

What's In A Face: Accuracy Of First Impressions Based Upon Facial Appearance

John Hubbard
Macalester College (Student)

Faculty Advisor:
Jack Rossmann
Macalester College

Abstract

The notion that first impressions are somewhat accurate is an intriguing possibility. Several studies have shown correlations between self and stranger ratings of personality, sometimes with as little exposure as a facial photograph. The reasons for this accuracy remain unclear. Possible explanations include everything from self-fulfilling prophecies to a genetic link between appearance and personality. One reason may be that emotional expression is a functional aspect of personality, and that people's expressions, over time, form lasting features on their faces. The accuracy of facial impressions would therefore increase with the age of the people being judged. An experiment was run that tested this claim. Target subjects were run in married couples, who rated themselves and their spouses on personality traits. Facial photographs were taken of each target and shown to judge subjects, who rated the targets on the same personality traits. Self, spouse, and stranger ratings of personality were compared. Although the correlations between stranger and other ratings were minimal, they were higher for older targets. These results are discussed and reasons for the accuracy of first impressions are further considered.

What's in a Face: Accuracy of First Impressions Based Upon Facial Appearance

The captain of the H.M.S. Beagle believed that personality is shown in facial characteristics. When he interviewed Charles Darwin for a post aboard his ship, it was not the young man's qualifications that deterred him, but the shape of his nose. The captain thought that Darwin's large nose was a sure sign of a sluggish personality (Fancher, 1990).

First impressions are formed when people observe others for the first time, and then further ascribe personality traits based on those observations. First impressions play an important role in human interaction because they dictate the ways in which people anticipate reactions from others. How people relate to each other as well as what and how relationships are formed are therefore influenced by first impressions.

It is important for people to realize how first impressions are formed so that they may be aware of the effects of their own behavior and the possible inaccuracy of their judgments of others. To this aim, an experiment was run concerning the accuracy of first impressions from facial appearance.

Theoretical Framework

Relevant Variables

The variables in an experiment on interpersonal perception are the judges, the targets, the inputs, and the judgments (Gage & Cronbach, 1955). These judgments could be compared with other ratings to determine consensus and accuracy. It is difficult to isolate one aspect of impression formation in a natural setting, so inputs such as photographs of targets are required. This is because many other factors, such as nonverbal behavior (Burgoon, Woodall, & Buller, 1989), eye contact (Knacksted, 1991), and style of dress (Hamid, 1968; Satrapa, et al., 1992) have all been found to have significant effects on impression formation.

Other factors that are present in facial appearance have also been shown to have effects on personality judgments. Among these are the gender (Albright, Kenny, & Malloy, 1988) and attractiveness (Albright, Kenny, & Malloy, 1988; Dion, Berscheid, & Walster, 1972; Kenny, Horner, Kashy, & Chu, 1992) of the target. Cosmetic use (Workman & Johnson, 1991), wearing glasses (Hamid, 1972; Thornton, 1944), the appearance of facial maturity (Berry & McArthur, 1986; Bronlow, 1992; Meerdink, Garbin, & Leger, 1990), and baldness or facial hair (Wogalter & Hoise, 1990) have also been shown to have effects upon personality judgments.

Defining Accuracy

An accurate judgment is one that is agreed upon by others and corresponds to a criterion (Funder, 1987). A consensus in judge ratings is needed due to individual differences among judges. Because a person is not viewed in the same way by all people, the perspective of the judge is an important part in the judgmental process (Schneider, Hastorf, & Ellsworth, 1979). The gender (Feingold, 1991), cultural background (Keating, et al., 1981), individual beliefs (Dion & Dion, 1987), expectations (Osbourne & Gilbert, 1992), and mood (Forgas & Bower, 1987; Tesser & Collins, 1988; Tesser, Pilkinton, & McIntosh, 1989; Toner & Gates, 1985) of the judge has an effect of how impressions are formed. Because of the individual variances of the judges, from an individual standpoint, it is still hard to find general laws of first impressions. Depending on who they are, each person will judge others a little differently. In order to cancel out any individual differences in judge ratings, a reliable average of ratings must be used.

There is a problem of what criterion to use because there is no purely objective measure of personality (Zebrowitz, 1990). One objection to using self-ratings is that they may reflect

inflated self-views, and self/judge correlations are found due to generous rating by judges (Paunonen, 1991). One study compared reflective self performance ratings on a group task with staff assessments (John & Robins, 1994). The self ratings often showed a positive bias, but self rating accuracy existed and correlations were found between the rating methods. Another good way to empirically test ratings is to predict behavior. One analysis of many self vs. clinical predictions showed that self assessment is usually a more accurate predictor of life's events (Shrauger & Osberg, 1981). Thus, self rating is shown as just as accurate a psychological evaluation as any clinical test.

