Journal of Personality and Social Psychology
1992, Vol. 62, No. 1, 88-97

0022-3514/92/$3.00

Consensus at Zero Acquaintance: Replication, Behavioral Cues, and Stability

David A. Kenny, Caryl Horner, Deborah A. Kashy, and Ling-chuan Chu
University of Connecticut

That observers tend to agree in their ratings of a target even if they have never interacted with that target has been called *consensus at zero acquaintance.* The basic finding that consensus is highest for judgments concerning a target's degree of extraversion (EV) and somewhat weaker for judgments of conscientiousness is replicated. Several potential observable cues that might be used by judges when rating targets are examined. The finding that ratings of physical attractiveness correlate with judgments of EV is replicated. In Study 1, rapid body movements and smiling were also found to correlate with EV judgments. The level of consensus declined when initially unacquainted Ss interacted one-on-one (Study 2), but did not decline—and even increased—when Ss interacted in a group (Study 3). Ss judged as extraverted at zero acquaintance were also seen as extraverted after interacting with others.

There has been considerable interest in the question of consensus: When two judges evaluate the same target, to what extent do the judges agree with one another? Evidence concerning consensus is theoretically important for two reasons (Kenny, 1991). First, social psychologists have long wondered about the extent to which person perception is driven by the stimulus or merely reflects assumptions of the perceiver. Beginning with the classic study by Dornbusch, Hastorf, Richardson, Muzzy, and Vreeland (1965), the level of consensus has been used to assess the extent to which person perception is target-based. Second, personality psychologists (e.g., Kenrick & Funder, 1988) have argued that consensus in personality judgments implies that traits are important factors in determining behavior.

An obvious question in the study of consensus is the relationship between acquaintance and consensus. It would seem logical that greater acquaintance leads to greater consensus. If one is to study consensus between interacting pairs of people, one needs first to know the baseline level of consensus between people who have never interacted. Albright, Kenny, and Malloy (1988) described *zero acquaintance* as a condition in which one person observes another, but the two have never engaged in social interaction. Somewhat surprisingly, it seems that people's first impressions of a stranger on some traits are remarkably similar.

To facilitate comparisons across studies, we consider only studies of consensus at zero acquaintance that used the five Norman (1963) personality factors: Extraversion, Agreeableness, Conscientiousness, Emotional Stability, and Culture (or Intelligence). There are four studies: Norman and Goldberg (1966), Albright et al. (1988, Study 3), Watson (1989), and DiPilato (1989). In each study, the investigators measured the extent to which the ratings of a single stranger made by 2 people agreed, and so they measured consensus at zero acquaintance. The measure of consensus was the proportion of variance in the trait rating due to the target. This proportion can be viewed as the correlation between 2 judges' ratings of a common set of targets. The proportions of target variance in the five Norman factors for the four studies are presented in Table 1.

It is apparent from Table 1 that judges do agree in their ratings of strangers. Only 2 of the 16 variances are zero. Consensus is highest on the Extraversion factor for all four studies. An average of 25% of the variation in ratings of Extraversion is due to the target being rated. Evidence from studies other than zero-acquaintance studies indicates that across varying levels of acquaintance, consensus is particularly high in ratings of Extraversion (Funder & Dobroth, 1987; Park & Judd, 1989). Thus, there is converging evidence both in ratings of strangers and acquaintances for consensus in judgments of Extraversion.

As can be seen in Table 1, the Conscientiousness factor shows the second highest level of consensus in three of the four studies. Both Albright et al. (1988) and Watson (1989) speculated that consensus on this factor is brought about by neatness in dress and grooming. The other three factors do not consistently exhibit consensus at zero acquaintance.

Although consensus at zero acquaintance has been consistently replicated for Extraversion and Conscientiousness, the procedures generally used in these studies may not have adequately controlled for prior acquaintance. In addition, exactly what cues bring about consensus on these two factors is not understood, and there are no data describing exactly what occurs when subjects interact with one another.

As was mentioned, prior studies of zero acquaintance may not have sufficiently controlled for the prior acquaintance of judge and target. Subjects may have had some degree of familiarity with one another before the study. For instance, it seems plausible that subjects may have taken other psychology courses together. In addition, as discussed by Watson (1989) and Al-

This research was supported in part by National Science Foundation Grant BNS-8807462 and National Institute of Mental Health Grant R01-4029501 to David A. Kenny.

