

Trait Inferences: Sources of Validity at Zero Acquaintance

Peter Borkenau and Anette Liebler
Universität Bielefeld
Bielefeld, Germany

Consensus between self-ratings and stranger ratings of personality traits was investigated. A sample of 100 adults was videotaped while entering and walking through a room, sitting down, looking into the camera, and reading a standard text. The targets then provided self-descriptions on 5 personality factors. A sample of 24 strangers who had never seen the targets before was given 1 of 4 types of information on the targets: (a) sound-film, (b) silent film, (c) still, or (d) audiotape. Strangers rated various physical attributes and 20 traits of each target. Level of information influenced the validity but not the reliability of the stranger ratings, which were most valid for extraversion and conscientiousness. Extraversion covaried most strongly with physical attributes, and implicit theories on the covariation of traits with physical attributes were more accurate for extraversion and conscientiousness than for agreeableness, emotional stability, and culture.

If judges indicate their impressions of other people, their judgments usually correlate with the self-reports of the ratees. Moreover, the correlations between self-ratings and ratings by judges tend to be higher the more the judges know about the targets (Cloyd, 1977; Funder & Colvin, 1988; Norman & Goldberg, 1966; Paunonen, 1989). The highest correlations are usually found between self-reports and ratings by spouses. McCrae and Costa (1989) report correlations between self-ratings and spouse ratings from .53 to .60, whereas they report somewhat lower correlations, ranging from .32 to .54, between self-reports and ratings by a single peer. These findings indicate that (a) personality ratings reflect actual attributes of the target persons rather than illusions that exist only in the eye of the beholder, and (b) self-ratings and ratings by others agree more the more that both reflect shared information concerning the targets' personality.

Consensus at Zero Acquaintance

This raises the question of how much information is required in order that ratings by others possess some validity. We use the term *valid* to indicate that ratings by strangers or acquaintances correlate substantially with the targets' self-reports. At first glance, this is a curious usage of the idea of validity because self-reports are usually more suspect than observer ratings and are therefore frequently validated against observer ratings. But we have a special case here: Because the observers are strangers, the self-ratings are the more dependable measure of the actual personality of the ratees.

There is evidence from several sources that very meager information is sufficient for judges to provide ratings of targets that are more valid than random judgments. Cleeton and Knight (1924), for example, reported substantial agreement not only among ratings by casual observers, but also between ratings by casual observers and those by close acquaintances of the ratees. Hunt and Lin (1967) reported better-than-chance agreement between self-reports and judgments by strangers, although the strangers listened to an audiotape of the targets' utterances only and thus inferred personality traits from speech alone. The validity of the stranger ratings thereby differed for various trait dimensions; higher validities were obtained for dimensions like forceful-gentle and assertive-reserved than for dimensions like orderly-casual.

Passini and Norman (1966) also found substantial correlations of ratings by strangers with self-reports. These authors let university freshmen who had no information about each other except having been together in the same room for about 15 min judge one another on 20 scales that were measures of the Big Five (Norman, 1963). A peer nomination technique was used. The correlations were highest, surpassing .30, for Extraversion and Conscientiousness, whereas they were lowest for Emotional Stability; the validities of the stranger ratings for Agreeableness and Culture¹ were between the others (Norman & Goldberg, 1966). Similar findings have also been reported more recently by Albright, Kenny, and Malloy (1988) and by Watson (1989). Albright et al. coined the term *consensus at zero acquaintance* to refer to findings like those reported by Norman and Goldberg (1966).

Thus it is likely that some physical attributes of persons provide cues that enable even strangers to accurately infer some

The research reported in this article was supported by a grant from the German Science Foundation to Peter Borkenau (DFG-Az. Bo 774/5-1).

We are indebted to Alois Angleitner, David M. Buss, John A. Johnson, Wolf Nowack, Rainer Riemann, and Enno Schwanenberg for their helpful comments on earlier drafts of this article.

Correspondence concerning this article should be addressed to Peter Borkenau, Abteilung für Psychologie der Universität Bielefeld, Postfach 8640, W-4800 Bielefeld, Germany.

¹ The appropriate labels of the five major personality factors are still subject to some controversy, and the label *Culture* for the fifth factor is not likely to be the most appropriate one (Digman, 1990). We use this label, however, because the last four trait dimensions that were used in this study had been selected by Norman to measure the factor called *Culture*.

aspects of the targets' personalities. More valid observable cues seem to exist for Extraversion and Conscientiousness than for Agreeableness, Culture, and Emotional Stability. Note that the higher validity of ratings for Extraversion and Conscientiousness is less pronounced if knowledgeable informants like spouses and peers provide the judgments (Amelang & Borkenau, 1982; McCrae & Costa, 1989; Norman & Goldberg, 1966). Thus a higher level of acquaintance is required to accurately infer Emotional Stability than to infer Extraversion and Conscientiousness. A corresponding interaction between trait observability, level of acquaintance, and the validity of ratings by others has recently been reported by Paunonen (1989) for the traits assessed by Jackson's Personality Research Form. Whereas it is clear now, however, that some personality traits are accurately inferred from physical cues alone, it has not been clarified which attributes in a person's physical appearance enable strangers to infer these traits accurately.

Reasonable Sources of Consensus at Zero Acquaintance

There is a large body of research concerning which personality traits are inferred from various physical attributes of strangers. Dion, Berscheid, and Walster (1972), for instance, identified a stereotype implying that "what is beautiful is good." Furthermore, Zuckerman and Driver (1989) demonstrated a "what sounds beautiful is good" phenomenon. Albright et al. (1988) found correlations beyond .70 between strangers' ratings of targets' physical attractiveness and their ratings of the targets' Extraversion. Moreover, they found similar correlations between strangers' ratings of the targets' formality and neatness of dress and their ratings of the targets' Conscientiousness.

Note, however, that such studies reveal only which attributes subjects rely on to infer strangers' personality traits. Unless these attributes correlate also with the targets' self-reports, they do not contribute to any agreement between self-ratings and stranger ratings. Moreover, there has to be a match between strangers' trait inferences from physical attributes and the relations between physical attributes and self-reports of personality. This is best illustrated with Brunswik's (1956) lens model. A somewhat modified version of the lens model that was adapted for the present purposes is depicted in Figure 1.

The center of the model refers to the physical attributes of the targets. Only those attributes that the strangers can perceive may mediate the correlation between self-reports and ratings by strangers. In Passini and Norman's (1966) study, for example, the judges could perceive the targets' visible attributes, such as their sex, stature, build, physiognomy, and neatness and manner of dress. Moreover, the targets' visible movements, that is, miming and gestures, could also be perceived. No acoustic information, however, was available to the judges. Thus, the term *physical attributes* is used in a very broad sense in this type of research. It refers to all attributes that can be perceived by strangers in the setting under study. It does not refer only to bodily attributes of the targets.

The right half of the lens model refers to the trait inferences by strangers. The single-headed arrows from the physical attributes to the strangers' trait ratings indicate that the strangers actually infer the traits from the physical variables. By contrast,

double-headed arrows are used in the left half of the model, which refers to the targets' self-ratings, to indicate that there are several likely sources of correlations between physical attributes and self-reports of personality: (a) Personality may be expressed through observable cues (Allport & Vernon, 1933); (b) physical variables, such as attractiveness and physique, may influence personality development; or (c) physical and personality variables may covary because of a third factor, for instance, a genetic factor. Thus the direction of effects cannot be specified. But it is only necessary to assume here that physical and personality attributes are correlated, as correlations between physical variables and self-reported personality traits are sufficient to explain the consensus between targets and strangers. For instance, if strangers infer level of Conscientiousness from the targets' neatness of dress, and if neatness of dress correlates with the targets' self-rated Conscientiousness, then neatness of dress may mediate the correlation between self-ratings and stranger ratings of Conscientiousness.

