 18-25; middle aged people, aged 37-59; and older people aged 60-93). There were 100 participants in the young adult cohort, 86 participants in the middle-aged cohort, and 94 participants in the older adult cohort. Participants were recruited from various community organizations. Younger participants were recruited from universities in the Delaware Valley and throughout the northeast. Middle aged participants were recruited from churches, community groups, and businesses. Older participants were recruited from several senior centers and unassisted living communities for elderly in the Philadelphia area. Older participants from assisted living communities or who experienced cognitive impairments were not recruited. Participants were excluded only if they had trouble completing the first questionnaires even when read aloud to them.

The sample was comprised of 190 women and 90 men. The total sample was predominantly comprised (80%) of participants who identified themselves as Caucasian. The sample included 224 participants who identified themselves as Caucasian, 26 who identified themselves as African-American, and 22 who identified themselves as Asian. Eight participants would not provide race/ethnicity information. Middle-aged and older cohorts had a higher proportion of Caucasians. The younger adult sample contained the largest percentage of Asian-Americans, while the older adult sample did not contain any Asian-American participants.

Measures

Emotional Responses: Positive and Negative Affect. We used the Positive and Negative Affect Schedule (PANAS) as one measure to assess emotional response in this sample (Watson, Clark, & Tellegen, 1988). The PANAS is a widely used measure of affect. It includes ten negative and ten positive adjectives that describe ways people may feel. Participants used a five-point scale to rate themselves on the degree to which they felt a certain emotion during the course of the day. Higher scores for the ten negative adjectives indicate higher negative affect, while higher scores for the ten positive adjectives indicate higher positive affect. By phrasing the question with an emphasis on the short term (i.e. today) rather than the long term, the PANAS is sensitive to fluctuations in mood (Watson, et al, 1988). These scales have been proven highly internally consistent.

Emotional Responses: Depressive Symptoms - We used the Center for Epidemiologic Studies Depression Scale (CES-D) to assess depressive symptoms (Radloff, 1977). The CES-D is a twenty-item scale that places an emphasis on the negative affective component of depressive mood. Participants mark the degree (on a scale from 0 to 3) to which they have felt or behaved the way described during the past week for each item. Four of the items are reverse coded, and a high score is indicative of a risk of depression or necessity for treatment. The CES-D was found to have a very high internal consistency, acceptable test-retest reliability (r=.54), and good concurrent validity by clinical and self-report criteria (Radloff, 1977). The measure also appears to be suitable for a host of demographic groups (e.g., socio-economic statuses and ethnicities).

Life Satisfaction - We used the Satisfaction with Life Scale (SWLS) to measure global life satisfaction (Diener, Emmons, Larsen, & Griffin, 1985). The SWLS is a five-item scale where participants rate themselves on the degree to which they agree with five statements regarding global life satisfaction. Higher scores indicate higher life satisfaction. The SWLS focuses explicitly on global life satisfaction and excludes positive affect and loneliness (Diener, et al., 1985). The measure appears to have high internal consistency and high temporal reliability.

Procedure

Participants completed the baseline interview in a one-on-one or small group setting with an interviewer from the research team. This interview included self-reported measures of positive affect, negative affect, depressive symptoms, and life satisfaction. When participants had difficulty reading questionnaires, the interviewer read the items aloud and recorded the participants' responses. The interview included measures of life events and emotional response. While this study is part of a larger longitudinal study, this paper will focus exclusively on the baseline interview. Participants were then paid ten dollars at the completion of the baseline interview.
RESULTS

Perceived health was lowest in the older adult cohorts. Participants rated their present health on a five-point Likert scale from 0 (poor) to excellent (4). Participants in the young adult age/cohort group had an average perceived health score of 3.3 (with a maximum score of 4), while participants in the middle-aged age/cohort group had an average health score of 3.01, and the participants in the older adult age/cohort group had an average health score of 2.29. Thus, there are noticeable age related differences in perceived health. Average years of education differed among the three cohorts. Middle-aged adults reported the most years of education, followed by young adults, and finally the older adults. The middle-aged cohort had an average score (determined by the number of years of education) of 17.69, the young adults had an average score of 14.77, and the older adults had an average score of 12.39. This difference between the middle-aged adult cohort and the older adult cohort may be attributed to the cohort effect that many participants from older generations were schooled to a lesser extent than those participants in the middle-aged cohort. The slightly lower educational average among the young adult sample (17.69) may be attributed to the large number of college students recruited that were in the process of completing their education.

Age groups differed significantly on mean values of depressive symptoms, F (2, 276) = 4.27, p < .05. (Mean levels are provided in Table 1). The mean levels were the highest for the younger adult cohort, followed by the older adults, and finally the middle aged cohort. When using Tukey’s Studentized Range Test, only the pair-wise difference between the young adult age group and the middle-aged group emerged as significant.

Age groups differed significantly on mean values of positive affect, F (2, 273) = 6.97, p < 0.01. The middle-aged group had the highest mean level and the younger adult group had the lowest mean level of positive affect, while the older age group fell somewhere in the middle (see Table 1). The Tukey comparisons revealed that only the pair-wise difference between the young adult age group and the middle-aged group emerged as significant. Therefore, middle-aged people were found to have significantly higher levels of positive affect when compared to the younger adults. However, there was no significant difference between older adults and the other age groups when looking at the more conservative analysis provided by Tukey’s Studentized Range Test.

Age groups differed significantly on mean values of negative affect, F (2, 273) = 17.33, p< 0.01. The mean levels show that younger people have the highest levels of negative affect, followed by the older people, and the middle aged group. Analysis by the Tukey comparisons show that the younger adults are significantly higher in negative affect when compared to the older adults and the middle aged adults, while the middle-aged people and the older people are not significantly different from each other.

Mean levels did not differ significantly in life satisfaction across age groups, F (2, 274) = 1.94, p<0.10. Analysis by the Tukey’s Studentized Range Test further shows that there is no significance for the young people, middle-aged people, or the older people.
in life satisfaction, as there are no group comparisons with a significance level at a 0.05 alpha-level. There is, however, a non-significant trend, with younger adults and middle-aged adults at approximately the same level of life satisfaction, and older adults non-significantly higher in life satisfaction.

DISCUSSION

By using these measures of SWB, the current study was able to tap into the components of SWB and find some pair-wise differences between age groups that emerged as significant. We found converging evidence that some components of SWB seem to be increasing with age (emotional response), and that another component of SWB (life satisfaction) does not appear to be increasing or declining, but rather staying at the same level.

These data support the idea that positive affect appears to be at relatively higher levels in middle adulthood when compared to younger adult cohorts. This finding is justified by cross sectional data. When we look at these findings in light of Isaacowitz & Smith's (1999) findings that positive affect was not as high in advanced old age, we might expect the decline in mean levels of positive affect in the older adult cohort. The mean level of positive affect in the older adult cohort was non-significantly lower than the middle-aged cohort. Nonetheless, it is still worthy to note this non significant age difference may be attributed to the growing health problems in the older adult cohort and the fact that this study did not separate participants into an old age group and an oldest-old age group. This cross-sectional data on positive affect was in accord with the hypothesis that there are significantly higher levels of positive affect as adults age.

Similar to what was found in terms of positive affect, we found that there was a significant pair-wise difference between the younger adults and the middle aged adult groups in levels of negative affect. These findings suggest that there is a significant decline in negative affect into middle age. My hypothesis also predicted that there would be lower levels of negative affect as people age. A further direction for this research might be to divide this old age cohort into two groups of older people and oldest-old people to test the claim of an upswing in negative affect in advanced old age (Isaacowitz & Smith 1999). These findings seem to be consistent with recent research that shows negative affect declining throughout middle age, specifically up until age 60 (Carstensen, Pasupathi, & Mayr, & Nesselroade, 2000).

Depressive symptoms seemed to be significantly lower in middle-aged cohorts as compared to younger cohorts. There is a non-significant upswing in depressive symptoms in the older adult cohort, which is at best a trend. In light of Nolen-Hoeksema's literature, our data are in accord with the tendency for depressive symptoms to be at the highest levels as young adults. The claim that cohort effect may explain this difference is applicable to our research here as well, since it is easy to see that parental divorce and parental depression may explain higher levels of negative affect in this cohort. The hypothesis was validated by these data since there was a decline in negative affect. The resur-
gence of the older adult cohort in negative affect was evidenced by only a non-significant trend.

The current study was not able to show a significant difference in life satisfaction between age groups. If one were only to look at the mean levels, it might appear that life satisfaction exhibits some highs into old age (see Table 1); however it bears repeating that a conclusion on such a trend is drawn from non-significant group or individual pair-wise comparisons. These findings on life satisfaction, and the absence of pair-wise differences between any age groups, suggests that life satisfaction is staying the same. This finding is converging evidence for the summary provided by Diener et al. (1999), which claimed that life satisfaction was at least the same, or higher in old age. The data seem to support parts of the hypothesis that SWB does increase with age. The data also suggest that the middle-aged cohort was typically high or stable in their levels of SWB.

Limitations

Where the data have not been able to explain declines or highs in old age, it may be useful to consider a range of confounding variables. The current study over-sampled females and under-sampled males, which is important in considering certain gender differences in emotional response. Obviously this study can only claim to make conclusions about SWB in terms of a cross-sectional analysis. We are not able to see if the young adult cohort approaches questions of negative/positive affect and life satisfaction with more honesty overall, thus having higher levels of negative affect and lower levels of positive affect. This age group may be significantly different from older cohorts and the young adult cohort may even show their consistency in affect and life satisfaction over time if a longitudinal study is to be conducted in the future. Cohort effect is always a possible confound when we are not able to see changes in SWB in longitudinal terms.

Another limitation of this study may be the sample. It may be difficult to generalize these findings to all socio-economic statuses when our study barely sampled any younger adults in lower middle class settings. Since the bulk of the younger participants were recruited from a particularly wealthy Ivy League institution, and a host of other universities and colleges with high SES students, it is hard to generalize the findings on SWB when this cohort may have been skewed in terms of their representative wealth for average peers in their age group. Since this does not represent the average American, and the sample might have come from a disproportionate amount of middle to upper class participants, this sub-sample might not have been representative of the average American young adult.

Another limitation that may have affected this young adult group is their familiarity with questionnaires in a university setting. It is likely that university students are accustomed to completing these surveys and are obviously more exposed to an academic environment, thus de-emphasizing their importance as younger adults are continuously completing them. Furthermore, middle aged adults and older adults might take greater time with the questionnaires due to their decreased familiarity and they might be less likely to report more emotional responses that are indicative of higher levels of negative affect. Younger adults may be more accustomed to these surveys and more willing to make themselves vulnerable and report their negative emotional responses. In addition, young people are often undecided in their career and life paths, and these opportunities to answer questionnaires might be potentially overwhelming and thus evocative of higher negative affect and lower levels of positive affect. Middle-aged adults and older adults may have been exposed to all of the important life domains, and thus the remainder of their life is much less daunting and less likely to be interpreted more negatively and less positively.

Future directions on the topic of age and SWB might include analyzing this data set longitudinally, by analyzing the young adult age group when they become middle aged adults, and the middle aged group when they become older adults. In addition, the current study might also better divide the age groups studied; specifically the older adult group. This age group combines individuals in "advanced old age," an age group that is prone to health conditions that are not normally correlated with high levels of SWB. Therefore, these data could be analyzed by creating a fourth age group that might show significant comparisons in upswings or declines that occur in advanced old age. Furthermore, the combination of these two diverse old aged groups is clumsy and may explain why the Tukey comparisons showed no significant pair-wise differences in the old age group.

Future research might consider both more diverse recruitment strategies and help to disentangle high levels of SWB in light of more favorable health conditions. Health seems to be a particularly crucial variable that explains, in part, differences in age and SWB. A further direction of this research might be to conduct this study, but exclude all those individuals suffering from health related problems. These ailments
may explain why older adults showed higher negative affect and higher depressive symptoms. While this type of recruitment seems to be a likely remedy, it would not be an ultimate solution given the limited amount of older adults who are completely healthy. In fact, such an exclusion criteria might result in a non-existent sampling of older people. Whenever dealing with health, it is difficult to control for health related problems; they are found at a higher prevalence in the older age groups. It is not easy to argue that health must be factored out of the equation, when decreased health and functioning is symptomatic of old age. Because deteriorating health and old age are related, it is debatable to call health declines confounding variables when they are an integral part of aging.