Thanks are due Rebecca Hodgkins, Christine McManus, Scott Parisi, and Claudine Gill, who assisted us in the data collection. We would also like to thank Linda Albright, Diane Berry, Thomas Malloy, and David Watson, who provided us with comments on an earlier version of this article.

Correspondence concerning this article should be addressed to David A. Kenny, Department of Psychology, University of Connecticut, Storrs, Connecticut 06269-1020.

Table 1
Proportion of Target Variance From Four Zero-Acquaintance Studies

Factor	Norman & Goldberg (1966)	Albright, Kenny, & Malloy (1988)	Watson (1989)	DiPilato (1989)	*M*
Extraversion	.18	.20	.36	.27	.25
Agreeableness	.07	.03	.21	.00	.08
Conscientiousness	.16	.10	.29	.09	.16
Emotional stability	.08	.09	.09	.10	.09
Culture	.12	.00	.14	.09	.09
n	84	169	250	96	—

Note. *M* = mean of all proportions of target variance by factor.

bright et al. (1988), some verbal and nonverbal interaction may have occurred during the study itself. Consensus at zero acquaintance needs to be studied in more controlled situations in which the judge is unacquainted with the target and unable to interact with the target during the rating process.

A second important issue needing clarification concerns how consensus arises. That is, what characteristics of a target bring about the judges' similar impressions? Logically, if judges agree in their ratings of personality, there must be some observable stimulus characteristic that creates the consensus (Kenny & Malloy, 1988). The task, then, is to isolate these observable stimulus characteristics. There are two classes of cues that judges can use to make personality inferences at zero acquaintance: physical appearance cues (e.g., sex, clothing) and nonverbal behavior. Of the four prior studies that investigated the zero-acquaintance effect, only Albright et al. (1988) examined the processes used in inferring personality traits at zero acquaintance. They found that physical attractiveness correlated with judgments of Extraversion. Building on the previous literature in person perception and nonverbal behavior, in this study, we investigate in greater detail the process by which consensus is achieved at zero acquaintance.

In none of the prior studies were the subjects allowed to interact with each other after zero acquaintance ratings. So another unresolved issue concerns the stability of consensus in first impressions. After a judge interacts with a target and thus is exposed to more information about the target, does the level of consensus increase or decline? Consensus at zero acquaintance is based on a set of stereotypes (an assumption that an observable stimulus characteristic is associated with a certain personality characteristic). At issue is whether these stereotypes influence judgments after the judges have interacted with the target. For the first time, we study the stability of judgments at zero acquaintance.

There are two aspects of consensus that may change over time. First, the level of consensus may change over time. At zero acquaintance, judges may to some extent agree on which targets are extraverted and which targets are introverted. After interaction, the judges may show even higher agreement concerning the extraversion of a set of targets. However if the level of consensus rises over time, this does not imply that the targets are seen differently at the two times. The same factor (e.g., Physical Attractiveness) could be determining consensus at both times, but that factor may be more influential at the second time point.

The second aspect of consensus that may change over time is the actual impression. The level of consensus in ratings may remain the same over time, whereas the standings of targets on the traits may change substantially. For instance, it may be that a target is rated by all the judges as quite extraverted at zero acquaintance. After interaction, however, that target might be rated as quite introverted by all of the judges. The level of consensus may not have changed; however, the impression of the target has. If the impression has changed, then there is some indication that different factors are determining the ratings at the two times. For instance, appearance cues may drive consensus at zero acquaintance, whereas behavioral information may be more important when the judge is acquainted with the target.

Social Relations Model

An elegant yet simple model that can be used to describe and analyze interpersonal perception data is the Social Relations Model (Kenny & La Voie, 1984). It has been applied to a number of studies in social perception: Albright et al. (1988); DePaulo, Kenny, Hoover, Webb, and Oliver (1987); DiPilato (1989); Kenny (1988); Kenny and Albright (1987); Malloy and Albright (1990); and Park and Judd (1989). In the perception context, the Social Relations Model proposes that any rating score can be partitioned into three basic components: the perceiver effect, the target effect, and the relationship effect.