Earlier Research

We are aware of only one study in which a design like the one we use has been realized successfully (Gifford, Fan Ng, & Wilkinson, 1985), whereas numerous studies have focused on the right half of the lens model only (that is, on the relations between physical attributes and trait inferences by strangers). Moreover, there are some studies that have been focused on the two tails of the lens model, that is, on the consensus between targets and strangers concerning the targets' personality traits (Albright et al., 1988; Hunt & Lin, 1967; Kenny, Horner, Kashy, & Chu, 1992; Passini & Norman, 1966; Watson, 1989). But these studies were not investigations of how the validity of stranger ratings reflects the strangers' trait inferences from valid physical cues. Thus these studies cannot explain *why* ratings by strangers correlate with self-ratings to some extent.

Some other researchers have tried to explain the consensus between targets and strangers with valid cues in the targets' physical appearance, but they failed to identify the physical variables that actually mediate this consensus. Cleeton and Knight (1924) collected various measures of the targets' physiognomy but did not find any relations between these measures and trait ratings by strangers. Amelang, Köhler, and Gold (1983) investigated physical attractiveness as a reasonable mediating variable but found that individual differences in attractiveness did not mediate the consensus between targets and strangers.

In contrast, Gifford et al. (1985) considered the entire lens model and were successful in identifying variables that actually mediate the consensus between targets and strangers for one personality variable. These authors videotaped interviews with 34 applicants for a job as a research assistant and let strangers infer the applicants' motivation and social skill from their viewing of the videotapes with the sound turned off. They also obtained self-appraisals of the applicants' motivation and social skill. They found consensus between self-appraisals and stranger ratings to be $r = .09$ for motivation and $r = .29$ for social skill. Furthermore, these authors measured various nonverbal behaviors of the applicants. They found that the consensus between applicants and judges concerning social skill was me-

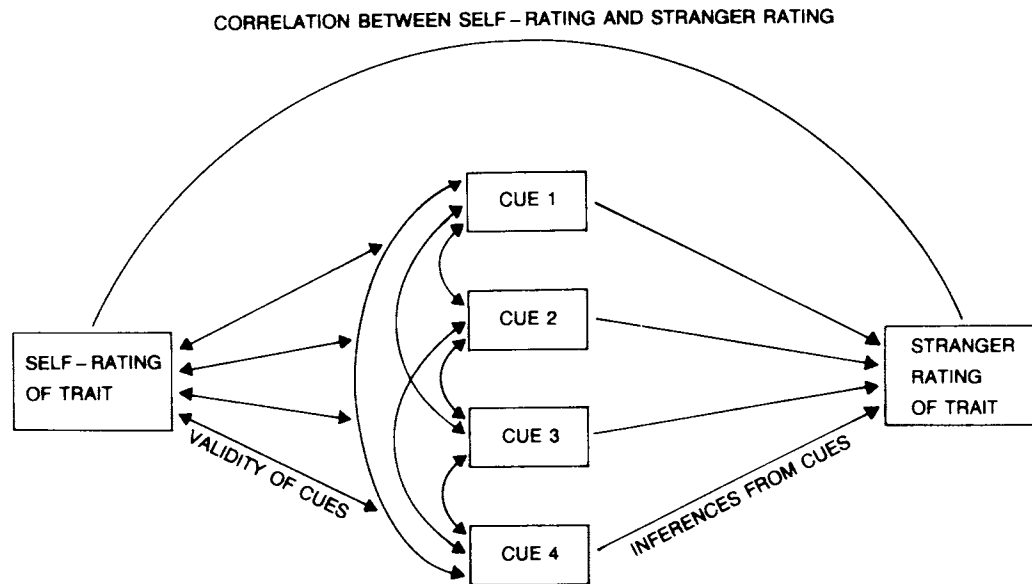


Figure 1. An adapted version of Brunswik's lens model.

diated by the applicants' gesturing, time spent talking, and formality of dress: These variables correlated with self-reported social skill as well as with the stranger ratings of social skill.

This Study

The present research differs from the study by Gifford et al. (1985) in several respects. First, we used a larger sample of targets ($N = 100$) to obtain more stable estimates of the various correlations. Second, we used a comprehensive system of personality traits, namely measures of the five major factors of personality (Digman, 1990). Third, we took several precautions to guarantee that the strangers really did not know the targets (Gifford et al. did not report how they precluded earlier contacts between judges and targets). Finally, following a suggestion by Amelang et al. (1983), we combined an experimental and a measurement approach to identify some mediating variables. The experimental approach consisted of randomly assigning the judges to four experimental conditions: Six judges saw the videotapes with the sound turned on (sound-film condition). Six other judges saw the same videotapes but with the sound turned off (silent film condition). Another six judges were presented only stills of the targets (still condition). Finally, the last six judges were presented an audiotape of the targets' utterances and were thus provided with acoustic information only (audiotape condition). The measurement approach consisted of having judges and targets indicate their impressions of various physical attributes of the targets on a specially constructed rating sheet.

Method

Subjects

The rationale of the present study made it mandatory that the judges had not experienced any prior personal contact with the targets. The

targets and judges were therefore recruited at different places: The targets were recruited outside and the judges inside the University of Bielefeld. Applicants were not accepted as targets if they had studied at this university. The mean age of the 100 accepted targets (50 women and 50 men) was 26.32 years ($SD = 8.01$ years). The youngest target was 17, and the oldest target was 54. Most targets were either high school students, students at a technical college in Bielefeld, Germany, or housewives who responded to an announcement in a supermarket.

The judges were recruited by means of flyers distributed around the university. They were paid for their participation. Of those who were interested in participating, 24 (13 men and 11 women) were selected who were all students at the University of Bielefeld but had not gone to school in this city. Bielefeld has a population of over 300,000, and the university is situated somewhat outside the city. Thus it was quite unlikely that the judges had ever met the targets before. Nevertheless, when watching or listening to the tapes, the judges were asked whether they had ever seen a target before. In 11 cases (among 2,400 judge-target combinations) judges reported that they had seen targets, and the judges' ratings for those targets were then excluded from further analyses.

Measures

Judgments of the targets' personality traits as well as of their physical attributes were provided by the strangers and by the targets themselves on similar instruments.

Personality measures. All personality measures were constructed within the framework of the five-factor model of personality. The targets described themselves, and were described by the strangers, on scales that were German translations of Norman's (1963) 20 marker scales for the five factors. Thus each of these factors was referred to by 4 rating scales. The two poles of each scale were labeled with two opposite trait-descriptive terms (e.g., *talkative* vs. *silent*), and a 7-point rating scale was provided for each pair of opposite traits to indicate the judgment. High scores were assigned to subjects who described themselves as extraverted, agreeable, conscientious, emotionally stable, and cultured.

Moreover, the targets were administered the 60-item NEO Five-Fac-

tor Inventory (NEO-FFI; Costa & McCrae, 1989). The NEO-FFI is a short version of the NEO Personality Inventory (Costa & McCrae, 1985). Twelve of its items refer to each of the factors Neuroticism, Extraversion, Openness to Experience, Agreeableness, and Conscientiousness. The appropriate labels of the five factors are still subject to some controversy (Digman, 1990). For reasons of convenience, however, we assume that Norman's (1963) factor Culture and Costa and McCrae's (1989) factor Openness to Experience refer to highly similar constructs, whereas Norman's factor Emotional Stability is the opposite of Neuroticism. This view is justified by findings reported elsewhere (Borkenau & Ostendorf, 1990).

Measures of physical attributes. The target persons were videotaped while entering a room, walking around a table toward a chair behind the table, sitting down, looking into the camera, reading a standard text (actually a weather forecast), standing up, and walking around the table again to leave the room. This procedure took about 90 s per target. The rating sheet that referred to the targets' physical appearance comprised 45 attributes that could probably be perceived in this setting.

The 45 attributes fell into four major classes. Category 1 comprised overall impression variables, namely the estimated age of the target and whether he or she was more masculine versus feminine, unsympathetic versus sympathetic, and unattractive versus attractive. Category 2 comprised nine attributes that could be inferred from acoustic information only. Category 3 comprised 24 attributes that could be inferred from the static visual information that was provided in the still condition. Finally, Category 4 comprised eight attributes that could be inferred from the dynamic visual information that was only provided in the sound-film condition and the silent film condition. All 45 physical attributes are listed in Table 1. They were chosen on the basis of a review of the research literature on the physical attributes that influence trait attributions by strangers.