Measuring Personality

The personality test used in this experiment has an extensive historical foundation. An early step in describing personality was the cataloging of 4504 trait adjectives (Allport & Odbert, 1936). This list was reduced to 171 variables, then to 35 bipolar scales that described 12-15 personality factors (Cattell, 1945, 1946). Fiske (1949) and Tupes & Christal (1961) next showed that five robust factors were sufficient to represent the structure of personality. These five factors are Neuroticism, Extroversion, Conscientiousness, Agreeableness, and Culture. For a historical review of the early development of the five factor model of personality, see Wiggins (1973). The initial testing of the five factor design showed that using 20 bipolar scales could adequately describe the five factor structure (Norman, 1963, 1969; Norman & Goldberg, 1966; Passini & Norman, 1966). The robustness of the five factor model has been further replicated more recently (Digman & Inouye, 1986; McCrae & Costa, 1985, 1987; Noller, Law & Comrey, 1989), leading to the conclusion that the five factor model is a valid taxonomic scheme for personality (McCrae & Costa, 1986). For a review of the evidence relating to the comprehensiveness of the five factor model of personality, see McCrae & John (1992).

Accuracy

Stereotypes

People draw many conclusions about a person's character from things as limited as facial appearance. These conclusions are usually agreed upon by other observers (Alley, 1988). Consensus in personality judgments at zero acquaintance would not be expected if people each guessed in their own way. Zero acquaintance is an important condition that causes judgments to be made only from physical appearance. Consensus may be due to a combination of accuracy and common stereotypes.

This points out the role of learned cultural stereotypes in social perception. It is necessary to look at the aspects of impression formation in a cultural context because of the varying social mores of different societies. For example, results from an attractiveness halo effect may differ across cultures not only because of the different standards of beauty, but because different traits are more socially desirable in different societies.

Accuracy with the five factor model

One study used the five factor model for measuring self and judge views of personality (Watson, 1989). Classes of 1540 unacquainted students rated themselves and each other on Extroversion, Agreeableness, Conscientiousness, Emotional Stability, and Culture at the beginning of the first day of class. A significant correlation was found between the self and judge ratings on all traits except Emotional Stability. Judgments of Extroversion were similar between judges and targets ($r = .41$). Another study showed a similarity between stranger and peer ratings on the same five personality traits (Passini & Norman, 1966). In a third study, groups of five strangers rated themselves and each other on the above five personality traits. Self/judge correlations were .33 for Extroversion and .46 for Conscientiousness (Albright, et al., 1988). The correlation of self and stranger ratings was further investigated with the five trait model of personality (Borkenau & Liebler, 1992). Based on only still photographs of a person seated at a desk, the self and judge ratings correlated $r = .33$ for Extroversion and $r = .32$ for Conscientiousness. Ratings of physical attributes correlated with Extroversion and Conscientiousness ratings.

Physiognomy

Perhaps even such a thing as facial wrinkle patterns are signs of personality. It has been suggested that personality can be interpreted as the tendency to experience certain emotions, and that emotional tendencies are a functional aspect of personality (Plutchik, 1980). In a study of older adults, supposedly neutral expressions showed the effects of emotion expression biases, and posed emotions were inaccurately judged as the expression of other self-rated frequently expressed emotions (Malatesta, Fiore, & Messina, 1987). The adage that children should not make faces because they might freeze that way may be somewhat true.

Phrenology was a popular belief fostered during the 1800s that nearly all personality characteristics were shown by the brain's shaping of the skull (Mainwaring, 1980). This idea was expanded to facial characteristics, but physiognomy failed to show that particular traits were isolated in specific facial characteristics, so a more holistic approach was taken (Liggett, 1974). Physiognomy currently exists as a pseudoscience which attempts to read personality in facial appearance. There is not a great deal of scientific respect for physiognomy these days because of its overextended applications (Brandt, 1980). However, the following experiments suggest that some type of facereading is possible, but just not on such a scope as early physiognomists believed.

Facial appearance and personality

Self and stranger ratings of facial babyishness relate to self views of personality traits. Facial babyishness correlates positively with warmth and negatively with power (Berry & Bronlow, 1989). Based on facial photographs, judge ratings of honesty negatively correlated with the targets' likelihood to volunteer to take place in a deceptive experiment (Bond, Berry,

& Omar, 1992). The rated dominance from facial photographs of graduating West Point cadets served as an accurate predictor of rank achievement: the rated facial dominance of cadets had a .54 correlation with cadet rank in senior year (Mazur, Mazur, & Keating, 1984).

To further test the correlation of judgments from facial appearance and personality, first impressions were measured against the judgments of acquaintances (Berry, 1990). Seminar psychology students were asked to rate fellow classmates on the first day of class, then during the fifth week of the semester, and again during the ninth week. A rating scale was used for personality traits such as dominance, honesty and warmth. During the semester, a color facial photograph of a neutral expression was taken of each of these students. These photographs were shown to other subjects, who rated the students on the same personality scales. The photographic impressions paralleled the students' impressions consistently throughout the semester. Also, the rated impressions from the photographs corresponded to the targets' self-ratings on personality tests. In another comparison of selfreports and first impressions, judges were given either facial or vocal impressions (Berry, 1991). Results showed that ratings from either facial photographs with neutral expressions or neutral voice recordings could predict self views on the traits of power and warmth. It appears that facial impressions are a somewhat accurate judgment of personality.