REFERENCES


INTRODUCTION

Why are so many more women than men depressed? Meta-analysis of studies conducted in various countries has shown that women are roughly twice as likely as men to experience depression (Nolen-Hoeksema, 1990). The reason for this sex difference is not entirely clear, although most researchers today believe that it is a combination of several factors, including: the effects of estrogen on the stress hormone, cortisol (Leibenluft, 2001), the prevalence of the victimization of women (Roesler & McKenzie, 1994), and the tendency of women to ruminate over their problems (Nolen-Hoeksema, 1990). These explanations point to differences between men and women biologically, environmentally, and psychologically. While all of these factors seem to play roles in accounting for the higher rates of depression in women, they do so in various ways, and arguably, to various degrees. Before we begin to evaluate these different explanations, however, we must first look at the evidence for sex discrepancies in rates of depression between men and women, and how these differences are manifested cross-culturally and throughout the life span. It is also necessary to consider the possibility that men and women actually share similar rates of depression, but express depression in gender specific ways. According to this theory, men mask their depression through externalizing acts, such as excessive drinking. If this is the case, then the difference in rates of depression between men and women is actually illusionary - men and women are equally likely to be depressed, and there is no need to account for sex differences in rates of depression.

WOMEN AND DEPRESSION

There have been numerous studies conducted within cultures and cross-culturally in order to identify depressive symptoms in both men and women. In 1994, Blazer et al. conducted the National Comorbidity Study assessing the rates of depression in each of the fifty states of the U.S. Of the randomised sample of people aged 15 to 54, 4.9% met the criteria for major depressive disorder. Of the women sampled, 6% were found to be clinically depressed compared to 3.8% of the men sampled. Gender differences in lifetime prevalence of depression were also found to be significant, with 21.3% of women and 12.7% of men having had experienced depression in their lifetimes (Blazer et al., 1994). The methods used in this study provide strong
evidence that women in America are roughly twice as likely to be diagnosed with major depression than men. Researchers such as Phillips & Segal (1969) have argued that women are more likely to be diagnosed with depression because they are more likely to seek professional help for their depressive symptoms. By taking a random sampling of Americans from each state, this bias is avoided. However, studies focusing on the sex differences of people seeking treatment for their depressive symptoms are also useful, because data are relatively easy to obtain.

In an analysis of studies of treated cases of depression, Nolen-Hoeksema (1990) compared the rates of depression for men and women from numerous countries and cities outside of the United States, including the United Kingdom, Egypt, Hong Kong, India, and Kenya. While each country and city varied in its rates of overall depression, women were consistently found to be approximately twice as likely to be treated for depression than were men. The only places where sex differences in treatment rates of depression were not reported were in Nigeria and in the rural parts of Iran. Nolen-Hoeksema (1990) also reviewed studies that identified subgroups of the American population that did not show gender differences in depression. These subgroups included children up until the age of puberty, college students, and the Old Order Amish. These exceptions, though few, are noteworthy and bring up several important questions. For example, what is particular to these populations that results in equal rates of depression between men and women? In the case of children, are the rates of depression similar for sociological or biological reasons? Do certain environments protect women from depression, or do they simply attract depression-resistant females? It is unclear whether the college environment promotes equal rates of depression in men and women, or if women who attend college are protected from depression in other ways. According to a study conducted by Lloyd and Miller (1997), medical students at the University of Texas share similar rates of depression, while medical students at the University of Edinburgh show higher rates of depression in females. Perhaps, then, the similar rates of depression that have been found in college students are limited to certain kinds of colleges and to certain countries. By identifying subgroups of populations that share similar rates of depression between men and women, we can learn more about what is particular to these groups in terms of demographics, environments, or psychologies that results in shared rates of depression. This information, in turn, may help us understand why there are sex differences in the greater population.

For now, it is important to note that despite these exceptions, women are overwhelmingly more likely to be depressed than men, at least in terms of meeting the diagnostic criteria for depression found in the DSM-IV. Whether these criteria are adequate measurements of depression is debatable, however, this leads some to question the idea of true sex differences in rates of depression. It has been argued that men are equally as likely as women to be depressed, but they are not as likely to express depression in terms of sadness, fatigue, loss of interest in activities, feelings of worthlessness, etc., since it is considered "unmanly" to do so (Real, 1997).

MEN AND DEPRESSION

According to Dr. Terrence Real, author of the best-selling novel, *I Don't Want to Talk About It: Overcoming the Secret Legacy of Male Depression*, men who experience depression carry two stigmas, one associated with mental illness and the other, “feminine” emotionality (Real, 1997). Among college students, Hammen and Peters (1978) found that when depressed female students reached out to their roommates and were met with concerned and nurturing reactions. Depressed male students who did the same, however, were met with social isolation and in some cases, outright hostility. Fourteen years later, Joiner et al. (1992) replicated the study and found similar results. Since men are taught that it is not acceptable to express their feelings of depression to others, they may seek comfort from other sources, such as alcohol. Indeed, one argument in support of the idea that the different rates of depression between men and women are illusionary is that alcoholism is twice as common in men than in women. If we think of alcoholism as the male version of depression, then we would not need to account for any differences in rates of depression between genders. There is, however, limited evidence in support of this theory.

For example, researchers have identified low serotonin levels in the brains of alcoholic men and depressed women, indicating similar biological components between the two disorders (Leibenluft, 2001). However, low serotonin levels have also been found in the brains of men and women with anxiety disorders (Leibenluft, 2001). It is unclear whether this is a result of comorbid depression in anxious patients, as well as
in alcoholic patients, since both anxiety disorders and alcoholism are highly correlated with depression, as well as with each other. However, even if low serotonin levels are found in alcoholic and anxious patients without depression, we cannot then assume that all disorders that involve low serotonin levels in patients in general are indicative of underlying depression. There seems to be very real differences between individuals diagnosed as clinically depressed, individuals with alcoholism, and individuals with anxiety disorders, such as obsessive-compulsive disorder.

It has also been argued that depression and alcoholism are genetically linked, since families with high rates of alcoholism in men also show high rates of depression in women. In 1974, Cadoret and Winoker found significantly higher rates of depression in the families of alcoholics, as well as significantly higher rates of alcoholism in the families with high rates of depression. However, recent studies, such as the Merikangas, Weissman and Pauls' study in 1985, have shown no connection between depression and alcoholism in families. Therefore, we should exercise caution in accepting the idea that vulnerabilities to depression and alcoholism are genetically linked. More importantly, even if they prove to be genetically linked in the future, we cannot then assume that they are two different manifestations of the same disorder. What would prevent us from grouping eating disorders, which are primarily diagnosed in females and have also been shown to run in families, under depression? To equate depression rates between men and women, we would be forced to look for other mental disorders that are predominantly diagnosed in men to "catch up" to the increased rate of depression in women.

Furthermore, depression is just as likely to be a consequence rather than a cause of alcoholism. Triffleman et al. (1995) found that in men alcohol disorders were more likely to precede depression than to follow depression. However, in women, this pattern was reversed. Also, while alcoholic men were twice as likely to develop depression than non-alcoholic men, alcoholic women were three times more likely to develop depression than non-alcoholic women. Perhaps then, it would be more accurate to say that alcoholism masks depression in some men and women, although women are more likely to have a primary diagnosis of depression, while men are more likely to have a primary diagnosis of alcoholism. Even among alcoholic individuals with a primary diagnosis of alcoholism, women are significantly more likely to develop depression in reaction to their alcoholism and its interpersonal and professional consequences.

In short, there does not seem to be any real evidence to support the idea that there is something particular to alcoholism that equates it to depression. These two disorders have distinct symptoms and can be viewed as different responses to similar difficult circumstances. Still, it is worth considering the possibility that some alcoholic men suffer from an underlying depression that triggered their alcoholism, and that the rates of depression in men and women are closer than they first appear. This still does not account for such a large, universal discrepancy in depression rates, however. The idea that this discrepancy is an illusion is not convincing and we must consider other explanations for this difference.

**BIOLOGICAL EXPLANATIONS**

Since the nineteenth century, doctors have been attributing mental disorders in women to their reproductive systems. In the middle of the 1800's, it was even thought that menstruation detracted blood from the brain, causing stupidity and temporary insanity. Women were told to stay in bed during their menstrual cycles, avoiding all physical activities possible (Nolen-Hoeksema, 1990). Even though medical research has advanced significantly since then, the role that women's reproductive systems play in their mental health is still unclear. While some studies support the idea that women's hormones are responsible for their higher vulnerability to depression, other studies have found no evidence for such a conclusion. The idea that biology plays an important role in explaining the different rates of depression between men and women seems to be a logical one, since different gender rates of depression have been proven cross-culturally. The fact that this gender difference does not emerge until puberty, when girls experience significant hormonal changes, also supports the idea that there is something particular to a woman's biology that makes her extremely susceptible to depression. While some researchers argue that a woman's hormones are responsible for her increased vulnerability to depression, others point to genetic factors.

Most of the research that has been done on the effects of hormones on women's moods has been done on estrogen and progesterone. Women experience periods of hormonal changes in these two hormone levels during the pre-menstrual period, pregnancy, the postpartum period, and during menopause. Evidence for an
increased risk of vulnerability to depression during these periods is weak at best. For example, Brook-Gunn and Warren (1987) assessed the moods of 103 girls, ages of 10-14, and how these moods correlated with girls’ various hormonal levels, the extent of their secondary sex characteristics development, and the type of life events each girl had experienced. They found that changes in hormone levels accounted for approximately 1% of the changes in girls’ mood levels. Life events, however, were strongly correlated with changes in moods. Specifically, girls who had experienced a negative life event were significantly more likely than girls who had not experienced a negative life event to experience depressed moods.

In another study by Eccles and colleagues (1988), the relationship between hormonal levels and moods varied. Some children experienced high levels of hormones, such as progesterone, androgen, estradiol, luteinizing hormones, and follicle-stimulating hormones with positive moods, while others experienced high levels of the same hormone with negative moods. This study suggests that increased levels of certain hormones may create increased sensitivity to moods in general, rather than increased vulnerability specifically to depression. Similarly, Alagna and Hamilton (1984) demonstrated that women who experience pre-menstrual syndrome show depressive symptoms throughout the month, although their moods are exacerbated before menstruation. It seems that changes in levels of estrogen and progesterone affect only those women with underlying tendencies to experience depression. Thus, the biological effects of hormones on women's moods and levels of depression cannot explain the large gender discrepancy in rates of depression between men and women. While female hormones may exacerbate depression in women who are already prone to depression, we cannot conclude that female hormones directly cause depression in women. Female hormones may, however, contribute to the higher rates of depression in women indirectly.

It has been demonstrated that close to half of all severely depressed people have high levels of cortisol, which the body releases in reaction to stress (Leibenluft, 2001). Vamvakopoulos et al. (2000) have found some evidence that estrogen increases cortisol secretion in women, as well as lengthening the amount of time it takes for high levels of cortisol levels to return to normal. This has important health implications in women, especially when we consider that women are more likely than men to experience stressful events in the form of low economic status and single parenthood. Marcott (1999) also found that upper-middle class adolescent girls were more likely than upper-middle class adolescent boys to report experiencing stressful events in their lives. Therefore, since women are said to experience a greater number of stressful events in their lives and also react to those stressful events with longer periods of cortisol secretion, they will then be more susceptible to the effects of cortisol in their blood. However, the relationship between cortisol and depression in unclear, since depression may cause cortisol levels to rise in the first place instead of the other way around. What is more, half of all depressed patients show normal levels of cortisol secretion and maintenance. Overall, hormonal explanations for the gender differences in depression are not convincing.

Other proponents of a biological theory for the different rates of depression between men and women argue that depression is passed down to women genetically. In a study done by Kendler et al. (2000), 2060 female twins with and without a family history of depression were assessed. When women without family histories of depression had undergone recent traumas, such as divorce or a death of a loved one, their likelihood of developing depression was increased by 6%. Those women with a history of family depression, however, were 14% more likely to develop depression. According to the study, women (and possibly, men) inherit the propensity to become depressed in the wake of a crisis. However, while there does seem to be a genetic link in depression, there is no evidence that this genetic link is passed down through the X chromosomes, which would make women more susceptible to depression than men. Merikangas et al. (1985) found that the relatives of male depressives and female depressives were equally likely to be diagnosed with depression. Therefore, there is no strong evidence that gender differences in depression can be explained through genetics, and we must consider other explanations.

ENVIRONMENTAL EXPLANATIONS

Environmental explanations for the higher rates of depression in women than in men are based on the assumption that women experience more stressors in their lives that lead to depression. Proponents of biological theories of sex differences in depression argue that the widespread nature of such differences point to a biological explanation, since women are biologically...
the same in Hawaii as they are in Hong Kong. Proponents of environmental explanations, however, point out that women are oppressed across the world as well. If women in some societies were to share equal roles and status to the men in their societies and still displayed twice the rate of depression as men, we would be able to say confidently that biology plays a major role in accounting for this sex difference. However, almost all societies have designated different, unequal roles for women than for men. The exceptional societies where women are considered to be completely equal or superior to men have not been studied in terms of depression rates, although such research might give us important further clues into this debate. What we do know, however, is that the environment plays a significant role in determining rates of depression in both men and women. This can be seen in the fact that depression rates vary from culture to culture, as well as between the subgroups within cultures. Research shows that rates of depression are higher in populations with low social status and income levels (Blazer et al., 1994). However, the relationship of social class, gender, and depression is not straightforward.