Five versions of this physical attributes rating sheet were compiled. Two versions asked for judgments concerning all 45 attributes and differed in the instructions only; one was filled out by the targets themselves to describe their own physical attributes and the other one was filled out by the six judges who saw the sound-film. The other three versions comprised only a subset of the 45 attributes. The six judges in the silent film condition had to report the 36 attributes in Categories 1, 3, and 4, the six judges in the still condition the 28 attributes in Categories 1 and 3, and the six judges in the audiotape condition the 13 attributes in Categories 1 and 2. All ratings of physical attributes were provided on 7-point bipolar rating scales. The two poles of each scale were labeled with opposite terms (e.g., *dark hair* vs. *light hair*).

Procedure

Before being videotaped, the targets were shown the room where they would be filmed and were instructed what to do next. After the videotaping, they were led to an adjacent room and were administered the physical attributes rating sheet, Norman's (1963) 20 rating scales, and the NEO-FFI. They were then paid for their participation and excused.

From the videotapes thus produced, two additional tapes were derived. A 60-s still of each target person looking into the camera was produced and rerecorded onto another videotape. This tape was presented to the six judges in the still condition. Finally, the sound from the original videotape was rerecorded onto an audiotape that was presented to the six judges in the audiotape condition. The original videotape was watched by the judges in the sound-film condition with the sound turned on and by the judges in the silent film condition with the sound turned off.

The judges sat alone in a room with either a tape recorder (in the

audiotape condition) or with a video recorder and a monitor (in the other three conditions). After having seen the information about one target, they stopped the recorder and filled out the rating sheets that referred to the physical attributes and the traits of that particular target. Then they restarted the recorder to view or listen to the next target, and so on, until each stranger had provided ratings of all 100 targets. This rating task took several days per judge.

Results

Reliabilities

Reliability of self-ratings. The internal consistencies of the NEO-FFI scales were .86 for Neuroticism, .81 for Extraversion, .69 for Openness to Experience, .71 for Agreeableness, and .81 for Conscientiousness. Moreover, the internal consistencies of the responses to Norman's (1963) four marker scales for each of the five factors were .78 (Extraversion), .67 (Agreeableness), .60 (Conscientiousness), .54 (Emotional Stability), and .27 (Culture). Finally, the correlations between the NEO-FFI scales and the composite ratings on Norman's four marker scales for the same factor were .77 (Extraversion), .45 (Agreeableness), .59 (Conscientiousness), -.59 (Neuroticism vs. Emotional Stability), and .35 (Openness to Experience–Culture). Thus, all internal consistency estimates were lowest for the Openness to Experience–Culture factor. A similar phenomenon has previously been reported by Cheek (1982).

Reliability of stranger ratings. Concerning the physical attributes, the rater agreement was separately estimated for the six judges in each of the four experimental conditions but did not systematically differ across these conditions. We therefore report only the reliabilities (coefficient alpha) in the sound-film condition in the first data column of Table 1. The consensus tended to be higher for the general impression variables (mean $\alpha = .81$) and the static visual attributes (mean $\alpha = .75$) than for the dynamic visual attributes (mean $\alpha = .63$) and the audible attributes (mean $\alpha = .63$). Note, however, that some variables, such as feminine impression, high voice, and made-up face, reflect mainly sex differences and would probably be less reliably estimated within each sex.

There was an interaction between experimental condition and personality domain concerning the consensus among the strangers for the targets' personality traits. Thus the stranger agreement is reported in Table 2 separately for the four conditions. The entries in the rows labeled *mean Extraversion* and so forth refer to the agreement among the six judges concerning the mean position of each target on the four scales that mark that particular personality factor.

The strangers' consensus tended to be lowest in the still condition, and this reflected the extremely low consensus in this condition concerning aspects of Emotional Stability. There are five rows in Table 2 that refer to aspects of Emotional Stability, and in each of these rows the lowest alpha is found in the column that refers to the still condition. The probability of such a finding occurring purely by chance is 1 in 1,024. Moreover, each of the five alphas that refer to aspects of Emotional Stability is lower than any other alpha in the still condition. The probability of such a finding occurring purely by chance is 1 in 53,130. Thus the agreement among the strangers was particularly low if aspects of Emotional Stability had to be inferred

Table 1
*Reliability (Alpha) of Stranger Ratings of Physical Attributes in the Sound-Film Condition
 and Correlations Between Self- and Stranger Ratings*

Attribute impression	Consensus among strangers	Consensus of targets and strangers			
		Sound- film	Silent film	Still	Audio- tape
General impression					
Estimated age of target	.96	.89	.74	.85	.82
Unsympathetic-sympathetic	.59	.03	.01	.10	.09
Unattractive-attractive	.79	.33	.07	.22	.29
Masculine-feminine impression	.90	.75	.77	.82	.79
Audible cues					
Effortless-effortful reading	.19	.18			.13
Soft-voiced-loud-voiced	.67	.15			.15
Deep-high voice	.80	.40			.45
Weak-powerful voice	.72	.10			.20
Unpleasant-pleasant voice	.58	.21			.23
Haltingly-fluently speaking	.77	.44			.44
Difficult-easy to understand	.62	.04			.17
Calm-hectic speaking	.57	.19			.29
Strong dialect-standard language	.76	.35			.30
Static visible cues					
Unrefined-refined appearance	.78	.36	.33	.30	
Not made-up-made-up face	.89	.75	.74	.72	
Light-dark garments	.92	.67	.70	.61	
Plain-showy dress	.77	.43	.47	.31	
Formal-informal dress	.71	.41	.43	.35	
Unfashionable-fashionable dress	.77	.38	.25	.34	
Short-long hair	.95	.85	.83	.76	
Unstylish-stylish hair	.78	.50	.57	.44	
Dark-light hair	.94	.76	.73	.75	
Slim-stout physique	.91	.72	.74	.59	
Short-tall stature	.88	.70	.66	.40	
Less-more muscular physique	.61	.30	.32	.30	
Ill-well-proportioned body	.68	.13	.15	.09	
Thin-round face	.85	.60	.58	.46	
Thin-full lips	.50	.34	.44	.34	
Childlike-mature face	.73	.38	.50	.40	
Soft-hard facial lineaments	.55	.31	.37	.28	
Grumpy-friendly expression	.80	.33	.33	.17	
Serious-unconcerned expression	.52	.04	.10	.08	
Timid-self-assured expression	.66	.22	.12	.17	
No smiling-extensive smiling	.89	.40	.42	.10	
Controlled-relaxed sitting	.54	.34	.20	.19	
Open-closed arms while sitting	.64	.38	.36	.23	
Avoided-looked into the camera	.78	.39	.40	.19	
Dynamic visible cues					
Fast-slow movements	.76	.17	.16		
Infrequent-frequent hand movements	.65	.22	.18		
Infrequent-frequent head movements	.71	.28	.22		
Touched own body rarely-frequently	.46	.31	.36		
Dragged-lifted feet while walking	.65	.43	.32		
Degree of arm swinging while walking	.53	.00	.14		
Small steps-long strides	.69	.35	.25		
Relaxed-stiff walking	.57	.22	.11		

Note. Correlations beyond .19 are statistically significant ($p < .05$).

from static visual cues alone. But there was no systematic relation between level of information and the consensus among the strangers. In particular, the consensus among strangers was not systematically lower in the audiotape condition or the silent film condition than in the videotape condition. Thus level of information was generally unimportant for the consensus among strangers.