The perceiver effect measures a person's general response level across all targets. For example, a person who tends to rate others at high levels of sociability can be considered to possess a high perceiver effect for sociability. The perceiver effect can be viewed as a response set. By response set, we mean the tendency for some subjects to give high ratings to all targets and other subjects to give low ratings. This tendency either could be a measurement artifact or might reflect certain psychological processes.

The target effect measures the level at which judges generally rate a particular target on a specific trait. For example, a person consistently rated by judges as unsociable would possess a low target effect for sociability. However, just because judges agree about a person does not necessarily imply that the judges are accurate.

The relationship effect measures the unique perception that one judge has of a particular target. That is, if the judge rates the target as especially sociable (more sociable than that judge tends to rate all other targets and more sociable than that target tends to be rated by all other judges), then that judge–target dyad would have a high relationship effect for sociability.

The Social Relations Model is well suited to the study of consensus in interpersonal perception because target effects directly measure the degree to which a set of judges agree in their ratings of a particular target, after partitioning out the perceiver and relationship effects. Thus, the amount of consensus present is captured by the amount of variance due to the target. In this article, we report the proportion of target variance (i.e., the target variance divided by the total variance) as our measure of consensus.

Present Studies

This series of studies is presented to further extend work in the area of consensus at zero acquaintance. Previous studies have not always controlled for the precise level of prior acquaintance or interaction during the experimental session. One purpose of Study 1 was to exert control over the presentation of the stimulus targets to the judges. Because videotapes are used in Study 1, judges and targets could have no interaction (verbal or nonverbal) during the experiment. In addition, because the targets are videotaped, it is possible to code the physical and behavioral cues of the targets to examine which cues judges use to make their ratings.

Studies 2 and 3 were designed to see how consensus levels might change after unacquainted subjects engage in social interaction and become acquainted with each other. Study 2 examines this process after brief one-on-one interactions, whereas Study 3 examines the stability in consensus after a series of group interactions. These studies also involve an examination of several possible observable variables that judges might use as cues when making their ratings.

Results of the three studies are presented as follows: The Social Relations Model variance partitioning is discussed for each study individually. The variance partitioning yields evidence concerning consensus effects. Next, various physical and behavioral cues that might be used by the judges are discussed. Finally, the stability of consensus effects for Studies 2 and 3, each of which involves measurement over time, is discussed.

Study 1

Overview

One goal of Study 1 was to replicate previous zero-acquaintance findings while exerting control over the degree of acquaintance and interaction between judges and targets. A second goal of Study 1 was to isolate the cues that judges use when making their ratings. Briefly, in this study, a set of judges viewed a videotape of 32 targets, and each judge rated each target on the Norman (1963) factors. Independent raters also rated each target on a set of behaviors to ascertain behavioral cues that produce target effects.

Method

Subjects

There were 73 subjects in Phase 1, and a different set of 113 subjects (52 men and 61 women) in Phase 2 of this study. All of these subjects were in an introductory psychology course at a large, eastern state university. Participation in the study was voluntary, and all subjects received experimental credit for their participation in the study. The Phase 1 subjects were run in the spring, and the Phase 2 subjects were run in the fall, 2.5 years later. Four upper-division psychology students, 2 men and 2 women, served as raters in Phase 3.

Procedure

Phase 1. At the beginning of each experimental session, small groups of subjects were told that the purpose of the experiment was to measure a set of personality traits. Subjects were informed that all results would remain confidential. In addition to completing various personality questionnaires, the subjects were informed that they would individually rate five TAT cards in a separate room. The questionnaires completed by all subjects included self-ratings on a set of traits selected from the Norman (1963) factors, the Self-Monitoring Scale (Snyder, 1974), the Marlowe-Crowne Social Desirability Scale (Crowne & Marlowe, 1964), and a demographic questionnaire.

After completing ratings on the traits from the Norman (1963) factors, individual subjects were randomly selected to—one at a time—leave the laboratory and accompany the experimenter to a small, private office. This room contained a two-way mirror, partially hidden by a bookcase; the curtain was drawn, and only a small part of the mirror was visible to the subject. The subject was asked to sit in a chair facing the wall that contained the mirror. Each subject was told that he or she would be shown five TAT cards by another experimenter and then would be asked to state his or her impressions of the five drawings. The subject was informed that after the ratings of the TAT drawings were completed, he or she would return to the original experimental room to finish the remaining questionnaires. The subject was then asked whether he or she had any questions, and then the experimenter left the room. While the subject waited for the second experimenter, he or she was secretly videotaped for 1 min through the two-way mirror.