Brown and Harris (1978) identified four precipitating factors for depression in American women. These factors include the loss of a mother before the age of eleven years old, having three or more children under fourteen living at home, lack of employment, and lack of a close confidant or source of support. Interestingly, Brown & Harris found no class difference in depression rates among women without children. This is in spite of the finding that women in the working-class (68%) were more likely to have experienced a negative life event than women in the middle-class (38%). This suggests that particular stressors are more likely to increase the likelihood of depression in women, regardless of social class and environmental triggers. However, this study has not been replicated to determine whether men share the same precipitating factors. What this study does show us is that the environment plays a key role in determining the likelihood of developing depression in women.

Radloff (1975) found that married, divorced, and separated women were more likely to be depressed than men, while widowed men were more likely to be depressed than women. Radloff also found that unmarried men and women shared similar rates of depression. This may be due to the kind of woman who chooses to remain unmarried, although it is an interesting exception to the general trend of gender differences in depression. Radloff’s findings suggest that men and women react differently to romantic relationships, with more women experiencing depressive symptoms due to negative relationships. These findings were replicated in a survey of 900 couples, with women more likely to report depressive symptoms due to marital troubles than men (Gale Group, 2000). The study also evaluated the levels of hostility exhibited from each partner in the areas of cynicism, aggressive reaction to problems, and negative emotions about others. Interestingly, women’s levels of depression were closely related to their husbands’ hostility ratings, while men’s depression levels were not linked to their wives’ hostility ratings. It seems that one explanation for the sex differences in depression can be attributed to women’s greater sensitivity to negative relationships.

Furthermore, women are also more likely to be victims of physical and sexual abuse than men. In a national population-based sample, 16% of men and 27% of women reported having experienced childhood sexual abuse (Finkelhor et al., 1990). It may be that men are not as willing to admit to childhood sexual abuse than are women. However, it cannot be disputed that women are much more likely than men to be the victims of other forms of adult sexual abuse, such as rape. Studies have shown that sexual abuse, in particular, leads to the development of depression in both men and women, sometimes years later (Roesler & McKenzie, 1994). Other factors in the environment that may contribute to women’s higher rates of depression include: conflicting roles in the home and workplace, sexual discrimination in the home and workplace, and the burden of "the double shift," with women continuing with their roles as homemakers despite working full-time outside of the home. Overall, it has been found that women are more likely to experience stressors in their environments in terms of abuse and dual role expectations. Along with the finding that women are more psychologically vulnerable to the effects of negative relationships, it is clear that the environment plays a major role in explaining the sex differences in rates of depression between men and women. Since our different environments also shape our thinking, one question that we can now ask is how a woman’s environment shapes her psychology in ways that increase her risks of developing depression.

**PSYCHOLOGICAL EXPLANATIONS**

In 1975, Seligman and his colleagues devel-
oped the theory of learned helplessness. According to this theory, repeated exposure to negative situations where one has little control produces feelings of helplessness and symptoms of depression. Since women are more likely to experience situations where they are made to feel helpless, such as through sexual abuse and single parent situations, women would then be more likely to develop a sense that they are not in control of their environments, which may lead to the development of depression. There is some evidence in support of this theory that offers us a non-biological explanation for why sex differences in depression emerge in adolescence. Marcott (1999) found that male adolescents’ confidence in their problem-solving skills increased with age, while female adolescents’ confidence in their problem-solving skills decreased with age. Girls also reported experiencing a greater number of stressful events and a lower sense of feeling in control than did boys. According to learned helplessness theory, these feelings of powerlessness should result in greater rates of depression in adolescent girls than in adolescent boys, which is indeed the case. Learned helplessness theory does not explain, however, why some women are more likely than others to develop depression. Also, under learned helplessness theory, men who experience trauma would be just as likely to develop depression than women in similar situations, which does not seem to be the case.

Some researchers argue that different cognitive coping styles between the sexes offers us a better explanation for the different rates of depression in men and women. According to Nolen-Hoeksema (1990), men are more likely to react to emotional distress by trying not to think about it, while women are more likely to react to emotional distress by ruminating over their problems. In a two-part phone interview of 1,132 randomly selected California residents taken one year apart from each other, Nolen-Hoeksema (2000) found that respondents who scored highly on rumination levels at time 1 were more likely to meet criteria for major depressive disorder both at time 1 and time 2. More importantly, time 1 rumination scores significantly predicted level of time 2 depressive symptoms both before and after time 1 depressive symptoms were controlled for. Therefore, rumination was indicative of future development of depression.

The reason why women are more likely to engage in rumination can be attributed to the gender socialization process. Parents tend to engage in elaborative styles of speaking to their daughters and pragmatic styles of speaking to their sons when talking about the past (Fivush, 1993). Elaborative styles of speaking are rich in detail, connecting events to other events and the feelings that are involved with those events. Pragmatic styles of speaking, however, are non-detail-laden and not linked to other events. Davis (1999) found that non-depressed women were faster at accessing emotionally charged autobiographical memories than non-depressed men, typically recalling more memories in greater detail. According to this study, women who are faced with negative events should be faster than men at retrieving memories of other negative events, experiencing these memories and their accompanying negative feelings in greater detail and to greater degrees. Women would then be more likely to develop and to maintain feelings of depression. Presumably, women would also maintain feelings of happiness for longer periods of time than men, although this has not been demonstrated.

Overall, the relationship between rumination and depression is still unclear, since depression may precede rumination. Also, Gold and Wegner (1995) have shown that ruminative thought often follows a traumatic experience and the attempt to suppress the experience from memory. Since trauma is highly indicative of the future development of depression, high correlations between ruminative thought and the development of depression may be a consequence of the relationship between trauma and depression, rather than a relationship between rumination and depression. It could, however, also be the case that ruminative thought is a mediating factor between experiencing trauma and developing depression. Further research needs to be done in order to understand the role of rumination in the development and maintenance of depression. What we can conclude, however, is that women, for whatever reasons, are more likely to engage in ruminative thoughts, which are highly correlated with and have been shown to precede and maintain depression.

**CONCLUSION: AN INTEGRATED APPROACH**

In this essay, I divided the possible explanations for the sex difference in the rates of depression between men and women in terms of biology, environment, and psychology. I want to stress, however, that these categories are not independent from each other. For example, the higher rates of childhood sexual abuse in females (an environmental explanation) may affect...
brain development in ways that increase the risks of depression in the predominantly female victims of sexual abuse (a biological explanation). A brain in the early stages of development is highly susceptible to environmental insults, which can negatively affect the size of the structures of the brain, including the prefrontal cortex. Brain imaging has revealed that people who are depressed often have overactive right cortexes, which govern the physiological loop that produces positive emotions (Robbins, 2000). Since young girls are more likely to undergo sexual and physical abuse than young boys, Robbins (2000) and other researchers speculate that a greater number of girls may experience slight brain alterations due to trauma, which may then increase their risks of developing depression in the future. Further studies will need to be conducted in order to understand the relationship between childhood abuse, brain development, and mental illness. However, this example illustrates the fact that biology, the environment, and a person's psychology interact with each other in complicated ways. Because of this, there is no simple answer as to why women are roughly twice as likely to experience depression than men. Women and men differ in how these factors interact with each other. Let us consider, for a moment, a fictional case scenario:

Mary and Craig are married. Depression runs in Mary's family. Therefore, she has inherited a greater vulnerability to experience depression in reaction to negative life events, including a stressful marriage. Mary and Craig are spending a lot of time arguing with each other, since Mary feels overburdened with the duties of her job and the amount of household chores that she performs. Unlike Mary, Craig does not dwell on his arguments with his wife. Instead he spends more time at work or goes out and drinks with his friends, temporarily forgetting about his problems at home. Meanwhile, Mary sits at home thinking about how horrible she feels and what this means for her future. It is no wonder that Mary is twice as likely to develop depression than her husband!

Of course, this is a very simplistic scenario and there are many additional factors that will determine whether or not Mary or Craig will develop depression. It does, however, provide us with a more vivid picture of how a married couple may react differently to each other and their situations, and why this may lead to differences in depression levels between a man and a woman.

Finally, even though it has been demonstrated that depression rates in women are consistently about twice as high as depression rates in men cross-culturally, these rates are in no way fixed. They may converge over time, or grow farther apart. The differences in the amount and severity of stressors between men and women may become less extreme, or coping styles may become less gender-specific. If such were the case, the environmental and psychological theories which have been proposed and reviewed in this essay would predict closer rates of depression in men and women. Further research in neuroscience should also provide us with a clearer understanding of the physiology of depression and by what processes negative life events and rumination contribute to this physiology. This research may in turn lead to more effective treatments for depression.

In summary, it is most likely a combination of factors that account for the extent and the cross-cultural nature of the discrepancies in depression rates between genders. Females emerge with higher rates of depression than males during adolescence, as they experience different hormonal changes and greater feelings of helplessness, perhaps due to these changes. The greatest predictor of adolescent female depression is the experiencing of a negative event, which underlies the significant contribution of the environment to the development of depression. Women tend to experience more traumas in their lives than men. They also are more likely to ruminate over their traumas, which may increase their likelihood of developing depression. Conversely, it may be that men protect themselves from depression by engaging in avoidant behaviors. This does not imply that mental illnesses stemming from avoidant behaviors are other forms of depression. If alcoholism is the male version of depression, perhaps depression is the female version of alcoholism in a society where women are not as encouraged to drink alcohol. Simply stated, men and women experience different degrees and types of negative events in their lifetimes, and are taught to react to these events in gender-specific ways. This leads to greater susceptibilities to different mental illnesses, such as depression in women.

REFERENCES


Blazer, D. G., Kessler, R. C., McGonagle, K. A. & Swartz, M. S.


INTRODUCTION

Over the past few decades, the increasing amount of eating disorders has become a topic of concern. Many attribute this "phenomenon" to societal factors, such as the standards of beauty set by the media, fashion and entertainment industries. However, most people today are misinformed about eating disorders and underestimate the life-threatening dangers involved. Many people do not realize that the mortality rate of anorexia nervosa has been found to be between 5-20% (Nielsen, 2001). Funding for health care services, especially eating disorder programs, has decreased over the past few years. Thus, it has become even more important to develop efficient and cost-effective ways to inform, treat, and prevent eating disorders. The emergence of the Internet has become a hopeful medium for the implementation of these programs.

Eating Disorder Websites

The advent of the Internet has given individuals the ability to access all kinds of information from the comfort and privacy of their homes and offices. Not only does the Internet supply a limitless range of information, but it also introduces us to millions of other users from around the globe, allowing us to talk to anyone, anywhere, at any time. Clearly, with the Internet's fast-paced growth, it was only a matter of time before the mental health world would begin taking advantage of its resources. Amidst the wealth of information the Internet offers, thousands of websites have been devoted to the issue of eating disorders.

It has been difficult to make people fully aware of the dangers of eating disorders because of society's misconceptions and ignorance about them. Although the Internet can educate individuals about eating disorders, vast amounts of false information on this subject also exist, as is the case with most everything else on the internet. Sites that may seem to be educational can sometimes be damaging to uninformed individuals. In allowing for fast distribution of information, the Internet also allows for fast circulation of inaccurate information. Because the freedom of the Internet allows any individual to post a website, there are also websites that advocate eating disorders. (Christensen and Griffiths, 2000). While most eating disorder sites have been developed to help individuals learn the truth behind eating disorders, there exist "Pro-Anorexia Websites" which blatantly and shamelessly advocate anorexia (Shafran, 2002). "Inspirational quotes" are posted to keep girls on the path of anorexia to the supposed "ideal body." It is unknown how many sites exist that promote eating disorders, but it is obvious that there is a need to evaluate Internet-based health information with what is known to be scientifically accepted. Nevertheless, although ensuring the accuracy of information on the Internet is a recurring problem, indi-
Online Eating Disorder Support Groups

Along with many of the misconceptions about eating disorders, a great deal of stigmatism also exists regarding eating disorders. Having illnesses that are perceived as embarrassing or socially disgracing, such as eating disorders often are, lead people to seek support of others who share their condition. Thus, it is no wonder that eating disorder support groups have evolved over the Internet. Online support groups can be accessed through all sorts of services such as AOL and USENET (Finfgeld, 2000). In chat rooms people can come together to talk and share stories about their common disorder. They often find support from those who can sympathize with their experiences, and they can pass on information about the best treatments and how they and their families can best cope with it. Today, several eating disorder newsgroups and chat rooms exist for individuals to share their personal experiences and learn information on how best to deal with the disorder.