Level of Consensus Between Targets and Strangers

The correlations between the ratings of physical variables by targets and strangers are reported in the second to fifth data columns of Table 1. Not surprisingly, the consensus was highest concerning the age and masculinity-femininity of the targets. This indicates that the judges filled out the rating sheets seri-

Table 2
Consensus (Cronbach's Alpha) Among Strangers Concerning the Targets' Personality Traits (Separately for the Four Experimental Conditions)

Trait dimension	Sound-film	Silent film	Still	Audio-tape
Talkative-silent	.76	.69	.67	.68
Frank-secretive	.72	.71	.72	.68
Adventurous-cautious	.75	.76	.68	.68
Sociable-reclusive	.77	.71	.65	.52
Mean Extraversion	.81	.82	.78	.73
Good-natured-irritable	.56	.61	.44	.51
Not jealous-jealous	.50	.62	.60	.59
Gentle-headstrong	.59	.56	.45	.56
Cooperative-negativistic	.52	.51	.53	.49
Mean Agreeableness	.62	.67	.58	.59
Fussy-careless	.73	.72	.71	.76
Responsible-undependable	.66	.61	.61	.70
Exact-ing-unexact-ing	.72	.67	.78	.76
Persevering-quitting	.63	.62	.61	.64
Mean Conscientiousness	.72	.77	.77	.73
Poised-nervous	.49	.46	.21	.70
Calm-anxious	.67	.65	.30	.71
Composed-excitable	.10	.42	.03	.59
Not hypochondriacal-hypochondriacal	.53	.68	.22	.68
Mean Emotional Stability	.65	.63	.19	.76
Artistically sensitive-insensitive	.50	.59	.49	.57
Intellectual-narrow	.58	.56	.60	.70
Polished-crude	.57	.75	.72	.72
Imaginative-simple	.65	-.09	.62	.45
Mean Culture	.66	.71	.67	.73

ously. Moreover, the correlations between self-ratings and stranger ratings tended to be higher for the static visual attributes (average correlation .42) than for the dynamic visual attributes (average correlation .23) and the audible attributes (average correlation .25). Finally, the correlations tended to be lower for the static visual attributes in the still condition than for the static visual attributes in the sound-film and the silent film conditions. There were 24 static visual attributes, and for 18 of them the consensus between targets and strangers was lowest in the still condition, whereas it was lowest for three of them in the sound-film and the silent film condition, respectively, $\chi^2(2, N = 100) = 18.75, p < .001$. This reduced consensus in the still condition may reflect that the judges saw only portraits of the targets here, whereas they saw the targets walking, sitting, and gesturing in the other two conditions.

The correlations between the ratings of personality traits by targets and strangers are reported in Table 3² for the entire sample, and for female and male targets. The rows labeled *mean Extraversion*, *mean Agreeableness*, and so forth refer to the correlation between the mean of the self-ratings on the four scales that mark that particular factor and the mean of the stranger ratings on the four scales that mark that particular factor. The latter score is the composite of 24 raw scores, as it is the aggregate of the ratings on four scales by six independent

judges. Furthermore, the rows labeled *NEO-FFI Extraversion*, *NEO-FFI Agreeableness*, and so forth refer to the correlations between the NEO-FFI scales and the composite of the 24 stranger ratings for that particular personality domain mentioned earlier.

The findings reported in Table 3 by and large replicate earlier findings that strangers infer Extraversion and Conscientiousness most accurately and Emotional Stability least accurately from physical cues. The design of the present study, however, made it feasible to investigate also some sources of this now well-established phenomenon.

Sources of Consensus Between Targets and Strangers

Effects of age and sex. Because the target sample was heterogeneous in age and sex, we examined the extent to which the correlations in Table 3 might be explained by stereotype accuracy: Strangers might appropriately expect different trait levels among men and women as well as among old and young people, and these expectancies might account for those correlations in Table 3 that refer to the entire sample. If this were true, it would have been group membership rather than physical or behavioral information that accounted for the agreement between targets and strangers. But the within-sex correlations were not systematically lower than those for the mixed-sex group. Thus the contribution of sex differences to the agreement between targets and strangers was small.

To control also for any stereotype accuracy on the basis of age differences, we calculated second-order correlations with the effects of sex and age partialled out from the correlations between self-ratings and ratings by strangers. These partial correlations were quite similar to the zero-order correlations reported in Table 3. In particular, the consensus for traits in the Extraversion domain was not diminished at all. For Conscientiousness, the partial correlations were somewhat lower than the zero-order correlations, but 8 of the 24 partial correlations in the mixed sex group were still significant. Thus, only a marginal proportion of the consensus between targets and strangers was explained by the effects of age and sex. The further analyses therefore refer to the zero-order correlations, and age and sex will be treated as two possible mediating variables among numerous others.

Level of information. One of the most perplexing features of Table 3 is that even a still or an audiotape was sufficient for strangers to infer various traits with substantial accuracy. But at the same time, the expected relation between level of information and the validity of stranger ratings was also found. The ratings by strangers tended to be most valid in the sound-film condition. There are 30 rows in Table 3, and in 18 of these rows the highest correlation for the mixed sex group was found for the sound-film condition, in 6 rows for the silent film condi-

² The entries in Table 3 report correlations between self-reports and mean ratings by six strangers. The ratings by single strangers were obviously less reliable and therefore less strongly correlated with the targets' self-reports. For example, the mean correlations between the NEO-FFI scales and the ratings by single judges in the sound-film condition were .35 (Extraversion), .20 (Agreeableness), .17 (Conscientiousness), .05 (Neuroticism), and .13 (Openness to Experience).

Table 3
Correlations Between Self-Ratings and Stranger Ratings for Personality Traits

Trait dimension	Sound film			Silent film			Still			Audiotape		
	All targets	Women	Men	All targets	Women	Men	All targets	Women	Men	All targets	Women	Men
Talkative-silent	.30**	.22	.37**	.26**	.19	.32*	.28**	.12	.37**	.22**	.27*	.18
Frank-secretive	.23*	.16	.32*	.39**	.27*	.54**	.19	.14	.23	.25*	.12	.34*
Adventurous-cautious	.37**	.43**	.33*	.30**	.41**	.19	.31**	.46**	.14	.29**	.39**	.19
Sociable-reclusive	.38**	.24	.55**	.27**	.15	.46**	.20	.08	.31*	.24*	.26	.25
Mean Extraversion	.42**	.36**	.52**	.41**	.32*	.54**	.33*	.24	.40**	.36**	.34*	.37**
NEO-FFI Extraversion	.51**	.55**	.46**	.47**	.41**	.56**	.33**	.28*	.36**	.33**	.40**	.24
Good-natured-irritable	.08	.01	.14	.00	.03	-.02	.03	.01	.05	.01	-.12	.09
Not jealous-jealous	.05	-.02	.21	.09	.02	.21	-.06	-.11	.04	.07	.02	.12
Gentle-headstrong	.18	.12	.22	.06	-.01	.09	.04	.07	.01	.07	.09	.04
Cooperative-negative	.19	.22	.13	.01	-.12	.15	.14	.14	.12	.08	.07	.08
Mean Agreeableness	.13	.05	.21	.06	-.05	.17	.06	.01	.09	.07	.02	.12
NEO-FFI Agreeableness	.35**	.26	.33*	.17	.09	.17	.19	.24	.10	.21*	.19	.21
Fussy-careless	.33**	.23	.25	.28**	.19	.21	.24*	.29*	.11	.21*	.25	.06
Responsible-undependable	.10	.00	.10	.18	.14	.17	.27**	.32*	.20	.00	-.01	-.08
Exacting-unexacting	.23*	.37**	.18	.21*	.12	.34*	.21*	.35*	.11	.07	.24	-.09
Persevering-quitting	.12	.12	.21	.24*	.37**	.09	.14	.08	.24	.06	.18	.04
Mean Conscientiousness	.25*	.28*	.25	.33**	.35*	.32*	.27*	.37**	.17	.09	.14	.02
NEO-FFI Conscientiousness	.25*	.25	.21	.25*	.20	.25	.32**	.31*	.29*	.10	.05	.08
Poised-nervous	.19	.09	.32*	.09	.11	.06	-.08	-.25	.17	.08	.15	.00
Calm-anxious	.21*	.16	.13	.16	-.02	.24	-.05	-.20	.04	.18	.21	.04
Composed-excitable	-.08	-.06	-.14	-.01	.02	-.18	.02	-.21	.13	.05	-.05	.02
Not hypochondriacal-hypochondriacal	-.10	-.48**	.23	-.04	-.25	.13	.00	-.10	.09	-.04	-.11	.04
Mean Emotional Stability	.04	-.09	.05	.02	-.15	.05	-.18	-.39**	.06	.04	.04	-.07
NEO-FFI Neuroticism ^a	.10	-.01	.09	.27**	.08	.35**	.06	-.14	.32*	.14	.09	.08
Artistically sensitive-insensitive	.15	.11	.18	.04	-.03	.05	.07	.15	-.05	.12	-.04	.22
Intellectual-narrow	.08	.18	.02	.10	.32*	-.05	.13	.05	.18	-.04	.27*	-.23
Polished-crude	.38**	.19	.37**	.44**	.48**	.30*	.44**	.48**	.30*	.19	.36**	-.05
Imaginative-simple	.15	.05	.30*	-.18	-.16	-.23	.06	.02	.15	.05	.15	-.03
Mean Culture	.12	.01	.12	.05	.09	-.11	.10	.14	-.03	.08	.09	.01
NEO-FFI Openness to Experience	.20*	.25	.31*	-.07	.03	.02	-.08	.10	-.05	.14	.01	.35

Note. NEO-FFI = NEO Five-Factor Inventory.