Review

These results show that we are hardly infallible in detecting personality traits based on little exposure. Two statistical analyses, Norman & Goldberg (1966) and Paunonen (1991), have pointed out how little accuracy is needed to achieve such experimental results. For example, one simulated study had ten judges per target, and a third of the judge groups contained two perfect raters, and all other judges guessed at random. The computed self/judge correlation was .25 (Paunonen, 1991). However, the standard effect is that a self/judge correlation is found; stranger ratings are significantly higher than chance. A meta-analysis of 44 studies measuring the accuracy of first impressions yielded a self/judge r of .39 (Ambady & Rosenthal, 1992). The causes of such accuracy cannot be readily examined, but several hypotheses seem to work with each other as explanations.

Hypotheses

Genetic causes

The first hypothesis deals with biology and genetics. This hypothesis says that the genes that affect personality are correlated with the genes that influence physical appearance. There may even be an evolutionary basis for this correlation: just as the Gila Monster has developed bright coloring to warn other animals of its poison, people may have evolved to show their personalities in their faces. Therefore, there is bound to be some relationship between physical appearance and personality, and subsequent accuracy of first impressions.

Facial development

Modern face reading places more emphasis on expressions than structural features (Street, 1990), but these two may be related. Different from the genetic causes of the bony structure of the face is its muscular composition, a possible sign of habits of expression (Allport, 1937). The second hypothesis says that facial expressions, over time, form lasting marks on facial structure. This effect would cause the face to become an illustration of the types of emotions it is used to showing. The accuracy of first impressions could be partially due to the face showing personality in this manner. This is the specific hypothesis that was chosen to attempt to isolate and test.

Social influences

The third concept that may be at work is a selffulfilling stereotype. An example of such an effect was shown by having men talk with women who they believed to be either attractive or unattractive (Snyder, Tanke, & Berschied, 1977). These beliefs elicited behavior that caused other listeners to judge the women just as the other subjects believed them to be, either attractive or unattractive. Because this shows how beliefs can lead to reality, this is support that a consensus in judgment may be related to accuracy. Such a consensus could even arise from a shared belief in one of the above theories.

Proposal

Specific research on one of these three hypotheses is difficult because their singular effects are difficult to isolate as such, so any attempt to analyze a singular hypothesis will not yield conclusive results. It seems prudent to realize that there are many variables involved in the accuracy of first impressions. However, an experiment was run which emphasized the facial development hypothesis. Also, the accuracy of first impressions from facial appearance still needed to be examined in general because it is not a widely accepted phenomenon (Brandt, 1980; Hinton, 1993).

What was proposed, was to give people (judges) stimuli of others (inputs of targets), and to ask them about their first impressions of the targets (judgments). By comparing these first impressions with selfviews of personality, the accuracy of the judgments could be tested. This accuracy could also be compared with peer accuracy by having peers rate targets as well. Because the accuracy of personality judgments from facial appearance is not a widely studied phenomenon, this topic was chosen for experimentation.

Hypothesis

To relate to the hypothesis that personality becomes illustrated in facial appearance over time, the effects of the age of the target on the accuracy of personality judgments were measured. It was hypothesized that the face becomes more influenced by the expression of personality over time. Therefore, if this is a main criterion that is correctly gauged in

judgments of personality, then the accuracy of judgments would increase with the age of the targets. This was a novel experimental idea because the hypothesis that the face becomes molded by its expressions over time is only suggested by the results from Malatesta, Fiore, & Messina (1987). Also, the only previous experiments on the accuracy of first impressions from facial appearance, Berry (1990) and Berry (1991), have both used college-aged targets.

Method

Materials

The questionnaire used is based on the five factor model, and is similar to the inventories used in Norman (1963) and Watson (1989). The version used has twenty questions. Each question asks for a numbered response from 1 to 5, with adjective pairs (such as talkative and silent) on the endpoints. Subjects were instructed to circle the number that they felt best described the person that they were rating. For example, for the talkative/silent question, subjects would circle 1 if the person is very talkative, 4 if the person is somewhat like silent, and so on. Four questions are devoted to each of the five traits that the test measures. The test measures the following traits: Extroversion, Agreeableness, Conscientiousness, Neuroticism, and Culture. These five traits and their component adjective pairs are listed in Table 1.

Table 1 <u>Factor Scales and Items</u>
<u>Extroversion</u> Talkative vs. Silent Frank, open vs. Secretive Adventurous vs. Cautious Sociable vs. Reclusive
<u>Agreeableness</u> Goodnatured vs. Irritable Not Jealous vs. Jealous Mild, gentle vs. Headstrong Cooperative vs. Negativistic
<u>Conscientiousness</u> Fussy, tidy vs. Careless Responsible vs. Undependable Scrupulous vs. Unscrupulous Persevering vs. Quitting, fickle
<u>Neuroticism</u> Nervous, tense vs. Poised Anxious vs. Calm Excitable vs. Composed Hypochondriacal vs. Not so
<u>Culture</u> Artistically sensitive vs. Insensitive Intellectual vs. Unreflective, Narrow Polished, refined vs. Crude, boorish Imaginative vs. Simple, direct

Target Subjects

Targets were recruited through a variety of personal acquaintances. In order to obtain a specific type of peer rating, only married couples were recruited. Couples were asked if they would volunteer fifteen minutes of their time to participate in a research study. After signing a general consent form, targets were first asked to fill out a first-person version of the questionnaire. Subjects then were asked to rate their spouses, using a third-person questionnaire. Subjects were next asked if they would agree to be photographed for future research. Consenting subjects were asked to remove all visible jewelry and a black cloth placed around their shoulders to conceal any visible clothing. Pictures were then taken of them from the shoulders up against a black background. Targets were asked to maintain a neutral facial expression while being photographed. The ages of the subjects and how long the couple had been married were then recorded.