Advantages and Disadvantages of Online Eating Disorder Support Groups

The Internet is accessible at any time of the day, all year round. Messages can be posted and e-mails can be sent at any time, by anyone in the world, from the privacy of one's own home. It is relatively cost efficient and easy to use. There is no set time that people must write e-mails or enter chat rooms, and individuals do not have to work their schedule around online group sessions as they do for face-to-face sessions. In one study on an eating disorder newsgroup, Winzelberg (1997) found that two thirds of the messages were posted between 6pm and 7am, a time when support from healthcare professionals or face-to-face support groups would have been least accessible.

Many hospitals and community clinics offer support groups, yet many individuals turn to Internet support groups. Internet support groups offer anonymity; and in cases like eating disorders, being able to seek support for a disease while still being able to remain anonymous is highly desirable. For some, remaining anonymous is a way of reaching out for help while not having to admit to those close to them that they have a problem. Girls can share their similar experiences and can often help one another, regardless of whom they are talking to.

Anonymity also allows for elimination of socio-demographic factors like age, race, and socioeconomic status (Finfgeld, 2000). As a result, differences in social status, which are normally more visible in face-to-face sessions, can be minimized, and issues of physical attractiveness and social skills are neutralized (Davison, 2000). Anonymity reduces the shame and embarrassment that people sometimes associate with their eating disorders.

Though many make use of the online support groups available, much controversy has surrounded their efficacy and validity. There have been claims that online chat groups can have harmful or even dangerous effects on individuals (Christenson and Griffiths, 2000). Accuracy of the information shared is a big concern because it is unknown how many individuals rely on the information they learn in these support groups.

Newsgroups and chat rooms without moderators fall under greater scrutiny because of the increase in risks they project. Wanderers can enter newsgroups, sometimes making negative comments and causing more damage on group participants than would otherwise have been there. False information is not always corrected, and while this is not exclusive to newsgroups about eating disorders, inaccurate information is more dangerous in these groups because of the lack of knowledge most individuals have about eating disorders. Thus, inaccurate information about eating disorders is more likely to be accepted as truth.

Allowing people to join whenever they choose also lets people leave whenever they choose. Even with moderator-led newsgroups, the moderator cannot control the regularity or length of a member's participation. They also cannot control what goes on in the privacy of each person's home. Distractions, such as television, radio and other people, which are not present in face-to-face sessions, are out of the control of the moderators (Zabinski, 2000).

Internet-based Eating Disorder Therapies

The difficulty in treating eating disorders and the amount of relapse (for review, see Herzog, Nussbaum, & Marmor, 1996) has led physicians to start using e-mail as an adjunct to therapy. Its uses can range from the simple task of scheduling appointments to more in depth correspondence in which weekly email reports are required of the patient.

A case study done by Yager (2000) illustrates the correspondence between him and four of his patients. After each treatment, patients submitted eval-
sations of the mandatory use of e-mail to supplement treatment. All the patients found it helpful, most importantly because it met a "demand feeding" schedule, meaning they were able to write whenever they felt the need to "speak" to the doctor. They could write the mandatory e-mails as their schedules permitted, and having to do them almost daily kept them in constant awareness of their behaviors. The constant contact via e-mails increased the frequency and amount of time with their clinician, forging a stronger relationship with the doctor and making it more comfortable for them to say whatever they wanted either in the e-mails or in the therapy sessions.

Though the simplicity of e-mail as an adjunct to therapy limits the dangers of its usage, some risks can be noted. On the part of the physician, failure to respond to patient's e-mails in a timely or sufficient manner can hurt the doctor-patient relationship, prompting the patient to lose trust in the doctor and causing unnecessary damage to the relationship and negatively affecting the therapy. The use of e-mail also increases the risk of violation of privacy. One of the hallmarks of therapy is the doctor-patient confidentiality, but as discussed earlier, improper logging out of e-mail accounts enables others to read physician's e-mail responses, thus breaching one's privacy.

A more risky venture has been for physicians to use e-mail or chat rooms as the only means of therapy and communication. One might ask, why offer eating disorder therapy over the Internet? Eating disorders generate a significant degree of shame and denial, and often, because of this, many sufferers do not seek treatment (Muscari, 1998). What many do not know is that for some girls, this disorder can be effectively treated using self-help treatments. Cognitive Behavioral Therapy is a sound candidate for online treatment because its effectiveness has been proven in face-to-face sessions as well as self-help formats (Zabinski et al., 2000).

One study done by Robinson and Serfaty (2001) studied the effectiveness of therapy delivered entirely by e-mail to 23 female students from the University of London. These students, recruited online by an e-mail sent only one time to the student and faculty members of the University of London, exhibited Bulimia Nervosa, Binge Eating Disorder or Eating Disorder Not Otherwise Specified, initially measured by BITE (Bulimia Investigatory Test Edinburgh) and BDI (Beck Depression Inventory). Half of the subjects were treated with CBT, Cognitive Behavioral Therapy, and the other half were treated with a more traditional treatment, using diet records, advising on regular eating habits and confronting of cognitive determinants of the eating disorder. At the 3 month follow up, subjects from both groups returned questionnaires which showed there was an overall significant reduction in scores of depression and of bulimic symptoms and severity.

For some girls, there are advantages to using e-mail as the only means of communication. Using e-mail eliminates much of the status disparity between therapist and patient. Along with this, some patients may feel more comfortable with their therapist and may interact less formally than they otherwise might feel comfortable doing in an office, face-to-face. For some, avoidance of face-to-face confrontation makes it easier for them to be honest about their discussions. Knowing that the physician will not be focusing on their physical appearances is comforting and allows them to feel less pressured. In this experiment, anonymity was maintained, to a great extent. This is very important for sufferers since many do not seek treatment because of shame and guilt. However, it should be noted that in this study, physicians were given their college e-mail addresses so that, in case of an emergency, they would be able to obtain proper medical care for the patient.

As would logically be the case, there are risks to this form of treatment. Because of the difficulty in expressing non-verbal cues across e-mail, failure to perceive urgent communications from troubled patients is a very significant risk. In the case of Robinson and Serfaty's study, attaining valid college e-mail addresses from which proper information of the subjects could be attained, controlled for the risks of complete anonymity. Even in this case, physicians were taking responsibility for patients they had never met before.

However, there are online counseling services such as the Samaritans (www.thesamratians.co.uk) in which complete anonymity is ensured. This becomes a great risk for patients and physicians. Troubled patients who may be a threat to themselves or others cannot be reached because of their anonymity. This can endanger the patient and can compromise the physician who is liable for the patient.

Another method of Internet-delivered therapy is treatment provided through newsgroups and chat rooms. One study done by Zabinski et al. (2000) studied the potential of IRC-delivered (Internet Relay Chat) treatment. IRC, a form of synchronous chat, is one instrument for people to login to chat rooms. Zabinski and colleagues provided a 7-week program based on a cognitive-behavioral approach to 4 female students.
who had high body image dissatisfaction as measured by the BSQ, Body Shape Questionnaire. Students met once a week in a limited access chat room for structured activities; homework was assigned weekly, and responses to the readings were posted in a newsgroup. Questions were e-mailed to the group moderator.

Significant improvements were observed on the EDI-DT (Drive for Thinness Subscale), EDE-Q (Eating Disorders Examination-Questionnaire), BSQ, and RSE (Rosenberg Self-Esteem Scale). Subjects reported high satisfaction with the program, and exhibited positive change in their attitudes about their shapes and weights. They also stated that the intervention helped them to recognize and challenge the thoughts that triggered their negative feelings.

The results of this study give support for moderator-mediated chat room therapy in the treatment of eating disorders. The accessibility of the Internet for so many sufferers makes this treatment a very valuable approach. The IRC-delivered treatment proved to be effective and more helpful for those who otherwise would not have sought treatment.

However, this form of treatment assumes the most risks. Clients must be forewarned of the potential breaches of privacy, such as the interception of Internet transmissions or improper logging out of chat rooms. Practitioners assume great risks in liability in delivering this form of therapy. Unless in complete certainty, practitioners cannot assume that their insurance carrier will cover the "telemetry-mediated services." As with e-mail delivered therapy, the clinician is accepting responsibility for clients he has not actually met, and any mishaps resulting in injury to the patient could leave the therapist vulnerable to civil action. Backup services in case of an emergency can be difficult for the physician to arrange if the patient maintains complete anonymity.

Implications of this Research

In response to the diminishing health care resources, online support groups and therapies have grown rapidly, and yet research in this area remains in its infancy. At first glance at the principal findings, one might be very confident in implementing these cost-efficient treatments. However, the findings of the aforementioned studies, though promising, are not enough to solve the problem of eating disorders. Nevertheless, what they have done is validate further large-scale research to establish the efficacy of these new treatments.

Most importantly, there are three main weaknesses which need to be addressed in future studies. First off, much of the research thus far has been case studies. Randomized-controlled trials are almost nonexistent. Given this substantial lack, there has been insufficient data to produce valid conclusions. Secondly, though Internet-based therapies have the potential to reach a vast number of people, the studies so far have only investigated small sample sizes. Unfortunately, what may work on four students may not necessarily work for the population at large. Lastly, none of the studies up to now have assessed the long-term effects of these therapies. The longest follow-up so far has been 10-weeks, in the study done by Zabinski, Wilfley and colleagues (2000). It is a possibility that the positive effects of these treatments may be gone within a year. Given that these follow-up questionnaires are sent via e-mail, this simplified task of administering follow-ups should be very agreeable with physicians. Therefore, it is imperative that future studies incorporate larger sample sizes, randomized-control groups and long-term follow-ups. Until these three limitations are resolved, Internet-based treatment cannot be established as effective therapy in the treatment of eating disorders.

Qualitative Research on Internet Communities

Amid the wealth of information on the Internet, one of the most useful sources of information for physicians comes from support groups created in chat rooms and newsgroups. Qualitative analysis of Internet postings may help to determine peoples' needs and preferences, thereby helping clinicians identify the optimal blend of patient and professional needs to maximize all types of treatment success (Eysenbach, 2001)

There are two kinds of research methods: active analysis, in which researchers actually participate in the group, and passive analysis. In 1997, Winzelberg completed a passive observation study to analyze the content of messages posted in an electronic eating disorder support group. By means of an unobtrusive 3-month observation, Winzelberg collected 306 posted messages. He categorized the postings into seven categories, with "personal disclosure" and "provision of information and emotional support" being the most frequently posted messages. According to Winzelberg, members of the group used helping strategies similar to those found in face-to-face support groups.

While the findings of Winzelberg's study show that online support groups are as effective as face-to-face support groups, his study has many flaws. The length of observation was relatively short, and because...
none of the patients were contacted, it is unknown whether this specific eating disorder support group was typical of all eating disorder support groups. In addition, there are many ethical limitations to this kind of research. In particular, privacy of group members needs to be addressed.

In conditions of passive analysis, a big dispute is whether consent is needed for unobtrusive observation and analysis of posted messages. In the case of this study, none of the members of the support group were aware that their messages were being copied, and in some cases, being reproduced in Winzelberg's article. There is a huge risk in doing this because some search engines, such as Google.com, can index newsgroups such that using the direct quote as a query will release the e-mail address of the sender (Eysenbach, 2001). However, Winzelberg did in fact make some changes to the messages to insure the anonymity of the group members, but it must be noted that this action is the very least that must be done to ensure every members privacy.

In order for researchers to conduct active analysis studies, informed consent must be obtained from each member of the group. This can be done either prospectively or retrospectively. Prospectively, the chat room or newsgroup must be monitored such that upon entry of new members, each can be briefed on the ongoing research. Unfortunately, use of this method may influence the content and manner of discussion because the knowledge of being monitored may affect the members. However, obtaining consent retrospectively can be a hassle since it is unknown how many members may ask to be withdrawn from the analysis.

For future studies, another focus in this area should be to elucidate the differences in communication and disclosure styles between groups with moderators and groups without moderators (Davison, 2000). While it seems that having moderator-led groups is more beneficial in ensuring the accuracy and privacy of discussions, it is unknown how these moderators affect members. Another focus needs to be on the accuracy of postings and discussions, and how best to avoid the spread of false information. In his study, Winzelberg found that almost 12% of the shared information was inaccurate and beyond the accepted standards of mental health care. If this is the case with all chat rooms and newsgroups, it is most imperative that this situation be managed as quickly as possible.