^a The sign of these correlations was reversed.

* $p < .05$. ** $p < .01$.

tion, in 4 rows for the still condition, and in 2 rows for the audiotape condition, $\chi^2(3, N = 100) = 20.67, p < .001$. This indicates that the static visual, dynamic visual, and audible cues usually contributed valid and nonredundant information and were therefore more informative together than the cues in each single category alone. An exception to this rule was Conscientiousness: The stranger ratings were about equally valid in the sound-film, the silent film, and the still condition, but much less valid in the audiotape condition. For the mixed sex group, the validity of the stranger ratings was always lowest in the audiotape condition, and the likelihood that this finding occurred purely by chance is 1 in 4,096. This indicates that the acoustic cues conveyed hardly any information on Conscientiousness that was appropriately used by the judges.

The effects of level of information on the accuracy of trait inferences (Table 3) should be compared with its effects on the consensus among the strangers (Table 2). The strangers did not agree more among each other in the sound-film condition than in the silent film or the audiotape condition. Thus, if the judges received less information, this did not yield less similar trait inferences. Rather, if some valid cues were withheld, the strangers provided equally similar but less valid judgments. This finding is predicted by Kenny's (1991) general model of consensus and accuracy in interpersonal perception.

Mediating Physical Attributes

Self-perceived and stranger-perceived physical attributes. To identify the physical cues from which the strangers accurately inferred the personality traits of the targets, we correlated the physical attributes, as judged by the strangers, with the self-ratings and the stranger ratings of the traits. The stranger ratings of physical attributes, rather than the self-ratings of these attributes, were used because a causal effect from physical variables to trait inferences by strangers is assumed in the right half of the lens model. But only the strangers' perceptions of the targets' physical attributes are reasonable causal variables, whereas the targets' self-perceptions of their own physical attributes are not reasonable causes of the strangers' trait inferences. For instance, strangers may infer targets' gregariousness from their perceptions of the targets' physical attractiveness, but cannot infer targets' gregariousness from the targets' self-perceptions of their own physical attractiveness.

In contrast, the relations between physical attributes and self-reports of personality (referred to by the left half of the lens model) are purely correlational anyway. Thus a strong argument exists to prefer the stranger ratings of physical attributes, but there is no argument to prefer the self-ratings of these attributes. Note, however, that self-ratings as well as stranger ratings of physical attributes reflect actual physical differences among the targets to some extent (see Table 1).

Correlations between physical and personality variables. The physical variables that mediate the consensus between targets and strangers were identified for the sound-film condition only, because this was the only condition for which strangers had rated the entire set of 45 physical attributes. These ratings of physical attributes were correlated with (a) the targets' self-ratings on Norman's 20 trait dimensions, (b) the targets' NEO-FFI scores, (c) the ratings for Norman's (1963) 20 trait dimensions by

the six strangers in the sound-film condition, and (d) the composite ratings by these strangers for Extraversion, Agreeableness, Conscientiousness, Emotional Stability, and Culture. Thus $45 \times 50 = 2,250$ correlations between physical attributes and ratings of personality were established. The correlations of the physical attributes with the targets' NEO-FFI scores and the strangers' composite ratings for the five personality domains are reported in Table 4.

Not surprisingly, many ratings of physical attributes correlated substantially with the strangers' trait inferences. But there were also various significant correlations between the ratings of physical attributes by strangers and the targets' self-reports of personality. In particular, self-reported Extraversion correlated significantly with 23 of the 45 physical attributes. Some of these physical cues referred to the targets' voice (soft-voiced vs. powerful voice), whereas others referred to the targets' static visible attributes (e.g., showy dress, stylish hair, and friendly expression) and the targets' dynamic visible attributes (e.g., fast movements and arm swinging). This diversity of physical concomitants of Extraversion may explain why strangers inferred this trait accurately even from a still picture or an audiotape of the targets' voice.

Match of validity relations and inferences from cues. Brunswik's lens model illustrates that three preconditions must be met in order that self-ratings and stranger ratings of a trait correlate substantially with each other: There must be (a) valid cues, that is, substantial correlations between physical cues and the self-ratings of the trait; (b) trait inferences from physical attributes, that is, substantial correlations between physical cues and the trait ratings by strangers; and (c) a match in sign and strength of the correlations of the physical variables with the self-ratings and the stranger ratings of the trait.

Consequently, stranger ratings may be inaccurate for three reasons: (a) no valid observable cues for a trait may exist; (b) strangers may not infer a trait from physical attributes; or (c) there may be valid cues as well as trait inferences, but the strangers may infer the trait inappropriately from the present physical cues. For instance, if strangers infer Agreeableness from extensive smiling, whereas there is no correlation or even a negative correlation between extent of smiling and self-reported Agreeableness, this visible attribute does not contribute to a positive correlation between self-ratings and stranger ratings of Agreeableness. Obviously, the three factors may account for invalid stranger ratings singly or in combination.

To investigate the actual impact of these three sources of validity, we calculated for each of 25 pairs of self-ratings and stranger ratings (a) the mean absolute correlation (irrespective of sign) between the 45 physical attributes and the self-ratings of the trait; (b) the mean absolute correlation between the 45 physical attributes and the stranger ratings of the trait; and (c) the correlation between the two vectors of correlations, one informing about the relations between physical attributes and self-reports and the other one informing about the relations between physical attributes and trait inferences by strangers. In the last analysis, the sign of the correlations between physical attributes and trait ratings was considered.

The first statistic informs about the average cue validity of the physical attributes, the second statistic informs about the average strength of trait inferences from these attributes, and

Table 4
Correlations of Physical Attributes With Targets' NEO-FFI Scores and Trait Ratings by Perceivers