Target subjects numbered 25 married couples who filled out the questionnaires completely and whose pictures developed clearly. Targets were divided into two age groups. The 24 members of the older group were from 50 to 74 years old, and had an average age of 62.08(6.60). The 26 members of the younger group ranged from 28 to 49 years old, and had an average age of 39.77(5.75). Couples were also divided into two groups based on marriage length. The 12 newlywed couples had been married from 2-16 years, for an average of 9.83 (4.15) years. The 13 veteran couples had been married an average of 33.83(8.41) years, with a range of 22-46 years. A color slide was prepared of each of the targets.

Judge Subjects

Students in Macalester College psychology classes received course credit for volunteering to serve as judges. A total of 42 judges were run in 3 sessions of 14 judges each. Each session of judges viewed a different random third of the target subjects in a random order. Judges viewed a series of 16 or 17 slides of targets, and were given 3 minutes to complete a third-person questionnaire for each of the slides. Judges were asked to fill out each questionnaire completely. In this manner, fourteen stranger ratings were acquired for each target.

Experiments lasted approximately one hour. None of the judges admitted to having been previously acquainted with any of the targets. Because most target subjects lived several states away from the college, it can be assumed that none of the judges and targets were previously acquainted. At the end of the experiment, judges were asked to write their age and gender on their packet of questionnaires. Of the 42 subjects, there were 11 male and 31 female judges, and judges had an average age of 19.55(2.04). Nineteen of the 14,000 questions that the judges were given were skipped (.14 %), and were scored as a 3.

Results

Reliability

Judge reliability was measured for each of the five traits using Cronbach's Alpha. An Alpha increases with a higher variance of individual judges' total scores for all targets, and decreases with a higher variance of judge scores within a single target. Reliability scores were .52 for Extroversion, .31 for Agreeableness, .33 for Conscientiousness, .44 for Neuroticism, and .28 for Culture. The mean Alpha across all five traits was .38. For each target, the judge ratings were averaged together to be compared with the self and spouse ratings.

Self/Spouse correlations

Table 2		Correlations between self ratings and spouse ratings					* p < .05	
Group	n	Extroversion	Agreeableness	Conscientiousness	Neuroticism	Culture	Average	
<u>Marriage Length</u>								
2-16 Years								
22-46 Years	24	.78*	.18	.26	.41*	.45*	.42	
	26	.76*	.30	.51*	.48*	.34	.48	
<u>Gender</u>								
Male Targets								
Female Targets	25	.82*	.18	.49*	.47*	.66*	.52	
	25	.58*	.49*	.25	.38	.07	.35	
<u>Age</u>								
28-49 Years								
50-74 Years	26	.69*	.08	.21	.38	.18	.28	
	24	.82*	.51*	.57*	.49*	.60*	.60	
All Targets	50	.76*	.27*	.37*	.42*	.38*	.44	

Table 2 shows the correlations between self and spouse ratings of personality for each of the five traits and for specific target groups. The correlations between self and spouse ratings for all fifty targets were highest for Extroversion (.76), and lowest for Agreeableness (.27). The average correlation between self and spouse ratings for all targets across each of the five traits was .44.

Veteran couples scored only slightly higher on the average of these correlations than Newlywed couples (.48 vs. .42). However, the correlations between males' self ratings and their wives' ratings of them were on average higher than the correlations between females' self ratings and their husbands' ratings of them (.52 vs. .35). Also, older targets' self ratings corresponded more closely to their spouses' ratings of them than did with younger targets (.60 vs. .28).

Spouse/Judge correlations

Table 3		<i>Correlations between spouse ratings and stranger averages</i>					* p < .05	
<u>Group</u> <u>Criteria</u>	<u>n</u>	<u>Extroversion</u>	<u>Agreeableness</u>	<u>Conscientiousness</u>	<u>Neuroticism</u>	<u>Culture</u>	<u>Average</u>	
<u>Marriage</u> <u>Length</u>								
2-16 Years								
22-46 Years	24	.05	.13	.05	.02	.14	.08	
	26	.24	.24	.62*	.26	.10	.25	
<u>Gender</u> <u>Male</u> <u>Targets</u>								
Female Targets	25	.15	.18	.45	-.06	.03	.11	
	25	.01	.30	.26	.37	.18	.22	
<u>Age</u> <u>28-49</u> <u>Years</u>								
50-74 Years	26	-.08	.15	-.12	.08	.02	-.02	
	24	.34	.22	.72*	.26	.27	.36	
<u>All</u> <u>Targets</u>	50	.16	.20	.35*	.05	.12	.18	

Judge ratings were first compared with spouse ratings of personality. Table 3 shows the correlations between judge averages and spouse ratings for each of the five traits and for specific target groups. The correlations between spouse and judge ratings for all fifty targets was highest for Conscientiousness (.35) and lowest for Neuroticism (.05). The average correlation between judge and spouse ratings for all targets across each of the five traits was .18.