Prevention Programs

Less than 1% of the young adult female population suffers from anorexia while about 1-2% of the population suffers from bulimia nervosa. More than 10% of college women are thought to have sub-clinical bulimia nervosa, and 25% of college women are thought to be at risk for developing eating disorders (Winzelberg, 2000). For this reason, it is clear why finding an effective prevention program is imperative. Unfortunately, prevention programs for eating disorders have not had very satisfactory outcomes. Their effectiveness has been questioned, with some studies showing that it can in fact be more harmful than beneficial for individuals. Furthermore, funding for prevention programs has been scarce because, even as eating disorders is a widespread mental health problem, its prevalence is not comparable to such disorders as depression or drug abuse.

One Internet-based prevention program, however, has been shown to be effective and cost-efficient. Participants in an Internet-delivered computer assisted health education program (CAHE) called Student Bodies have exhibited significant decreases in body image dissatisfaction. The Student Bodies program, modeled from self-help treatments for eating disorders, incorporates a moderator-mediated electronic newsgroup used as a forum by participants to discuss readings and assignments. It was designed to decrease body image dissatisfaction, a probable risk factor for the development of eating disorders. Developed by eating disorder researchers at Stanford University, the program was first run as four separate trials between 1995 through 1998. Each successive trial, using a modified version of the program, showed improved results in the reduction of body image concerns and eating disorder psychopathology (Dev et al., 1999). After these first trials, researchers conducted more studies to provide more proof of the efficacy of the program.

Zabinski and colleagues (2000) did a controlled study to assess whether an 8-week intervention of the Student Bodies program would significantly decrease body image dissatisfaction and disordered eating in college women. The 8-week program consisted of weekly assignments and mandatory postings to the discussion group in response to the readings. The control group was put on a wait-list and given the opportunity to participate in the Student Bodies program after the follow-up period. Results of the 10-week follow-up assessment completed by 56 students showed that the intervention significantly decreased measures of body image dissatisfaction and eating disorder pathology, with students reporting that they felt most supported by
the electronic newsgroup component of the program. Winzelberg and colleagues (2000) also did a study evaluating the Student Bodies program. Using the same structure as the study done by Zabinski et al. (2001), Winzelberg and colleagues conducted an 8-week intervention on 60 college women. Questionnaires completed at the 3-month follow-up showed that participants in this study also exhibited significant changes in body image satisfaction as determined by the BSQ, EDI-DT and EDI - Bulimia scale.

Winzelberg and colleagues (2000) conducted another controlled study this time comparing the Student Bodies program with a classroom-delivered intervention program called Body Traps. The Body Traps program consisted of eight two-hour class meetings, weekly assigned readings and written reflections in response to the readings. A significant alteration was made to the Student Bodies program such that, in addition to the 8-week program, it included three two-hour face-to-face sessions with the same moderator who led discussions on the on-line discussion group. Measures of the EDE-Q Eating Concerns subscale, EDE-Q Restraint subscale and the EDE-DT by fifty-eight students at the 4-month follow-up assessment revealed significant differences between the Student Bodies condition and the wait-list and Body Traps conditions, with participants in the Student Bodies program showing the most progress. Significant improvements in body image satisfaction and disordered eating attitudes and behaviors were demonstrated by the Student Bodies program while no significant effects were produced by the Body Traps in comparison with the wait-list condition.

The data to support the efficacy of this program have been overwhelming. Modifications made after each trial, in accordance to the feedback of the students, proved to be even more effective than each revision before it. Nevertheless, trials of the program need to be implemented in more areas, particularly where the girls are at high risk. Eighty-five percent of anorexics are diagnosed between the ages of 13 to 20 years old with peaks occurring in the years when high school and college begin. For this reason, the Student Bodies program should also be introduced to high school students, especially those in boarding schools who are at even higher risk. The method should follow the studies already done by Zabinski et al. (2000) and Winzelberg (2000). Finally, it is imperative to determine the long-term effects of this program. By doing a prospective study with long-term follow-ups, researchers will be able to determine for whom the program works best and if those improvements are sustained.

Suggestions for Future Research

A randomized-controlled-trial of the Student Bodies program should be conducted on the students of a different university and another trial should be conducted on high school students. For each trial, a sample of about 400 female students should be randomly assigned to the Student Bodies program or a wait-list control. Much like the earlier trials, an 8-week intervention should be administered with the following modifications. Measures should be taken at baseline, post-treatment, 4-month follow-up and 12-month follow-up. Online questionnaires should include the BSQ, EDI Bulimia, EDI Drive for thinness, and EDE-Q: Global, Weight, Shape, Eating and Restraint. Analysis of the data should be done using analyses of variance (ANOVA).

It is obvious that the technologies of the Internet offer a very useful tool for mental health care providers to reach a wider range of girls suffering from or at risk for eating disorders. Perhaps the most important development is the success of the Student Bodies prevention program. Determining the factors that are most effective in the program will allow physicians to better understand the pathology of the disorder and perhaps aid them in modifying the current treatments. Hopefully, if this program can be safely provided to girls all around the country, the epidemic of eating disorders might come to a halt, or at the very least, exhibit a decrease in the number of incidences.

REFERENCES


Factors Influencing Women’s Attitudes Towards Abortion at the University of Pennsylvania

Danesh Modi

In 1982, nearly one in five American women had an abortion. The prevalence of such surgeries has steadily increased, and those women having the operation performed find themselves, along with many others, in a highly polarized debate concerning the ethics of abortion. Through collected self-report data, this study examines the factors influencing abortion attitudes of 187 female undergraduate psychology students at the University of Pennsylvania, and intends to explain the underlying forces driving abortion’s perceived morality. Results support four of the apparent causal hypotheses, leading one to believe that sexual mating, career, and birth control strategies are related to abortion attitudes. Results further suggest that proponents of either side share certain views of their opponents, yet there exist various and strictly defined boundaries between the two main pro-life and pro-choice ideologies.

INTRODUCTION

In the last three decades, few issues have agitated Americans as much as abortion. It is currently a heated, controversial debate that pervades regular citizens and lawmakers alike. While the two main pro-choice and pro-life sides have traditionally been thought to maintain diametrically opposed viewpoints, current research suggests that this may not be entirely true. Nevertheless, one fact is certain: there are strong correlational relationships among abortion and various other measures relating to sexual mating, reproductive strategies, and birth control, among others. Religiosity and educational level, for example, have traditionally been some of the most predictive factors in explaining women's abortion attitudes. Henshaw and Kost (1996), from the results of a 1994-1995 national survey of actual abortion patients, clearly showed that "women [...] who have no religious identification are 3.5-4.0 times as likely as women in the general population to have an abortion" (140). The study also examined 1988 National Center for Health Statistics (NCHS) data to show that women who "completed their education with a high school diploma have a much higher abortion rate than women of any other educational level" (144). The article further emphasizes that the number of abortion patients who have never used any contraceptive method, including the pill, is highest among women younger than 18 years of age, and more importantly, that nearly 90% of those patients aged 20 years or older have used a certain form of birth control. These results suggest that women who are older, more religious, and better educated are all positively inclined to resort to abortion in order to avoid carrying a child to term.

Although abortion is certainly very controversial, it is not as highly polarized as most people believe. In recent years, current research has emerged that refutes some of the most traditional views concerning abortion, and proposes other possible theories for what drives people to have the abortion attitudes they have. Cook et al. (1992) cite a study of pro-choice and pro-life activists by Kristin Luker, affirming that a "[...] surprising number of pro-life activists in California were converts to Catholicism" (133) even though Granberg (1982) maintains that, in general, pro-life supporters are opposed to contraception use. More general attitudes towards abortion have been found by Harris et al. (1985) in which the authors cite the adult population's viewpoint that 84% of adults believe that the number of teenage pregnancies in the United States is a serious problem and that open communication about sex education and birth control are two key things that parents should do to alleviate the problem of teenage pregnancy. Furthermore, as Goggin et al.
young women have various technologically advanced sexual morality attitudes in the first place. These technologies, which are the factors causing their abortion and integral part of their sexual mating and career strategies. Their attitudes towards abortion form an abortion legality is related to UPenn women's sexual abortion (see methodology).

factor (e.g. contraceptive use) and attitudes towards assess correlational significance between one particular composite abortion scale, [abort], which will be used to related variables from the same data set, I create a composite variables.) Additionally, by combining causally that dispose UPenn women to seek an abortion. (Table select group to better understand the underlying factors this rather large sample of scaled variables, I examine a coded into various, strictly defined variables. From population of undergraduate psychology students at the University of Pennsylvania - to be more or less inclined to resort to abortion. This study's experimental data draws upon a larger set of results obtained from questionnaires administered to UPenn psychology students. The responses to those questionnaires were scored and coded into various, strictly defined variables. From this rather large sample of scaled variables, I examine a select group to better understand the underlying factors that dispose UPenn women to seek an abortion. (Table 1 provides a brief description for each of these predictive variables.) Additionally, by combining causally related variables from the same data set, I create a composite abortion scale, [abort], which will be used to assess correlational significance between one particular factor (e.g. contraceptive use) and attitudes towards abortion (see methodology).

The underlying hypothesis of this study is that abortion legality is related to UPenn women's sexual strategies. Their attitudes towards abortion form an integral part of their sexual mating and career strategies, which are the factors causing their abortion and sexual morality attitudes in the first place. These young women have various technologically advanced forms of birth control available at their disposition, which allows them to be sexually active while at the same time delaying pregnancy. The hypothesis also states that those women in need of abortions - those that are sexually active and do not want a pregnancy - are pro-choice because they are protecting their main sexual strategy choices, such as delaying in order to achieve stronger overall fitness gains in the future, such as receiving a higher-paying and higher-respected job due to additional years of advanced post-graduate education. On the other hand, it is believed that non-sexually active women are pro-life, for the near exact opposite reason: because premarital sex is a strategy these women do not support and one in which sexually active women exploit for access to potential mates.

Consequently, the variables that I expect to be strongly correlated with the abortion scale are: [pill], [agekids], [hasfr], [spcareer], [spatt], [nsex23], [pbschool], [religious], [incedu] (see Table 1). In accordance with the hypotheses, it is believed that [pill] will be a strong correlate of [abort] because it is a good predictor of a woman's sexual activity and her desire not to have a child at the present moment. Likewise, [agekids] should serve as a good explanatory variable because it directly measures the extent to which women are delaying pregnancy. Quite simply, the more a woman delays the birth of her first child, the more pro-choice she is expected to be. I also expect [hasfr] to matter because women with boyfriends should, on the average, be more sexually active. All three of these variables are expected to give strong positive correlations, indicating strong pro-choice viewpoints. Additionally, both [spcareer] and [spatt] speak to a woman's desire to secure an "outstanding" (in terms of overall reproductive fitness) long-term mate, and they should, therefore, be highly correlated with abortion attitudes. The hypothesis in this case is that there is no real reason to delay pregnancy if an outstanding long-term mate has already been secured. These two variables should also give positive correlations with the abortion scale.

The variable [nsex23] is an already created composite scale within the database that represents the total number of intercourse partners expected by a respondent by the time she reaches 23 years of age. It is an item that combines actual sexual history with expected future partners while controlling for age - which is very important for a sexual intercourse item simply because greater experience (greater age) will usually lead to a greater value for this variable. This variable is expected to be positively correlated with
pro-choice abortion attitudes for it is a direct measure of sexual activity and thus of a woman's risk of needing an abortion. The last three explanatory variables are also composites, which I created from existing variables in the dataset. The variable [pbschool] asks whether sex education and birth control should be available in school, and consequently those strongly in favor of such a proposal should generally be pro-choice. The variable [religious] examines the overall religiosity of a respondent, and because it is negatively coded in the database, a strong positive association with [abort] is expected. That is, the less religious a respondent purports to be, the greater likelihood of that person being a pro-choice supporter. Lastly, [incedu] was created to examine future beliefs and goals about income and education. While there is much research showing that current socioeconomic conditions play an integral role in whether a woman receives an abortion or not, this variable examines whether future predictions of the same socioeconomic conditions will have the same effect. This variable should play an explanatory role because it is part of the reason why these women want to delay pregnancy - to build up their future careers in order to secure that outstanding long-term mate. To this extent, I also expect a strong positive correlation between greater income and higher education level and pro-choice beliefs.

**METHODOLOGY**

**Participants**

In this study, a particular subset of a data file comprised of 324 University of Pennsylvania undergraduate psychology students were analyzed. Participants in the study were females (n = 187), of ages 18 to 21, recruited from psychology courses during the fall of 1999 through the summer of 2001.