Physical attribute	Targets' self-reports (NEO-FFI scores)					Perceiver-rated personality dimensions				
	I	II	III	IV ^a	V	I	II	III	IV	V
Estimated age	-.11	.05	.26**	.03	-.11	-.03	.04	.40**	.11	-.02
Sympathetic	.24*	.23*	-.03	-.05	.20*	.47**	.76**	.02	.28**	.67**
Attractive	.37**	.08	-.11	-.04	.17	.47**	.31**	-.25*	.17	.43**
Feminine	.35**	.32**	.04	-.23*	-.01	.22*	.48**	.30**	-.15	.49**
Effortful reading	.10	.31**	.32**	-.06	-.06	.14	.42**	.46**	-.09	.33**
Soft-voiced	-.26**	.10	-.13	-.09	.23*	-.42**	-.06	.07	-.21*	.01
High voice	.19	.22*	.04	-.18	-.06	.05	.42**	.35**	-.17	.39**
Powerful voice	.20*	-.14	.06	.08	-.23*	.46**	.05	-.10	.38**	-.01
Unpleasant voice	-.14	-.13	.07	.00	-.06	-.42**	-.46**	-.08	-.48**	-.49**
Haltingly speaking	-.05	-.12	.00	.18	.12	-.21*	-.13	-.32**	-.29**	-.30**
Easy to understand	.14	.28**	.10	-.05	-.11	.30**	.35**	.38**	.38**	.35**
Hectic speaking	.05	-.25**	-.13	-.06	.07	-.16	.30**	-.32**	-.49**	-.21*
Standard language	.14	.17	-.12	-.02	.01	.06	.08	.03	.11	.29**
Unrefined appearance	-.28**	-.19	-.19	.10	.04	-.22*	-.35**	-.34**	.10	-.39**
Made-up face	.36**	.14	.13	-.16	-.15	.31**	.25*	.19	-.17	.31**
Dark garments	-.19	.02	-.13	-.22*	.11	-.03	-.19	-.21*	-.03	.08
Showy dress	.34**	-.09	.02	-.08	.05	.47**	.06	-.39**	.02	.15
Informal dress	.01	-.04	-.25*	.04	.24*	.18	.03	-.55**	.17	.03
Unfashionable dress	-.27**	.00	-.07	.02	-.01	-.31**	-.08	.10	.09	-.20*
Long hair	.23*	.16	-.15	-.11	.03	.16	.11	-.15	-.10	.18
Stylish hair	.33**	-.08	.01	-.05	.05	.38**	-.03	-.17	-.12	.11
Light hair	.15	-.05	.03	.18	.02	.17	-.01	-.25*	-.04	-.21*
Stout physique	.00	.02	.17	-.08	-.07	.13	.02	.12	.13	-.23*
Tall stature	-.16	-.27**	-.12	.09	.03	-.04	-.15	-.18	.17	-.06
Less muscular physique	.08	.15	.00	-.05	.00	.00	.20*	.17	-.28**	.31**
Well-proportioned body	.06	-.05	-.17	.05	.13	.06	.02	-.21*	-.01	.11
Round face	.04	.11	.11	-.08	-.02	.20*	.17	.18	.09	-.09
Thin lips	-.22*	-.06	.04	.10	-.09	-.17	-.09	.13	.04	-.10
Childlike face	.10	.12	-.16	-.04	.11	.01	.17	-.20*	-.27**	.07
Hard facial lineaments	-.27**	-.25*	-.01	.13	-.01	-.23*	-.52	-.19	.11	-.32**
Friendly expression	.39**	.23*	.04	-.02	.07	.62**	.84**	.07	.29**	.56**
Unconcerned expression	-.26**	.00	-.05	-.06	.07	-.57**	-.31**	-.14	-.42**	-.36**
Self-assured expression	.29**	-.05	.00	.09	-.03	.65**	.30**	.04	.64**	.36**
Extensive smiling	.33**	.12	.03	.04	-.05	.62**	.69**	-.08	.26**	.32**
Relaxed sitting	.08	-.15	-.29**	.10	.13	.33**	.07	-.61**	.38**	-.01
Closed arms while sitting	-.06	.11	-.12	-.04	.11	-.17	-.15	-.08	-.24*	-.04
Avoids the camera	-.33**	.02	-.16	-.20	.05	-.54**	-.37**	-.04	-.36**	-.20*
Fast movements	.36**	-.21*	-.16	.01	.06	.40**	.03	-.35**	.00	-.01
Frequent hand movements	.13	-.23*	-.11	.20	.11	.23*	-.01	-.30**	.12	-.05
Frequent head movements	.23*	-.25*	.15	.11	.05	.37**	.15	-.12	.02	.04
Touches oneself frequently	.15	-.11	-.22*	.10	.13	.25*	.04	-.28**	.06	.04
Lifts feet while walking	.23*	.01	.07	-.10	-.01	.11	.24*	.24*	-.06	.09
Lack of arm swinging	-.31**	.02	.11	-.04	.05	-.35**	-.25*	.08	-.24*	.01
Long strides	.07	-.32**	-.16	-.04	.06	.20	-.18	-.39**	.21*	-.24*
Stiff walking	-.22*	.15	.18	-.21*	-.11	-.54**	-.16	.52**	-.61**	-.04

Note. NEO-FFI = NEO Five-Factor Inventory. I = Extraversion, II = Agreeableness, III = Conscientiousness, IV = Emotional Stability, V = Culture-Openness to Experience.

^a The signs of the correlations in this column were reversed because the NEO-FFI scale is keyed for Neuroticism.

* $p < .05$. ** $p < .01$.

the third statistic informs about the extent that the strangers infer the traits appropriately from the available cues. The last three columns of Table 5 report these three statistics for 25 personality measures. Moreover, the correlations between the trait ratings by targets and by strangers are reproduced in the

first data column of Table 5. Note that some of the figures in the last three columns of Table 5 were derived from information reported in Table 4. For instance, the figure .20 in the second data column of the row labeled *NEO Extraversion* is the mean of the absolute values in the first data column of Table 4. Corre-

Table 5
Mean Correlation (Irrespective of Sign) Between Trait Self-Ratings and 45 Physical Attributes, Trait Ratings by Strangers and 45 Physical Attributes, and Match Between the Two Vectors of Correlations

Trait dimension	Validity of stranger ratings	Mean absolute correlation of physical variables with		Correlation between vectors of correlations
		Self-rating	Stranger rating	
Talkative-silent	.30**	.16	.28	.45**
Frank-secretive	.23*	.14	.26	.52**
Adventurous-cautious	.37**	.18	.25	.46**
Sociable-reclusive	.38**	.15	.27	.28
NEO-FFI Extraversion	.51**	.20	.28	.89**
Good-natured-irritable	.08	.08	.20	-.20
Not jealous-jealous	.05	.09	.21	-.12
Gentle-headstrong	.18	.09	.22	.17
Cooperative-negativistic	.19	.14	.21	.32*
NEO-FFI Agreeableness	.35**	.14	.22	.57*
Fussy-careless	.33**	.12	.22	.70**
Responsible-undependable	.10	.07	.22	.13
Exacting-inexacting	.23*	.09	.22	.32*
Persevering-quitting	.12	.11	.19	.37*
NEO-FFI Conscientiousness	.25*	.11	.23	.76**
Poised-nervous	.19	.07	.18	.01
Calm-anxious	.21*	.12	.22	.20
Composed-excitable	-.08	.12	.13	.18
Not hypochondriacal-				
hypochondriacal	-.10	.08	.18	.04
NEO-FFI Neuroticism ^a	.10	.09	.20	-.43**
Artistically sensitive-insensitive	.15	.08	.18	-.01
Intellectual-narrow	.08	.07	.15	.15
Polished-crude	.38**	.14	.21	.49**
Imaginative-simple	.15	.11	.25	.34*
NEO-FFI Openness to Experience	.20*	.08	.21	.02

Note. NEO-FFI = NEO Five-Factor Inventory.

^a The sign of this validity coefficient was reversed.

* $p < .05$. ** $p < .01$.

spondingly, the figure .28 in the third data column of the same row is the mean of the absolute values in the sixth data column of Table 4. Finally, the last figure (.89) in the row *NEO Extraversion* is the correlation across rows, between the entries in the first and sixth data columns of Table 4. The figures in the rows labeled *NEO Agreeableness*, *NEO Conscientiousness*, *NEO Neuroticism*, and *NEO Openness to Experience* were calculated from the corresponding entries in Table 4.

The 45 physical attributes indicated aspects of Extraversion most strongly, whereas aspects of the other personality domains were less strongly indicated by the physical attributes. The lowest mean correlation between physical attributes and the self-ratings for an aspect of Extraversion ($r = .14$) was as high as the highest mean correlation between physical attributes and the self-ratings for any other trait. Similar findings were obtained concerning strength of trait inferences; the stranger ratings of Extraversion correlated higher with the 45 physical attributes than did the stranger ratings that referred to the other personal-

ity domains. The lowest mean correlation between physical attributes and stranger ratings for an aspect of Extraversion ($r = .25$) was as high as the highest mean correlation between physical attributes and the stranger ratings for any other trait. A somewhat different situation, however, was found concerning the strangers' tendency to consider the validity relations of the physical attributes appropriately. The correlations in the last column of Table 5 are similar for Extraversion and Conscientiousness, but much lower for the other personality domains.