On average, Veteran spouses scored closer to judge ratings than Newlywed spouses (.25 vs. .08). Males' ratings of their wives were closer to judge ratings on average than were females' ratings of their husbands (.22 vs. .11). Also, older targets' spouse ratings corresponded more closely to judge ratings than they did with younger targets (.36 vs. -.02).

Self/Judge correlations

Table 4		<i>Correlations between self ratings and stranger averages</i>					* p < .05	
<u>Group</u> <u>Criteria</u>	<u>n</u>	<u>Extroversion</u>	<u>Agreeableness</u>	<u>Conscientiousness</u>	<u>Neuroticism</u>	<u>Culture</u>	<u>Average</u>	
<u>Gender</u> <u>Male</u> <u>Targets</u>								
Female Targets	25	.20	.08	.56*	-.07	-.19	.10	
	25	-.12	.19	-.05	.12	-.25	-.02	
<u>Age</u> <u>28-49</u>								

Years								
50-74	26	.06	.24	.23	.04	-.39	-.06	
Years	24	.25	.29	.38	.05	-.03	.19	
All								
Targets	50	.16	.12	.32	.03	-.22	.08	

Judge ratings were next compared with self ratings of personality. Table 4 shows the correlations between judge averages and self ratings for each of the five traits and for specific target groups. The correlation between self and judge ratings for all fifty targets was highest for Conscientiousness (.32) and lowest for Culture (-.22). The average correlation between self and judge ratings for all targets across each of the five traits was .08.

On average, males' self ratings were closer to judge ratings than were females' self ratings (.10 vs. -.02). Lastly, as predicted, older targets' self ratings correlated more closely with judge ratings than they did with younger targets (.18 vs. -.06). However, due to the low correlations and limited sample size, this difference was not significant at the five percent level.

Discussion

Operational Definitions

The accuracy of first impressions was defined within this design as the correlation of self and stranger ratings of personality. This was done by making the following four assumptions: (1) the personality inventory used can measure personality (2) subject responses were honest (3) self ratings were an accurate measure of personality (4) facial photographs could be isolated and function as the influence of facial appearance on first impressions. Each of these assumptions will be discussed in turn.

1. The first assumption made is that the personality questionnaire yields an accurate measure of personality, and can therefore accurately record self-views and first impressions. As covered in the theoretical framework section, the five factor model of personality has been extensively verified as a comprehensive model of personality. However, a problem lies in the unknown reliability of self ratings. If it were true that another group of self-ratings taken a week after the experiment only correlated .50 with the first self-ratings, then the strength of the measured accuracy of impressions would have to be reconsidered.

2. It can be assumed that all questionnaire responses are honest because they were taken anonymously. Demand characteristics (c. f. Orne, 1962) may have also caused the judges to not report their honest impressions, but this seems unlikely. It is possible that judges attempted to guess the self ratings of the targets rather than reporting their own first impressions. This could happen if the judges guessed the procedure of the entire experiment and attempted to meet the experiment's expectations. Not only does it seem unlikely that this happened, but trying to guess self views of personality may actually be a natural method of

forming impressions.

3. Even sincere target responses may not reflect true personality traits. However, as earlier mentioned in the theoretical framework, self-views are just as good a measure of personality as any clinical test. It can therefore be concluded that self-reports of personality are accurate.

4. Because of the absence prospect of interaction and the artificial setting, the elicited judgments may not be representative of real life impressions. Because facial appearance is relied upon in first impressions, it was assumed that this aspect could be isolated and would function as a realistic stimulus for first impressions. However, the effects of targets maintaining neutral expressions while being photographed remain unclear. Asking subjects to give a neutral expression may have caused them to compensate from common facial expressions that would otherwise naturally act as an influence upon the first impressions.

These four assumptions define the personality of the targets as their questionnaire responses, and the impressions as the judges' questionnaire responses. Once this is accepted, the correlations between the targets' self ratings and the judge averages are a measure of the accuracy of first impressions from facial appearance. Even if accepted, the experimental conditions and operational definitions of this experiment do need to be kept in mind.

Representativeness

The chosen targets and judges were not representative of the general population. All targets were kind enough to volunteer to participate in this experiment, and all judges were students in an introductory psychology course at Macalester College. Taking the average of judge ratings has its limits. Some judges were undoubtedly more or less accurate than the average ratings, as the low reliability scores indicates.

Reliability scores

The reliability scores obtained for the judge ratings were relatively low in comparison with previous research. Low reliability scores indicate that there was a high variance of judge scores within each target. A low amount of consensus helps to point out that different judges can view the same people differently. As covered in the review of judge characteristics, there are several reasons for the individual differences among judges. Among these differences are gender and cultural effects, as well as individual biases, beliefs, expectations, and moods.