**Materials**

Data acquisition of this particular subset involved a direct self-report process in which respondents were asked to complete a questionnaire containing various items. The items were strictly defined and clearly described to all participants, and were designed to ask questions and obtain representative information about each female respondent, their biological parents and siblings, their boyfriend (if applicable), and about certain aspects of their physique. The overall data come from the gathering of responses during numerous administrations of the surveys. Due to the fact that certain items were not asked in each session, some items do not have data for all the female respondents. In these cases, the pairwise correlations and multiple regression models could not be accurately calculated or analyzed, and thus, these women have been excluded from these particular tests, as indicated by Table 2.

**Procedure**

The surveys were administered to University of Pennsylvania undergraduate psychology students participating for course credit. The participants were interviewed to determine what relatives they have, and then, they filled out questionnaires concerning their political and moral attitudes, themselves, and their close relatives and romantic partners. Of these questionnaires, only those from the female respondents were examined and analyzed.

In order to decrease the effects of co-linearity and those of general statistical noise among the data, a new scale was created under the variable name [abort], as previously mentioned. This is a composite variable composed of the following variables: [abhealth], [abdefect], [abrape], [abnomore], [ablowinc], [abrelat], [abschool], [abany], [flpcact], [flplact], [abmoral], [abself], and [abfriend]. These 13 variables are among the most representative, in this data set, of UPenn students' attitudes towards abortion, and thus comprise the very essence of what this study tries to explain. All variables but [flplact], [abmoral], [abself], and [abfriend] were positively correlated with abortion - implying a pro-choice stance - and were thus positively coded into the formula defining [abort]. The other four variables mentioned, however, were negatively correlated, and were therefore adjusted through a negatively coded submission into the [abort] variable formula.

In accordance with the previously mentioned hypotheses, the following variables were examined in order to understand the underlying causality and basic beliefs of UPenn women concerning abortion: [pill], [agekids], [hasfr], [spcareer], [spatt], [nsex23], and three composite variables: [pbschool], comprised of [sexed] and [birthcon]; [religious], comprised of [relcons], [relstron], [church], and [pray]; and [incedu], comprised of [edu] and [inc]. In the first explanatory composite variable, both constituent variables are negatively coded so that the smaller the resulting datum (i.e. the more negative it is), the greater the correlation with a pro-choice stance. In the second, for all four religion-related variables, the greater the resulting value for each respondent, the more she is expected to
believe in the pro-choice side of the debate. Therefore, all four constituent variables are positively coded into the [religious] variable formula. In the last composite variable, both component variables are expected to give high degrees of positive correlation with [abort] and are consequently positively coded into the variable formula.

First, pairwise correlations were calculated between the dependent variable [abort], and all the hypothesized predictive variables, in order to look at overall relationships and strength of causality between UPenn women's views on abortion and each of the explanatory factors. Then, these variables were analyzed through multivariable regression models followed by forward stepwise regressions until the best predictive model was determined.

RESULTS

Table 2 shows the pairwise correlations for females among [abort] and all the response variables that attempt to explain the effects of abortion beliefs. These correlations are calculated after eliminating outliers.

<table>
<thead>
<tr>
<th>Table 1. Predictive Variables: Explanations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PILL</strong></td>
</tr>
<tr>
<td><strong>SPCAREER</strong></td>
</tr>
<tr>
<td><strong>SPATT</strong></td>
</tr>
<tr>
<td><strong>PBSCHOOL</strong></td>
</tr>
<tr>
<td><strong>RELIGIOUS</strong></td>
</tr>
<tr>
<td><strong>INCEDU</strong></td>
</tr>
<tr>
<td><strong>NSEX23</strong></td>
</tr>
<tr>
<td><strong>AGEKIDS</strong></td>
</tr>
<tr>
<td><strong>HASFR</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2. Pairwise Correlations: Females</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>pill</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>agekids</strong></td>
<td>r = .008 p = .92 n = 140</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>hasfr</strong></td>
<td>r = -.32 p = .0001 n = 141</td>
<td>r = -.15 p = .04 n = 184</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>spcareer</strong></td>
<td>r = -.15 p = .08 n = 141</td>
<td>r = -.01 p = .85 n = 165</td>
<td>r = .10 p = .21 n = 166</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>spatt</strong></td>
<td>r = -.14 p = .09 n = 141</td>
<td>r = -.07 p = .38 n = 165</td>
<td>r = .10 p = .19 n = 166</td>
<td>r = .40 p = .0000 n = 166</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>nsex23</strong></td>
<td>r = -.33 p = .0001 n = 141</td>
<td>r = -.14 p = .11 n = 140</td>
<td>r = .16 p = .07 n = 141</td>
<td>r = .10 p = .22 n = 141</td>
<td>r = .27 p = .001 n = 141</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>pbschool</strong></td>
<td>r = -.08 p = .34 n = 141</td>
<td>r = -.09 p = .21 n = 183</td>
<td>r = .12 p = .34 n = 186</td>
<td>r = -.03 p = .74 n = 186</td>
<td>r = .06 p = .45 n = 165</td>
<td>r = .32 p = .0001 n = 141</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>religious</strong></td>
<td>r = -.24 p = .005 n = 138</td>
<td>r = -.24 p = .32 n = 183</td>
<td>r = .08 p = .34 n = 186</td>
<td>r = -.07 p = .74 n = 165</td>
<td>r = -.05 p = .49 n = 166</td>
<td>r = .22 p = .01 n = 165</td>
<td>r = .21 p = .005 n = 141</td>
<td></td>
</tr>
<tr>
<td><strong>incedu</strong></td>
<td>r = -.07 p = .39 n = 141</td>
<td>r = -.09 p = .23 n = 184</td>
<td>r = .02 p = .79 n = 186</td>
<td>r = .37 p = .0000 n = 186</td>
<td>r = .22 p = .01 n = 166</td>
<td>r = .08 p = .10 n = 141</td>
<td>r = .12 p = .99 n = 183</td>
<td>r = -.00 p = .182 n = 141</td>
</tr>
<tr>
<td><strong>abort</strong></td>
<td>r = -.24 p = .005 n = 139</td>
<td>r = -.08 p = .30 n = 183</td>
<td>r = .17 p = .02 n = 183</td>
<td>r = .10 p = .20 n = 183</td>
<td>r = -.00 p = .99 n = 183</td>
<td>r = .38 p = .0000 n = 183</td>
<td>r = .51 p = .000 n = 183</td>
<td>r = .43 p = .30 n = 182</td>
</tr>
</tbody>
</table>
those respondents who lacked information for a particular item. Next, a simple regression model was run, analyzing all the predictor variables except for \(\text{religious}\) and \(\text{pbschool}\), because these items were not thought to be causal (see Table 3). They will be added after all the beginning regression models are performed, in order to first see the importance of issues relating directly to reproductive decisions. After this, a forward stepwise regression was performed, assuming a 0.15 probability to enter the consequent model. The analysis resulted in a regression model including only the following four variables: \(\text{nsex23}\), \(\text{agekids}\), \(\text{spcareer}\), and \(\text{hasfr}\), in that order. Table 4 presents a second multiple regression model that was run according to these same premises. Then, the last regression model was run - after adding in the measure of religiosity and \(\text{pbschool}\) - to see if these factors would influence the women's attitudes towards abortion (see Table 5). Finally, I ran one last forward stepwise correlation assuming a 0.15 probability to enter the model. The statistical analysis resulted in a regression model with the same variables as in Table 5, minus the rather insignificant \(\text{hasfr}\). Another multiple regression model was performed accordingly, and the results are displayed in Table 6.

### DISCUSSION

All of the nine initial factors assume that positively correlated relationships with the variable \(\text{abort}\) should not only make these women favor abortion, but they should also affect their moral judgments and lead them to prefer birth control policies. The variable \(\text{religious}\) was not tested in the first regression model because the underlying theory is that sexual activity combined with wanting to delay childbearing are causing all of these effects. Thus, \(\text{religious}\) may not be a causal variable in trying to understand \(\text{abort}\), and should therefore be tested only after those items that are assumed to be most certainly causally related to abortion attitudes. For this same reason, \(\text{pbschool}\) was also excluded from the initial tests. Therefore, the first model (Table 3) was run with only seven of the initial nine items. Overall, it explained 23% of the variance in the abortion scale, which is a relatively high amount, even though only two of the individual items themselves yielded a significant result: \(\text{nsex23}\) and \(\text{agekids}\) were both significant at the 1% level. The next multiple regression analysis conducted on the

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Std. Beta</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSEX23</td>
<td>.21</td>
<td>.007</td>
</tr>
<tr>
<td>AGEKIDS</td>
<td>.12</td>
<td>.09</td>
</tr>
<tr>
<td>SPCAREER</td>
<td>.16</td>
<td>.02</td>
</tr>
<tr>
<td>HASFR</td>
<td>.09</td>
<td>.19</td>
</tr>
<tr>
<td>RELIGIOUS</td>
<td>.29</td>
<td>.0001</td>
</tr>
<tr>
<td>PBSCHOOL</td>
<td>.33</td>
<td>.0001</td>
</tr>
<tr>
<td>R-squared for model:</td>
<td>.40</td>
<td></td>
</tr>
<tr>
<td>p-value:</td>
<td>.0001</td>
<td></td>
</tr>
<tr>
<td>n:</td>
<td>135</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Std. Beta</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSEX23</td>
<td>.21</td>
<td>.007</td>
</tr>
<tr>
<td>AGEKIDS</td>
<td>.12</td>
<td>.09</td>
</tr>
<tr>
<td>SPCAREER</td>
<td>.16</td>
<td>.02</td>
</tr>
<tr>
<td>RELIGIOUS</td>
<td>.29</td>
<td>.0001</td>
</tr>
<tr>
<td>PBSCHOOL</td>
<td>.33</td>
<td>.0001</td>
</tr>
<tr>
<td>R-squared for model:</td>
<td>.39</td>
<td></td>
</tr>
<tr>
<td>p-value:</td>
<td>.0001</td>
<td></td>
</tr>
<tr>
<td>n:</td>
<td>135</td>
<td></td>
</tr>
</tbody>
</table>
model obtained from the first forward stepwise regression (Table 4) yielded ambiguous results. The first two items were significant at the 0.7% level or better - a result that is significant under any normal circumstance - but the last two measures ([spcareer] and [hasfr]) were only significant at the 14% and 15% levels of significance, respectively. However, the variance only dropped one percentage point from the previous model, which is important because variance always decreases when additional items are removed from a model. Another interesting result is that both the first two standard beta values - the standardized regression coefficients, or the correlations that are left over after controlling for the effects of the other variables in the model - are relatively high, whereas the last two are not. Therefore, one can tell that both [nsex23] and [agekids] have relatively high correlations with [abort] after controlling for the effects of the other two variables. This result seems to accord with my findings of the individual significance values of each item.

The variance of the model in Table 5, considering only two additional variables were added, increased by a monstrous amount: just under 100%, from .22 to .40. Additionally, significant results were obtained for five out of the six items. Although [agekids] produced a borderline significance at just under the 10% level, all but one of the remaining variables yielded highly significant results. The variables [nsex23], [religious], and [pbschool] were all significant at or better than an astounding 0.7% level of significance and [spcareer] proved to be significant at the 2% level. Only [hasfr] yielded a non-significant result with a p-value = 0.19. It is also important to note that the overall model is also highly significant with a p-value of less than 0.0001. Lastly, I ran a forward step-wise regression to see if I could achieve significant values for all the individual predictive factors. The final model left me with the same variables as in the previous one, minus [hasfr]. After excluding that variable from the regression, the variance only decreased by one percentage point and now four out of the five variables were highly significant: [nsex23], [spcareer], [religious], and [pbschool] are all significant at or better than a 2% threshold level. Another interesting result is that [agekids] has become even less significant than in the last model, but the p-value is not nearly as inconsequential as that of [hasfr] from the model in Table 5. However, due to the relatively small sample size of this study, I am forced to conclude that a p-value of less than 0.12 is not significant, and therefore, I cannot use this factor to accurately predict abortion attitudes. Overall, the model itself was highly significant with a p-value less than 0.0001. For all these reasons, I concluded that this model was the best overall fit and that it will be of the most help in explaining the effects of the created abortion scale.