Thus all three sources accounted for the higher accuracy of stranger ratings of Extraversion. By contrast, only one source accounted for the higher accuracy of stranger ratings of Conscientiousness compared with stranger ratings of Agreeableness, Emotional Stability, and Culture: Neither did the physical attributes indicate Conscientiousness more strongly, nor did the strangers infer Conscientiousness more strongly from the physical attributes. Rather, the higher accuracy of Con-

scientiousness ratings reflected the strangers' tendency only to infer Conscientiousness from the available cues such that the sign and the strength of the trait inferences matched the actual validity relations of the cues.

The validity of the strangers' trait ratings not only varied across the five trait domains but also within these domains. We therefore checked how well differences among the 25 validity coefficients in the first data column of Table 5 might be predicted from (a) differences in the mean correlation between physical attributes and the self-ratings of the traits (second data column), (b) differences in the mean correlation between physical attributes and the stranger ratings of the traits (third data column), and (c) differences in the strangers' sensitivity to the validity relations between physical cues and traits (fourth data column). The entries in the columns of Table 5 correlated between .46 and .73 across rows. All correlations were highly significant. This indicated that the strangers inferred traits more strongly and more appropriately from the targets' physical attributes the more the traits were indeed indicated by physical cues.

Moreover, a multiple regression analysis was run with the validity of the stranger ratings (first data column) as the criterion and the entries in the other three columns as predictors. The multiple correlation was .82 and thus explained quite exhaustively why the stranger ratings were more valid for some traits than for others. The correlations among the three predictors and their beta weights are reported in Figure 2. All predictors accounted independently for some of the differences in the validity of the stranger ratings for the 25 traits.

Discussion

In the present study, the strangers must have inferred trait impressions from physical attributes because they had no other information at their disposal. It is not clear, however, whether (a) the strangers' ratings of the physical attributes directly reflect those perceptions from which they inferred their trait impressions, or (b) the strangers derived all their ratings, those of the physical attributes and those of the traits, from more gen-

eral impressions. So the derivation of the ratings of the physical attributes is debatable.

Two findings, however, suggest that strangers form mental images of physical attributes first and that their ratings of these attributes—at least of some of these attributes—directly reflect these mental images. First, the rater agreement is higher for some of the physical attributes than for the most reliable and valid trait ratings. It is thus unlikely that the ratings of these particular physical attributes are derived from more general impressions. The situation may be different, however, concerning variables such as sympathetic that are not judged that reliably. Second, a hypothesis that the ratings of the physical attributes are derived from more general impressions cannot account for the finding that the stranger ratings are more valid for those traits that correlate more strongly with the physical attributes (see Figure 2). Thus strangers probably form images of physical attributes first and then derive their trait impressions from them.

Implicit Personality Theory

Another issue that requires comment is that the "actual" personality traits are measured by self-report. This may be a problem because targets and strangers may both inappropriately expect covariations between physical and personality attributes that do not actually occur. Moreover, targets and judges may both infer their trait ratings from observable attributes although this is not justified by the actual relations between physical and personality variables (Buse & Pawlik, 1984; Chapman & Chapman, 1967). This view, however, is not supported by the present findings because (a) the correlation of stranger ratings and self-ratings varies considerably among the various traits, and (b) for some traits, the trait inferences by strangers bear no similarity at all to the correlations between physical cues and the targets' self-reports (see Table 5, last column).

Why, for example, should targets and strangers share only their co-occurrence expectancies concerning physical attributes and carelessness, but not share their co-occurrence expectancies concerning physical attributes and jealousy? The notions of physical and vocal attractiveness stereotypes suggest that attractive people, people who smile extensively, people who have a pleasant voice, and people who show a refined appearance, are all perceived as less careless as well as less jealous. Table 5, however, shows that the physical attributes correlate in a similar way with carelessness ratings by targets and strangers, whereas the physical attributes correlate in a dissimilar way with jealousy ratings by targets and strangers. The most parsimonious explanation of these findings is that the correlations between physical attributes and self-reports do not mainly reflect illusory co-occurrence expectancies between physical attributes and traits. Rather, these correlations reflect the actual covariation of trait level with physical attributes. Further analyses that point to the accuracy of implicit personality theory in general are reported elsewhere (Borkenau, in press; Borkenau & Ostendorf, 1987).

Why Do Physical and Personality Variables Co-Occur?

Co-occurrences of physical and personality variables probably reflect several mechanisms. First, personality may influ-

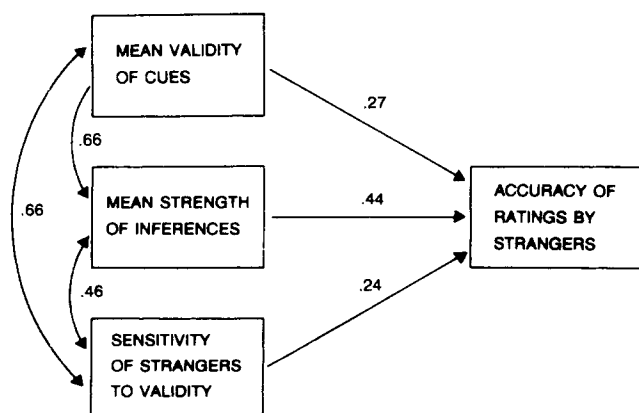


Figure 2. Sources of differences among validity coefficients for 25 traits.

ence attributes like friendly versus grumpy face by affect expression (Ekman, Friesen, & Ellsworth, 1972), and attributes like refined appearance and showy dress as signs of a preferred mode of living. Moreover, people sometimes dress and act in certain ways to convey a particular impression of their personality. But the influence of personality on appearance is not likely to be unidirectional. Rather, physical attributes are also likely to have an effect on personality development. More physically attractive people, for instance, are preferred as employees (Bardack & McAndrew, 1985) and as mates (Buss, 1989). Thus attractive people are likely to experience more success, and this probably has some side effects on personality development. Moreover, Snyder and Swann (1978) reported that the expected Extraversion of subjects influences their actual Extraversion. Symbolic interactionism goes even as far as to claim that we are entirely as others see us. But this view is probably too radical: Table 3 shows that only with respect to some traits do targets see themselves as others see them. Moreover, Plomin, Willerman, and Loehlin (1976) found that monozygotic twin pairs tend to be less similar the more frequently they are confused. However, the present research shows that lay people are aware of actual co-occurrences between physical and personality variables and that they use these co-occurrences to accurately infer personality traits of strangers from physical cues.

Differences Among Traits

The present study also provides information on why ratings by strangers are more valid for some traits than for others. Table 5 shows that the physical attributes investigated here bear a stronger relation to the targets' self-reported Extraversion than to self-reported Agreeableness, Conscientiousness, Emotional Stability, and Culture. The stranger ratings appropriately reflect this circumstance, as those for Extraversion are most strongly related to the physical attributes. A similar relation holds also across all 25 validity coefficients (see Figure 2): The more the physical attributes correlate with self-reports of a trait, the more the strangers infer this trait from the physical attributes.

Another issue is whether strangers use each attribute according to its actual validity relation, thus distinguishing appropriately among direct relations, inverse relations, and nonexistent relations between physical attributes and traits. The last column in Table 5 shows that strangers are most sensitive to the actual validity relations of cues concerning Extraversion and Conscientiousness. This may explain why stranger ratings tend to be most valid for these two domains. Moreover, a substantial correlation is found across the 25 single scales between the presence of valid cues and the sensitivity of strangers to the actual validity relations (see Figure 2). This shows how efficiently judges exploit the scarce information that is provided to them: If there are valid cues for a trait, strangers draw strong inferences from these cues that appropriately reflect the actual validity relations. In contrast, if there are less valid cues for a trait, trait inferences are less strong, and the actual validity relations are by and large ignored. Thus the judges focus their limited information-processing capacity on those traits that can be most accurately inferred from the physical attributes.