Self/Spouse Correlations

Table 2 shows that a definite relationship existed between self and spouse ratings of personality. That spouses know each other better than chance should come as no surprise. However, the overall average of the accuracy was not that great (.44), indicating that knowledge of one's spouse is far from perfect.

The average correlations for Newlywed and Veteran couples are very similar. It can be concluded that newlyweds knew their spouses just as well as couples that had been married for longer periods of time. A possible reason for this result stems from the idea that personality changes over time. This would cause people to make adjustments of their impressions of acquaintances over time. People could not increase their knowledge of their spouses' personality, but could only hope to maintain the same level of knowledge.

The average correlations shown in Table 2 were higher for male targets than female targets. This suggests that females know their husbands better than their husbands know them. The source of this difference can be accounted for by two possible factors. If females were better judges of character, they would know their husbands' personalities better than their husbands knew theirs. Also, in a marriage, females could be more difficult to know than males. A combination of these two forces could account for the found gender differences.

Judge Correlations

Tables 3 and 4 show that the relationship between judge ratings of personality and spouse or self ratings was minimal. Only very few of these correlations were significant. Considering how many correlations were measured, such a number can be expected to be significant solely by chance. There are many possible reasons for a low accuracy of judge ratings of personality. For example, judge subjects may have been poorly motivated to make the effort to form accurate impressions. The operational definitions may not be valid. Also, the judgments of strangers from facial appearance, however well thought out, may be just plain wrong. However, except for self/judge correlations for the trait Culture, most correlations were positive. This suggests that there is a definite, but very slight, accuracy of first impressions from facial appearance.

This accuracy can have several explanations. There may be a biological link between the genes that influence personality and the genes that shape the face. A more environmental reason may exist in the form of a self-fulfilling prophecy. If personality is influenced by the expectations of society, expectations that are influenced by a shared interpretation of facial appearance, then personality would develop according to facial appearance. A third possibility is that the face is a sign of personality because it becomes molded by character expressions over time.

Gender differences from judge accuracy are unclear. Judges were more accurate in predicting the spouse ratings of female targets, but more accurate in predicting self ratings of male targets. Also, judges were more accurate in predicting Veteran spouse ratings of personality than Newlywed spouse ratings. Because Newlywed and Veteran spouse ratings were equally accurate in predicting self ratings, the reasons for this difference remain unclear. One possibility is that these differences are related to the effects from the differences between older and younger targets.

Accuracy differences between older and younger targets

In comparison to both spouse and self ratings of personality, judge ratings were closer to the ratings of older targets than younger targets. However, because overall accuracy is very minimal, these differences are not significant at the five percent level. Nevertheless, the higher scores for older targets suggest that older targets are judged more accurately from facial appearance than are younger targets.

These differences may exist for many reasons. There may be more accurate stereotypes for older people. Older targets may also have rated themselves more according to how others view them. Because social influences on personality have worked longer on older targets, older targets may be more similar to their judged appearances than younger targets. It is also possible that younger targets were not as accurate in rating their own personalities. Any of these conditions would have caused the higher accuracy for older targets.

These results of a higher accuracy for older targets also support the hypothesis that the face shows personality because it becomes molded by expressions of character over time. It was hypothesized that the face becomes influenced by the expression of personality over time. If this is true, it was reasoned that the accuracy of judgments would increase with the age of the targets. This was predicted before the experiment was run.

The results of this experiment support the hypothesis that it was designed to test. However, the overall accuracy of first impressions from facial impressions was very small, so the results need to be viewed in this light. Yet it does appear that the accuracy of facial impressions may be due to "different persons bringing into frequent use different facial muscles, according to their dispositions; the development of these muscles being perhaps thus increased, and the lines or furrows on the face, due to their habitual contraction, being thus rendered more conspicuous." (Darwin, 1872, p . 364365).

Future Research

As indicated by the low reliability scores, there are definite individual differences in rating styles among judges. One possible area of research could examine how the individual differences of the judges relate to their personality judgments and accuracy. Also, because of the differences in target ratings, another area of research could examine how individual differences among targets influence personality judgements and their accuracy. As previously summarized, the isolation of traits such as attractiveness and babyfacedness have already shown them to be factors in impression formation.

Because the research hypothesis has been supported by the data from this experiment, more research in this area is needed. Due to the low overall accuracy of first impressions, the found differences found between target groups were not significant. A greater sample size could help to validate and replicate these findings. Another idea is the to modify the experimental conditions and operational definitions. In order to make sure that the elicited

judgments represent the true to life conditions, the personality questionnaire used must yield self ratings that are accurate and properly represent the judges' impressions. Other methods of parsing out these two views are possible. For example, Bond, Berry, & Omar (1992) used targets' willingness to volunteer to take place in a deceptive experiment as a measure of honesty. Using other such definitions in research can help to validate these findings to those who remain skeptical concerning the chosen operational definitions.