From this final multiple regression model, I am left with four highly correlated and highly significant predictors of the abortion variable: [nsex23], [spcareer], [pbschool], and [religious]. All of these factors yielded positive standard betas, which indicate a positively correlated relationship between each variable and [abort], as initially hypothesized. Additionally, all four variables produced p-values of .02 or lower, meaning that all four were significant at an astounding 2% level of significance or better. The overall model explained almost 40% of the overall variance at a 0.01% level of significance. Quite simply, this is a very solid resulting model, statistically speaking. It also had a final sample size of 135, which is a relatively large proportion of the initial 187 female respondents, given the fact of the multiple questionnaire distribution and collection sessions and the resulting incomplete data for certain participants. The model provides us with confirming results for four of my nine initial hypotheses, and the positive standard beta values also indicate relationships in the same directions as originally hypothesized. Statistical results do, however, give five neutral results for the remaining initial hypotheses - primarily because those five variables did not "reach" the final phases of analysis and did not enter into the stepwise regression models. Nevertheless, this does not necessarily mean that these factors are not correlated with abortion attitudes at all. In fact, examining the results shown in Table 2, we see relatively strong correlations between [abort] and a few of the variables that did not serve as good predictors according to this study. For example, the correlation between [abort] and [hasfr] is 0.17 (with p = 0.02), which suggests that [hasfr] might indeed be a good predictor of abortion values, when taken to be the sole explanatory effect. The same can be said for [pill], because the correlation between this contraceptive usage measure and [abort] is -0.24 (p = .005). This is a highly significant correlation that also seems to validate my initial hypothesis regarding this birth control factor - because lower values of this item reflect the contraception usage of a given woman, a negative correlation with [abort] does indeed show that women on the pill (low values) are more pro-choice (high values) in their beliefs. But once again, according to the final regression model of this study, these statements cannot be statistically supported.
What can be supported, however, is that I obtained significant (at a 2% level or better) confirming results for my four initial hypotheses involving \([nsex23]\), \([spcareer]\), \([pbschool]\), and \([religious]\). That is, the relationships of these variables in comparison with the abortion scale were all in the predicted directions. The results show that the total number of predicted male intercourse partners by age 23 is positively correlated with a woman's pro-choice beliefs. Similarly, the predicted ambition of a woman's future marriage partner in pursuing an income-generating career, her support for public school sex education and accessible birth control supplies, and her overall religiosity, are all significantly and positively correlated factors with her abortion attitudes. In examining the pairwise correlations of Table 2, one can see that 3 out of 4 of these "final" predictors of abortion attitudes have individual correlations with \([abort]\) of greater than 0.38. Thus, these results are exactly what the initial hypotheses predicted for \([nsex23]\), \([pbschool]\), and \([religious]\). As the expected number of sex partners increase, the more pro-choice a female is expected to be. Also, because these very women are engaging in sexual strategies designed to enhance future reproductive and sexual success, it makes sense that these women will delay pregnancy and look for a mate with a strong career, that generates large amounts of money and resources - this is exactly what these women are searching for in the first place, by utilizing these delaying and sexual strategy techniques. Similarly, if a woman supports sex education and available contraception in public schools, she should also support a woman's right to choose. A pro-life woman would not, according to the hypothesis of this study, support such open sexual education and apparent contraception support in public schools. The religiosity measure, also produced a relationship in the exact direction as that hypothesized. It was initially believed that the less religious a woman was, the less she would oppose contraception and the ability to choose, and thus, those women who are not highly religious are thought to be pro-choice. The less religious a person is - the less she goes to church, the weaker she considers her religious affiliation to be, the more liberal her beliefs are within her own religion, and the fewer the number of times she prays - the more pro-choice her viewpoints are hypothesized to be. This is exactly what the results show, confirming an initial hypothesis. All these results accord with the underlying sexual strategies theory that women delay pregnancy and practice sexual career and mating strategies in order to secure a more positive net gain in terms of sexual and reproductive success.

It is absolutely crucial to understand that the data set examined in this study includes only female undergraduates at the University of Pennsylvania. It is assumed that none of these participants are actively using government-funded programs for adolescents, such as Medicaid, and that the students are relatively all within the same socioeconomic class. Primarily for these reasons, then, the large variation in abortion attitudes among demographic subgroups that Henshaw and Kost (1996) had discussed, was not found in this study. This subpopulation of UPenn psychology volunteers is definitely not representative of any randomly chosen population eliciting other women's attitudes concerning the abortion debate, and thus makes this particular study especially difficult to interpret in a larger context.

Examining the results of the pairwise correlations in Table 2, an examiner finds another problem with this study: namely, co-linearity among various explanatory variables. If one takes \(r = .20\) to be an imaginary threshold value, inter-correlations are found among ten relationships of explanatory variables. For example, strongly correlated relationships are found between \([hasfr]\) and \([pill] (-.32)\); between \([spatt]\) and \([spcareer] (.40)\); between \([nsex23]\) and both \([pill]\) and \([spatt] (-.33 and .27, respectively)\); between \([religious]\) and each of \([pill]\), \([nsex23]\), and \([pbschool]\) (-.24, .22, and .21, respectively); and finally, between \([incedu]\) and both \([spcareer]\) and \([spatt] (.37 and .22, respectively)\). Nearly all of these correlated relationships are to be expected. For example, it is no surprise that having a boyfriend or being highly religious is correlated with contraception usage or that religiosity itself is correlated with the number of sexual partners expected by age 23. All these collinear relationships averse affect the overall variance and significance of the statistical models. This leaves the study with greater uncertainty in understanding the factors that affect the abortion attitudes of UPenn undergraduate women.

Yet another problem with the nature of this study is that the data were acquired through self-report mechanisms and are thus inherently susceptible to all associated errors, such as: lack of participant comprehension of a particular questionnaire item, misrepresentation or misinterpretation of a response or question, direct fabrication, or more simply, nonresponse - which seems to be one of the main inaccuracies of this particular type of research. Due to the different administrations and that certain questions were not asked at each session, nonresponse became an unrecoverable part of
the overall data set even before the respondents began
the questionnaires. In spite of these assumptions and
technicalities, however, the data set is indeed large
enough to accurately examine the population of UPenn
female psychology students and the underlying factors
influencing their abortion attitudes.

CONCLUSION

There is little doubt that students who take a
particular stand in the abortion debate share certain
ideas and beliefs with their opponents. As Cook et al.
(1992) points out, "Pro-life respondents do value per-
sonal freedom and autonomy, and pro-choice citizens
do have some concern for the fate of the embryo"
(156). Goggin et al. (1993) also show that among non-
activists, "many Americans […] endorse both sides of
the issue." The results of this study seem to support
this view among University of Pennsylvania female
psychology undergraduates, yet still produce four con-
firming results among my "partisan" hypotheses. This
study shows that even within this relatively homoge-
neous sample, four variables can be used as suitable
and influential predictors of abortion attitudes:
[nsex23], [spcareer], [pbschool], and [religious].

REFERENCES

absolutes: Public opinion and the politics of abortion.
Goggin, M. L. (1993). Understanding the new politics of
Granberg, D. (1982). "A comparison of members of pro- and
anti-abortion organizations in Missouri." Social Biology,
28, 239-252.
attitudes about sex education, family planning, and
abortion in the United States. New York: Louis Harris and
Associates, Inc.
abortions, 1982-1983." Family Planning Perspectives, 19,
5-9.
1995: Characteristics and contraceptive use." Family
Planning Perspectives, 28, 140-147, 158.
Major, B., Cozzarelli, C., Cooper, M. L., Zubek, J., Richards, C.,
responses of women after first-trimester abortion." Archives
of General Psychiatry, 57(8), 777-784.
INTRODUCTION

Language is a uniquely human phenomenon. People are able to communicate infinite ideas to one another through the use of words. Across the world people communicate in numerous different languages. Although people readily attain proficiency at any language they are exposed to as a child, the difficulties of learning a language later in life, past the critical period, are well known. While the cortical representation of language is fairly well understood, the neural substrates of bilingualism are still unknown. Researchers questioning whether multiple languages are represented by the same or distinct brain regions have yet to agree on an answer. Proponents of identical representations for multiple languages of bilingual speakers contend that the same brain regions mediate languages regardless of the type of language or whether they are native or foreign (Klein, 1994; Chee, 1999; Klein, 1999). Numerous neuroimaging studies have revealed similar brain regions active while subjects process both native and foreign languages. Other studies have found contrasting results. Neuroimaging studies have detected differential brain activation in response to tasks in various languages of a multilingual speaker (Dehaene, 1997; Kim, 1997; Perani, 1998). Other cases of bilingual aphasia, in which only one language of a bilingual is disrupted following brain damage, provide further evidence for separate neural representations of multiple languages (Gomez-Tortosa, 1995; Moretti, 2001; Paradis, 1989). Intraoperative methods, which provide the most direct measures of brain activity, have demonstrated regions of brain selectivity to each language of a bilingual (Ojemann & Whitaker, 1978; Pouratian, 2000; Simos, 2001).

CASE STUDIES

Neuroimaging studies of bilingual subjects have attempted to elucidate the localization of multiple languages within a single individual. A positron emission tomography (PET) study by Klein, Milner, Zatorre, Zhao, and Nikelski (1994) found no differences in the brain activation patterns as subjects processed words in either their first language (L1) or their second language (L2). All twelve subjects in this study were native English speakers who had learned French after the age of five and regularly used both languages. The left inferior frontal cortex showed a similar response to both English and French words. However, the authors warn that results from this study should not lead one to conclude common neural substrates for native and second languages because the similarity in brain activation between L1 and L2, in this case, may be due to the fact that English and French are both Indo-European languages.

To test the hypothesis that the extent of similar
activation patterns of L1 and L2 are due solely to the similar characteristics of L1 and L2. The structure of Chinese is drastically different from that of English - the first language being an ideographic script and the second an alphabetic script. Functional magnetic resonance imaging (fMRI) revealed the same pattern of activation as subjects performed cued word generation tasks in both Chinese and English. Subjects were visually presented with either the first or second half of a word and instructed to complete it. Both languages activated overlapping regions of prefrontal, temporal and parietal regions as well as the supplementary motor area. Based on these results, and supporting those of Klein et al. (1994), Chee and colleagues (1999) argue for a common cortical representation, even between two languages as disparate as Chinese and English.

A subsequent study, by Klein, Milner, Zatorre, Zhao, and Nikelski (1999), also testing Chinese/English bilinguals, found results consistent with those obtained by Chee and colleagues (1999). PET detected increased cerebral blood flow in the left medial temporal lobe, left parietal cortex and right cerebellum in response to verb generation in both Chinese and English. Since no differences were found in the activation patterns between Chinese and English verb generation, Klein et al. (1999) conclude that language specificity does not influence its organization in the brain.

More recently, based on the results from an event-related functional magnetic resonance imaging (ER-fMRI) study of Chinese-English bilinguals, Pu and colleagues (2001) argue for a shared neural mechanism for the processing of native and second languages. Investigation of the left inferior to middle frontal lobe revealed parallel neural activity-induced hemodynamic responses during verb generation tasks in each language. These findings suggest that not only are the same brain regions activated by each language, but moreover that they operate on a similar time scale. However, not all studies have unanimously concluded that multiple languages are represented and processed by common brain regions. There is an enormous body of evidence that demonstrates polyglot speakers recruit distinct areas for each language. A fMRI study by Dehaene et al. (1997) found a dissociation between cortical areas involved in first and second languages. Native French (L1) speakers who had acquired English (L2) through school after the age of seven listened to stories in each language. The regions of the left superior temporal sulcus, superior and middle temporal gyri showed consistent activation across subjects during presentation of L1. This finding is consistent with traditional language areas associated with language in monolinguals. On the other hand, L2 yielded a wide assortment of activation patterns that was not consistent between subjects. In fact, in some subjects, only right hemisphere activation was found in response to L2.

Electrocortical stimulation in bilingual patients undergoing surgery to treat severe epilepsy has offered insight into the neural substrates underlying multiple language representations. Ojemann and Whitaker (1978) found both overlapping and distinct areas of the cortex involved in each language of two bilingual patients using stimulation mapping with a naming task. More specifically, stimulation of the posterior temporal lobe disrupted both languages. However, stimulation of other cortical areas produced naming deficits in one language without any disruption of the other. Of the two patients used in this study, one was a native Dutch (L1) speaker who subsequently learned English (L2) and the other patient was exposed to both English (L1) and Spanish (L2) in infancy, but predominantly used Spanish. However, despite these differences, both subjects showed the same basic pattern of effects. In both patients, a broader range of cortex represents L2 than does L1.