Ratings by strangers tend to be most valid for Extraversion

and Conscientiousness. But they are sometimes more accurate for Extraversion than for Conscientiousness (Hunt & Lin, 1967; Watson, 1989), sometimes equally accurate for both trait domains (Norman & Goldberg, 1966), and sometimes more accurate for Conscientiousness than for Extraversion (Albright et al., 1988). The present study sheds light on why these discrepant findings occur. Whereas Extraversion seems to be inferred from visual cues (Albright et al., 1988) as well as from acoustic cues (Scherer, 1978), the physical attributes that indicate Conscientiousness seem to be almost entirely visual ones. The validity of stranger ratings of Conscientiousness decreases considerably in the present study if only acoustic information is provided. Correspondingly, stranger ratings were much more valid for Extraversion than for Conscientiousness in Hunt and Lin's study, in which personality traits had to be inferred from speech alone.

Conclusions

Although there was early work on the relations between physical and psychological variables by authors like Allport and Vernon (1933), Kretschmer (1921), and Sheldon (1940), trait inferences from physical attributes were most extensively studied in the literature on social stereotypes. Research like that by Kretschmer and Sheldon was first criticized for some methodological flaws (Anastasi, 1966) and was later dismissed as informing mainly about social stereotypes (Buse & Pawlik, 1984). The research then focused on the relations between physical attributes and trait inferences by strangers (the right side of the lens model), whereas the validity of the physical cues (the left side of the lens model) was no longer an important research topic. Indeed, there is much recent and successful work on the facial (Ekman, Friesen, & Ellsworth, 1972) and vocal (Scherer, 1986) expression of affect. The relations between physical attributes and *traits*, however, are still regarded as predominantly illusory.

The present study shows that such a view is not entirely justified. Rather, there are relations between physical attributes and personality traits, and subjects are quite aware of these relations. This obviously does not imply that judges could not be more accurate (see Table 5, last column). Further research on these relations may therefore also be important for the applied concerns of personality assessment: If more evidence about the actual relations between physical and psychological attributes will be available, this information may be used to train judges for providing more accurate judgments of persons they do not know much about. This skill is required among personnel managers, for instance, who have to make decisions on the basis of job interviews.

References

- Albright, L., Kenny, D. A., & Malloy, T. E. (1988). Consensus in personality judgments at zero acquaintance. *Journal of Personality and Social Psychology*, 55, 387-395.
- Allport, G. W., & Vernon, P. E. (1933). *Studies in expressive movement*. New York: Macmillan.
- Amelang, M., & Borkenau, P. (1982). Über die faktorielle Struktur und externe Validität einiger Fragebogen-Skalen zur Erfassung von Dimensionen der Extraversion und emotionalen Labilität [On the fac-

- tor structure and external validity of some questionnaire scales measuring dimensions of extraversion and emotional lability]. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 3, 119–146.
- Amelang, M., Köhler, B., & Gold, A. (1983). Physische Attraktivität, Ausdrucksverhalten, und Persönlichkeit: Über einige Zusammenhänge zwischen Selbst- und Fremdeinschätzungen [Physical attractiveness, expressive behavior, and personality: On some relations between self-ratings and ratings by strangers]. In G. Lüer (Ed.), *Bericht über den 33. Kongreß der Deutschen Gesellschaft für Psychologie*. Göttingen, Germany: Hogrefe.
- Anastasi, A. (1966). *Differential psychology*. New York: Macmillan.
- Bardack, N. R., & McAndrew, P. T. (1985). The influence of physical attractiveness and manner of dress on success in a simulated personnel decision. *The Journal of Social Psychology*, 125, 777–778.
- Borkenau, P. (in press). Implicit personality theory and the five-factor model. *Journal of Personality*.
- Borkenau, P., & Ostendorf, F. (1987). Fact and fiction in implicit personality theory. *Journal of Personality*, 57, 415–443.
- Borkenau, P., & Ostendorf, F. (1990). Comparing exploratory and confirmatory factor analysis: A study on the five-factor model of personality. *Personality and Individual Differences*, 11, 515–524.
- Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. Berkeley: University of California Press.
- Buse, L., & Pawlik, K. (1984). Kretschmers Konstitutionspsychologie als implizite Persönlichkeitstheorie: Selbst-Attribuierungseffekte in Abhängigkeit vom Körperbau-Persönlichkeits-Stereotyp. [Kretschmer's psychology of physique as an implicit personality theory: Effects of the physique-personality stereotype on self-attributions]. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 5, 111–129.
- Buss, D. M. (1989). Sex differences in human mate preferences: Evolutionary hypotheses tested in 37 cultures. *Behavioral and Brain Sciences*, 11, 1–49.
- Chapman, L. J., & Chapman, J. P. (1967). Genesis of popular but erroneous psychodiagnostic observations. *Journal of Abnormal Psychology*, 72, 193–204.
- Cheek, J. (1982). Aggregation, moderator variables, and the validity of personality tests. *Journal of Personality and Social Psychology*, 43, 1254–1269.
- Cleeton, G. U., & Knight, F. B. (1924). Validity of character judgments based on external criteria. *Journal of Applied Psychology*, 8, 215–231.
- Cloyd, L. (1977). Effect of acquaintanceship on accuracy of person perception. *Perceptual and Motor Skills*, 44, 819–826.
- Costa, P. T., & McCrae, R. R. (1985). *The NEO Personality Inventory Manual*. Odessa, FL: Psychological Assessment Resources.
- Costa, P. T., & McCrae, R. R. (1989). *The NEO PI/FFI Manual Supplement*. Odessa, FL: Psychological Assessment Resources.
- Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, 41, 417–440.
- Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology*, 24, 285–290.
- Ekman, P., Friesen, W. V., & Ellsworth, P. C. (1972). *Emotion in the human face*. New York: Pergamon Press.
- Funder, D. C., & Colvin, C. R. (1988). Friends and strangers: Acquaintanceship, agreement, and the accuracy of personality judgment. *Journal of Personality and Social Psychology*, 55, 149–158.
- Gifford, R., Fan Ng, C., & Wilkinson, M. (1985). Nonverbal cues in the employment interview: Links between applicant qualities and interviewer judgments. *Journal of Applied Psychology*, 70, 729–736.
- Hunt, R. G., & Lin, T. K. (1967). Accuracy of judgments of personal attributes from speech. *Journal of Personality and Social Psychology*, 6, 450–453.
- Kenny, D. A. (1991). A general model of consensus and accuracy in interpersonal perception. *Psychological Review*, 98, 155–163.
- 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, 88–97.
- Kretschmer, E. (1921). *Körperbau und Charakter* [Physique and character]. Berlin: Springer-Verlag.
- McCrae, R. R., & Costa, P. T. (1989). Different points of view: Self-reports and ratings in the assessment of personality. In J. P. Forgas & J. M. Innes (Eds.), *Recent advances in social psychology: An international perspective*. Amsterdam: Elsevier Science.
- 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., & Goldberg, L. R. (1966). Raters, ratees, and randomness in personality structure. *Journal of Personality and Social Psychology*, 4, 681–691.
- 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. (1989). Consensus in personality judgments: Moderating effects of target-rater acquaintanceship and behavior observability. *Journal of Personality and Social Psychology*, 56, 823–833.
- Plomin, R., Willerman, L., & Loehlin, J. C. (1976). Resemblance in appearance and the equal environments assumption in twin studies of personality traits. *Behavior Genetics*, 6, 43–52.
- Scherer, K. R. (1978). Personality inference from voice quality: The loud voice of extraversion. *European Journal of Social Psychology*, 8, 467–487.
- Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin*, 99, 143–165.
- Sheldon, W. H. (1940). *The varieties of human physique: An introduction to constitutional psychology*. New York: Harper & Row.
- Snyder, M., & Swann, W. B. (1978). Hypotheses-testing processes in social interaction. *Journal of Personality and Social Psychology*, 36, 1202–1212.
- Watson, D. (1989). Strangers' ratings of the five robust personality factors: Evidence of a surprising convergence with self-report. *Journal of Personality and Social Psychology*, 57, 120–128.
- Zuckerman, M., & Driver, R. E. (1989). What sounds beautiful is good: The vocal attractiveness stereotype. *Journal of Nonverbal Behavior*, 13, 67–82.

Received December 26, 1990

Revision received April 26, 1991

Accepted April 29, 1991 ■