References

- Albright, L., Kenny, D. A., & Malloy, T. R. (1988). Consensus in personality judgments at zero acquaintance. Journal of Personality and Social Psychology, *55*, 387-395.
- Alley, T. R. (1988). Physiognomy and social perception. In T. R. Alley, (Ed.), Social and applied aspects of perceiving faces. Hillsdale, NJ: Erlbaum.
- Allport, G. W. (1937). Personality: A psychological interpretation. New York: Holt.
- Allport, G. W., & Odbert, H. S. (1936). Trait names: A psycholexical study. Psychological Monographs, *47*, (1, Whole No. 211).
- Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. Psychological Bulletin, *111*, 256-274.
- Berry, D. S. (1990). Taking people at face value: Evidence for the kernel of truth hypothesis. Social Cognition, *8*, 343-361.
- Berry, D. S. (1991). Accuracy in social perception: Contributions of facial and vocal information. Journal of Personality and Social Psychology, *61*, 298-307.
- Berry, D. S., & Bronlow, S. (1989). Were the physiognomists right? Personality correlates of facial babishness. Personality and Social Psychology Bulletin, *15*, 266-279.
- Berry, D. S., & McArthur, L. Z. (1986). Perceiving character in faces: The impact of age related craniofacial changes on social perception. Psychological Bulletin, *100*, 3-18.
- Bond, C. F., Jr., Berry, D. S., & Omar, A. (1992). The 'kernel of truth' in judgments of deception. Manuscript submitted for publication, Texas Christian University and Southern Methodist University.
- Brandt, A. (1980). Face reading: The persistence of physiognomy. Psychology Today, *12*, 90-96.

- Borkenau, P., & Liebler, A. (1992). Trait inferences: Sources of validity at zero acquaintance. Journal of Personality and Social Psychology, *62*, 645-657.
- Bronlow, S. (1992). Seeing is believing: Facial appearance, credibility, and attitude change. Journal of Nonverbal Behavior, *16*, 101-115.
- Burgoon, J., Woodall, W. G., & Buller, D. B. (1989). Nonverbal communication: The unspoken dialogue. New York: Harper & Row.
- Cattell, R. B. (1945). The principal trait clusters for describing personality. Psychological Bulletin, *42*, 129-161.
- Cattell, R. B. (1946). The description and measurement of personality. New York: World Book.
- Cronbach, L. J. (1955). Processes affecting scores on "understanding of others" and "assumed similarity." Psychological Bulletin, *52*, 177-193.
- Darwin, C. (1872). The expression of emotions in man and animals. New York: Philosophical Library.
- Digman, J. M., & Inouye, J. (1986). Further specifications of the five robust factors of personality. Journal of Personality and Social Psychology, *50*, 116-123.
- Dion, K. K., Berscheid, E., & Walster, E. (1972). What is beautiful is what is good. Journal of Personality and Social Psychology, *24*, 285-290.
- Dion, K. L., & Dion, K. K. (1987). Belief in a just world and physical attractiveness stereotyping. Journal of Personality and Social Psychology, *52*, 775-781.
- Fancher, R. E. (1990). Pioneers of psychology. New York: Norton.
- Feingold, A. (1991). Sex differences in the effects of similarity and physical attractiveness on oppositesex attraction. Basic and Applied Social Psychology, *12*, 357-367.
- Fiske, D. W. (1949). Consistency of the factorial structures of personality ratings from different sources. Journal of Abnormal and Social Psychology, *44*, 329-344.
- Forgas, J. P., & Bower, G. H. (1987). Mood effects on person perception judgments. Journal of Personality and Social Psychology, *53*, 5360.
- Funder, D. C. (1987). Errors and mistakes: Evaluating the accuracy of social judgment. Psychological Bulletin, *101*, 75-90.

Gage, N. & Cronbach, L. J. (1955). Conceptual and methodological problems in interpersonal perception. Psychological Review, 62, 411-422.

Hamid, P. N. (1968). Style of dress as a perceptual cue in impression formation. Perceptual and Motor Skills, 26, 904-906.

Hamid, P. N. (1972). Some effects of dress cues on observational accuracy, a perceptual estimate, and impression formation. The Journal of Social Psychology, 86, 279-289.

Hinton, P. R. (1993). The psychology of interpersonal perception. New York: Routledge.

John, O. P. & Robins, R. W. (1994). Accuracy and bias in self-perception: Individual differences in self-enhancement and the role of narcissism. Journal of Personality and Social Psychology, 66, 206-219.

Keating, C. F., Mazur, A., Segall, M. H., Cysneiros, P. G., Kilbride, J. E., Divale, W. T., Komin, S., Leahy, P., Thurman, B. & Wirsing, R. (1981). Culture and the perception of social dominance from facial expression. Journal of Personality and Social Psychology, 40, 615-626.

Kenny, D. A., Horner, C., Kashy, D. A., & Chu, L. (1992). Consensus at zero acquaintance: Replication, behavioral cues, and stability. Journal of Personality and Social Psychology, 62, 889-907.

Knacksted, G. (1991). Eye contact, gender, and personality judgments. The Journal of Social Psychology, 131, 303-304.

Liggett, J. (1974). The human face. New York: Stein and Day.

Mainwaring, M. (1980). 'Phys/Phren' - Why not take each other at face value. Smithsonian, 11, 193-212.