Similar results were obtained by Pouratian et al. (2000) with a novel technique - intraoperative optical imaging of intrinsic signals (iOIS). Intrinsic cortical light reflectance is measured with iOIS using white-light illumination to detect changes in capillary beds in the brain. Language mapping on a Spanish-English bilingual patient undergoing surgery to remove an astrocytoma near the left perisylvian cortex demonstrated a double dissociation between the patient's two languages. Although the lesion was proximal to the classic language areas around the left sylvian fissure (Gazzaniga, Ivry, & Mangun 1998), language mapping with iOIS successfully spared all of the patients’ language areas and there were no impairments reported post-operatively (Pouratian, 2000). Both languages of this patient activated structures of the superior temporal sulcus and the superior and middle temporal gyri. Consistent with electrical cortical simulation results (Ojemann, 1978), areas were identified that showed activity for each language individually. Naming in Spanish distinctly activated the supramarginal gyri whereas naming in English revealed the precentral gyri as a language specific area.
Simos et al. (2001) used magnetic source imaging to determine the location of the receptive language maps in bilingual individuals. The subjects used in this study comprised of Spanish-English bilinguals, some of whom learned English as their native language while others learned English in school. In each bilingual subject, the posterior superior temporal gyrus comprised both languages' receptive maps; however notable differences were observed in the supramarginal and superior temporal gyri. In all subjects, both of these regions were found to contain language-specific maps; however, there was no consistency amongst individuals anent which language, Spanish or English, corresponded to which region. This suggests that the differences in location of each language-specific map cannot be attributed to any innate differences in the languages themselves. Each language of the bilingual speakers demonstrated a spatially discrete area of cortex that was not implicated in the other language. However, Simos et al. did not report whether there was any consistency in brain regions supporting each language within the two groups of bilinguals studied.

Evidence from bilingual aphasic patients provides compelling evidence in support of differential brain representation of multiple languages. Brain damage that selectively affects one language but not another in a bilingual supports the notion of separate neural substrates for each language. Gomez-Tortosa, Martin, Gaviria, Charbel, and Ausman (1995) describe the case of a bilingual woman who displayed an impairment following surgery in one language, but not the second. Pre-operatively, this woman had comparable command of both Spanish, her native language, and English, her second language. Two months subsequent to a left frontotemporal craniotomy to remove a perisylvian arteriovenous malformation (AVM), the patient showed decreased fluency in Spanish, but intact functioning of English. Hines (1995) suggests that these results do not provide evidence of a selective deficit in one language. The analysis performed by Hines on the data obtained by Gomez-Tortosa et al. (1995) revealed no statistically significant impairment in the patient's native language following resection of the AVM. But this claim is countered by the fact that the patient herself noticed and reported the difficulty in her ability to find words in Spanish. Thus, there was, at least to some degree, a selective deficit in language skills following surgery. Furthermore, the impairment of one language but not the other suggests differential brain organization of the two languages. An integral region of the Spanish lexicon was damaged by the surgery yet left the patient's English competence intact.

Another study of bilingual aphasia brings up a fundamental difference between native and second languages. Language learning in infancy does not depend on explicit procedures of memorizing vocabulary and grammar rules that the acquisition of a language later in life requires, but rather exceedingly on a more unconscious system (Paradis, 1994). Thus, a native language is subserved by the implicit memory system, while other languages are reliant upon the explicit memory system. In particular, the medial temporal lobe and diencephalon are crucial for explicit memory, whereas a variety of subcortical structures and the neocortex are used to mediate implicit memory functions (Gazzaniga, 1998). Moretti et al. (2001) described a Croatian-Italian bilingual patient who suffered severe impairments in her native language (Croatian) after an infarct of the caudate. Subsequently, the patient showed improvement of Croatian and a drastic decline in performance of Italian accompanied by an extension of the ischemic lesion to the cortex, particularly to the left frontal and temporal lobes. The performance of this patient is in line with the hypothesis that cortical structures used for explicit memory have a role in second language processing, whereas one's native language is processed predominantly by subcortical regions, such as the caudate. One issue that has yet to be resolved is how this patient was able to regain L1 abilities following the initial impairments.

Other studies have examined how the level of proficiency and age of acquisition of a second language affect its representation in their attempt to unravel the complex cerebral organization of multiple languages in the bilingual brain. One study found that age of acquisition of the second language affects its neural representation (Kim, Relkin, Lee, & Hirsch 1997). Overlapping brain regions support native and second languages when the second language is learned early in development. Subjects who learned a second language as young adults also showed similar regions of activation in the superior temporal gyrus (Wernicke's area). However, spatially distinct regions of activation were present in the inferior frontal gyrus (Broca's area) of all late bilingual subjects studied. Since differences were seen in Broca's, but not Wernicke's, area depending on age of acquisition of the second language, Kim et al. (1997) suggested that there may be a critical period for language learning that is determined by Broca's area. This area is set early in development by repeated exposure to a language and cannot be modified by an additional language learned later in life.
Perani et al. (1998) found similar brain activation in late bilinguals as compared to early bilinguals if a high level of proficiency was obtained in their second language. In addition, they found differential patterns of cortical activity depending on the degree of fluency of the late bilinguals. All factors except for degree of proficiency were constant across the two late acquisition groups. Thus, in contrast to the Kim et al. (1997) study, Perani et al. (1998) argue that age of acquisition of a foreign language does not have as much bearing as does the level of proficiency on its cortical organization.

Perani and colleagues (1998) suggest that the apparent discrepancy in the findings may stem from the fact that no proficiency test was given to the late bilingual subjects used by Kim et al. (1997). This raises the possibility that the differential cortical representation that Kim and colleagues (1997) attribute to age of acquisition may actually be a result of dissimilarities in fluency level between the two groups. In this case, the results of the two studies may actually support one another.

However, a case presented by Paradis and Goldblum (1989) of a trilingual aphasic patient (A.M.) demonstrates differential representation of languages in the brain even when the languages present identical acquisition histories and fluency levels. Of the three languages A.M. could speak, both Gujarati and Malagasy were acquired during infancy, while French was obtained at school. Immediately after the removal of a parasitic cyst in the right prerolandic area the patient developed a non-fluent aphasia in Gujarati. Meanwhile, he could easily and accurately express himself in either of the other two languages. Eight months after the surgery the pattern of impairments had shifted such that A.M.’s fluency in Gujarati was back to normal, while deficits manifested in Malagasy. Two years later all three languages were fully recovered. During this whole sequence, both of A.M.’s native languages were selectively impaired at one time or another, while French remained fully functional throughout. Although it is not clear exactly what induced each selective deficit, it is clear that the three languages were represented by A.M.’s brain in a slightly different manner.

**DISCUSSION**

Based on findings with electrical stimulation mapping that L2 is represented by a larger area of cortex than L1, Ojemann and Whitaker (1978) suggests that a large portion of neurons are recruited as one learns a language. However as proficiency increases, the area of cortex required for processing of that language decreases. This may be due to the increased automatization of language with continued use. This hypothesis is in concord with a later proposal by Dehaene and colleagues (1997). They suggest that there are differences in the cerebral microcircuitry between the languages of a polyglot, however these differences are too small to be detected with the current imaging techniques.

The methods and experimental design must be taken into consideration when interpreting the results from these studies. Variations in experimental tasks, imaging procedures, and protocols used are the likely reasons for the incongruity in the findings (Simos, 2001). The nature of PET studies requires that all subjects be averaged together. The use of fMRI yields higher resolution images and allows for data to be analyzed subject by subject. Thus, PET may not detect actual differences between L1 and L2 if the activation patterns are not consistent across subjects (Kim, 1997). The studies that reported differences in cerebral organization of native and foreign languages commented on the fact that no consistent pattern of cerebral organization could be attributed to L2 (Dehaene, 1997). Thus, a PET study that fails to find differences in L1 and L2 is not conclusive evidence against distinct structures supporting each language.

Even greater spatial resolution is obtained with the use of iOIS, which measures changes in intrinsic cortical light reflectance, than with PET or fMRI (Pouratian, 2000). Higher spatial resolution offers the opportunity to detect minute differences in activation levels. Pouratian et al. (2000) found differences on the scale of 4mm in language representations. Clearly this must be kept in mind while interpreting data obtained from techniques which do not provide such a fine spatial resolution.

Another element that influences the outcome of a study is the method of analysis. Pu et al. (2001) focused their analysis solely on the left inferior to middle frontal lobe and from this concluded common neural mechanisms for multiple language processing. Activation occurring on the same time course except at different places within this broad region of the left frontal lobe would not be detected with this method of analysis. Furthermore, other studies have demonstrated differential representation of multiple languages in regions other than the left inferior to middle frontal.
lobe. For example, hemispheric differences in the nature of native and foreign languages have been demonstrated (Dehaene, 1997). Thus, the fact that Pu and co-authors (2001) failed to find differences in the hemodynamic response to Chinese verses English stimuli may simply be because of the brain region they choose to investigate. They justify this constraint by claiming that this region encompasses several Brodmann’s areas known to be implicated in language processing of Chinese and English individually (Pu, 2001). Consequently, it is not possible to dismiss the evidence in favor of distinct neural regions for each language of a polyglot based on this study.

Further evidence for differences in languages come from data showing increased right hemisphere activity when processing a non-native tongue (Dehaene, 1997). L1 consistently produced greater activation in the left hemisphere than the right, however a different pattern was found of L2. In some subjects, although the left hemisphere showed dominance for L1, only the right hemisphere was activated by comprehension of L2. Dehaene et al. (1997) proposed that the method of acquisition and fluency level of L2 may be involved in the variation of cortical representations of L2. Other studies have confirmed the existence of acquisition history affecting processing strategies. The highly automated process of manipulating L1 does not depend on the explicit memory system in the manner that L2 requires (Moretti, 2001). Thus, the right hemisphere may initially be dominant for L2 in individuals who use different strategies to process L2 than they use for L1. As one becomes more proficient in a language, the left hemisphere may become more involved. However, the role of the right hemisphere does not dissolve, as demonstrated by patient A.M.’s pattern of language impairment after a resection of the right prerolandic area. Even though fluent in both Gujarati and Malagasy, the right hemisphere contributed elements to the processing of these languages. Yet even with part of the right hemisphere removed, A.M. was able to recover language functions.

Studies employed various linguistic tasks to measure brain activity while processing different languages. While there may be distinct regions of cortex implicated in each language, the nature of the task might not target the regions of difference. Kim et al. (1997) tested covert language production, whereas Perani and colleagues (1998) used a language comprehension task and obtained different results. The findings can be explained by the different requirements of the language tasks, which are known to recruit distinct brain areas. Wernicke’s area is believed to be involved in language comprehension (Gazzaniga, 1998), and subjects understand what they are producing regardless of age of acquisition. Kim et al. (1997) observed no differences in Wernicke’s area activation; however the nature of the language production task would not be expected to produce differences between early and late bilinguals in this region (Perani, 1998). Consequently, there may be spatially distinct regions within Wernicke’s area for multiple languages that were not detected by the covert generation language task exercised.

Cases of bilingual aphasia afford an excellent opportunity to study language processes. The pattern of aphasia following injury to a bilingual is very diverse and therefore results obtained should be waryly approached. Not all cases present selective deficits in the manner of previously described studies (Gomez-Tortosa, 1995; Moretti, 2001; Paradis, 1989). There are numerous reports of aphasia simultaneously affecting both of a bilingual patients languages following lesions of the left perisylvian area (Gomez-Tortosa, 1995). This conflicting data can be resolved when one keeps in mind that lesions are often widespread. Damage to a large region of the left perisylvian area would be expected to affect both languages if the left perisylvian area is implicated in both languages of a bilingual. On the other hand, a smaller lesion could selectively disrupt one language but not the other, such as case of the patient reported by Gomez-Tortosa et al. (1995). Thus, cases of similar disruption of multiple language functions following brain damage do not prove common cortical representation, while instances of incongruous impairments strongly argue against a single language system which encompasses all languages in bilinguals.

Despite claims that common brain regions subserve multiple languages, the bulk of evidence points towards some distinction in the representation of each language. The current literature provides compelling evidence that there are dominant language areas that are involved in all languages, yet segregated regions for each language also exist. The fact is, many irrefutable differences have been found, even when controlling for factors such as age of acquisition, level of proficiency, and actual language tested. However, the actual neural mechanisms underlying multiple language processing have yet to be fully understood. While there is evidence that differences occur even between languages with identical acquisition history and fluency, it is not the case that these factors do not have an effect on the organizational patterns of the lan-
guage. It is clear that there will be no easy solution to the understanding of such a complex process. The advent of new techniques to the field should aid this inquiry, however further research needs to be done that encompasses all aspects of language in order to elucidate the nature of second language processing and, in doing so, shed light on the nature of language in general.

REFERENCES

Chee, M W L, Tan, E W L & Theil, T (1999) “Mandarin and English single word processing studied with functional magnetic resonance imaging” Journal of Neuroscience, 19, 3050-3056


Paradis, M, & Goldblum, M C (1989) “Selected crossed aphasia in a thalamic aphasic patient followed by reciprocal antagonism” Brain and Language, 36, 62-75


Perspectives in Psychology
Spring 2002  Ψ  55
Perspectives in Psychology
The Undergraduate Psychology Journal of the University of Pennsylvania
c/o Department of Psychology
3815 Walnut Street
Philadelphia, PA 19104