Malatesta, C. Z., Fiore, M. J., & Messina, J. J. (1987). Affect, personality, and facial expressive characteristics of older people. Psychology and Aging, 2, 64-69.

Mazur, A., Mazur, J., & Keating, C. (1984). Military rank attainment of a West Point class: Effects of cadets' physical features. American Journal of Sociology, 90, 125-150.

McCrae, R. R., & Costa, P. T. Jr. (1985). Updating Norman's "adequate taxonomy": Intelligence and personality dimensions in natural language and questionnaires. Journal of Personality and Social Psychology, 49, 110-121.

McCrae, R. R., & Costa, P. T. Jr. (1986). Clinical assessment can benefit from recent advances in personality psychology. American Psychologist, 41, 1001-1002.

- McCrae, R. R., & Costa, P. T. Jr. (1987). Validation of the Five-Factor model of personality across instruments and observers. Journal of Personality and Social Psychology, *52*, 81-90.
- McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. Journal of Personality, *60*, 175-215.
- Noller, P., Law, H., & Comrey, A. L. (1989). Cattell, Comrey, and Eysenck personality factors compared: More evidence for the five robust factors? Journal of Personality and Social Psychology, *53*, 775-782.
- Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. Journal of Abnormal and Social Psychology, *66*, 574-583.
- Norman, W. T. (1969). "To see ourselves as others see us!": Relations among self perceptions, peer-perceptions, and expected peer-perceptions of personality attributes. Multivariate Behavioral Research, *4*, 417-433.
- Norman, W. T., & Goldberg, L. R. (1966). Raters, ratees, and randomness in personality structure. Journal of Personality and Social Psychology, *4*, 681-691.
- Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. American Psychologist, *17*, 776-783.
- Osbourne, D., & Gilbert, D. (1992). The preoccupational hazards of social life. Journal of Personality and Social Psychology, *62*, 219-228.
- Passini, F. T., & Norman, W. T. (1966). A universal conception of personality structure? Journal of Personality and Social Psychology, *4*, 44-49.
- Paunonen, S. V. (1991). On the accuracy of ratings of personality by strangers. Journal of Personality and Social Psychology, *61*, 471-477.
- Plutchik, R. (1980). Emotion: A psychoevolutionary synthesis. New York: Harper & Row.
- Satrapa, A., Melhado, M. B., Curado-Coelho, M. M., Otta, E., Taubemblatt, R., & Fayetti Siqueria, W. (1992). Influence of style of dress on formation of first impressions. Perceptual and Motor Skills, *74*, 159-162.
- Schneider, D. J., Hastorf, A. H., & Ellsworth, P. C. (1979). Person perception. Reading, MA: AddisonWesley.

- Shrauger, J. S., & Osberg, T. M. (1981).** The relative accuracy of selfpredictions and judgments in psychological assessment. *Psychological Bulletin*, 90, 322351.
- Snyder, M., Tanke, E. D., & Berschied, E. (1977).** Social perception and social behavior: On the self-fulfilling nature of social stereotypes. *Journal of Personality and Social Psychology*, 35, 656-666.
- Street, B. (1990).** Reading faces: Physiognomy then and now. *Anthropology Today*, 6, 11-12.
- Tesser, A., & Collins, J. (1988).** Emotion in social reflection and comparison situations: Intuitive, systematic, and exploratory approaches. *Journal of Personality and Social Psychology*, 55, 695709.
- Tesser, A., Pilkinton, C. J., & McIntosh, W. D. (1989).** Selfevaluation maintenance and the mediational role of emotion: The perception of friends and strangers. *Journal of Personality and Social Psychology*, 57, 442456.
- Thornton, G. R. (1944).** The effect of wearing glasses upon judgments of personality traits of persons seen briefly. *Journal of Applied Psychology*, 28, 203-207.
- Toner, H. L., & Gates, G. R. (1985).** Emotional traits and recognition of facial expressions of emotion. *Journal of Nonverbal Behavior*, 9, 48-66.
- Tupes, E. C., & Christal, R. E. (1961/1992).** Recurrent personality factors based on trait ratings. *Journal of Personality*, 92, 225-251.
- Watson, D. (1989).** Strangers' ratings of the five robust personality factors: Evidence of a surprising convergence with selfreport. *Journal of Personality and Social Psychology*, 57, 120-128.
- Wiggins, J. S. (1973).** *Personality and prediction: Principles of personality assessment*. Reading, MA: Addison-Wesley.
- Wogalter, M. S., & Hoise, J. A. (1990).** Effects of cranial and facial hair on perceptions of age and person. *The Journal of Social Psychology*, 131, 589-591.
- Workman, J. E., & Johnson, K. K. P. (1991).** The role of cosmetics in impression formation. *Clothing and Textiles Research Journal*, 10, 63-67.
- Zebrowitz, L. (1990).** *Social perception*. Pacific Grove, CA: Brooks/Cole.

This report was written by John Hubbard for the Macalester College course "PSYCH 96: Independent Project" in December 1994. It has not been significantly altered from

the original version; the last modified date shown below indicates when this Webpage was last uploaded in its present form.