The experimenter then returned to the private experimental room and informed the subject that the second experimenter would not be showing the subject the TAT cards. The subject was debriefed and asked whether he or she would agree to remain on tape so that the experimenter could study the perceptions of people. All subjects who declined were immediately erased from the videotape (9 subjects declined). Individual subjects were thanked for their participation and told that they had 1 week to change their minds about remaining on the videotape (none did). As each subject returned to the original experimental room, this procedure was repeated for each of the remaining subjects. At the conclusion of the experimental sessions, 32 women and 32 men were videotaped.

From this original tape, two edited tapes were made. Thirty-two people were placed on Tape 1, and the remaining 32 were placed on Tape 2. Each target was presented for 20 s on the edited videotapes. In editing the tapes, the earliest segment during which the target directly faced the camera was selected.

Phase 2. In Phase 2, subjects were told that they would be viewing videotapes of other students. Subjects were informed that they would be watching 20-s clips of 32 people who were not aware that they were being videotaped while they waited alone in a private office for 1 min.

Groups of subjects (approximately 10 per group) watched each of the 32 targets on a videotape for 20 s and then rated each target on the 10 traits from the Norman (1963) factors. The traits were sociable–reclu-

sive and talkative–silent for the Extraversion factor, cheerful–irritable and cooperative–negative for the Agreeableness factor, careful–careless and responsible–irresponsible for the Conscientiousness factor, calm–anxious and relaxed–tense for the Emotional Stability factor, and intelligent–unintelligent and imaginative–unimaginative for the Culture factor. Each trait was on an unnumbered 7-point scale with two anchors (e.g., *sociable* and *reclusive*). Subjects were also asked whether they had previously seen, talked to, or been in a class with the target.

At each experimental session, subjects viewed either Tape 1 or Tape 2. To control for order effects, both tapes were partially reversed, and each of these reversed tapes was shown to about half of the subjects. In the reversed tapes, the sessions began with Target Numbers 17–32 and concluded with Numbers 1–16.

After all the targets had been viewed and rated, the subjects were asked to rate themselves on the 10 traits from the Norman (1963) factors, and they also answered a set of demographic questions. All subjects were then debriefed as to the nature of the study. The experimental sessions for this phase of the study took approximately 45 min.

Phase 3. The purpose of this phase was to measure the effect that nonverbal behaviors and physical appearance on the part of the targets might have on subjects' ratings of the targets. Four raters (2 men and 2 women) were recruited from an upper-division psychology seminar course. These raters, in a group, viewed the same videotapes that were presented in Phase 2. They were instructed to rate independently each target on 17 nonverbal behavior Likert-type rating scales, ranging from 1 (*not at all*) to 7 (*all of the time*). The entire set of ratings were made after viewing each target and were based on the quantity and frequency of the nonverbal behaviors. Raters judged the targets on 17 nonverbal behaviors: smiling, frowning, forward gaze (toward the concealed camera), downward gaze, sideways gaze, information-seeking gaze (visual search), rapid movement, slow movement, arms or hands folded, erect posture, primping self-touch (grooming behaviors), nonprimping self-touch, attractive, jewelry, hair groomed, fashionable dress, and casual dress. All 4 raters viewed one tape per experimental session with a 10-min break during the middle of the tape.

Results

We first checked to see whether the judges were acquainted with the targets. Because the targets were primarily seniors and the judges were mostly freshmen, we expected low levels of acquaintance. For the 113 judges, only 1.8% of the targets had been seen, 0.9% had been spoken to, and 0.6% had been in a class with the judge. Thus, there is strong evidence that the judges were not familiar with the targets.

The variance partitioning for Study 1 is presented in Table 2. The design used in Study 1 is a half block (Kenny, 1990). The numbers in Table 2 are the proportions of variance: the variance in the Social Relations Model component divided by the total variance. Unfortunately, there is no standard test of statistical significance for the variance components of this design. On the basis of past experience, variance components of 10% or more are interpretable.

As can be seen in Table 2, a small amount of perceiver variance is present in the ratings on all of the factors. Consensus effects are apparent in the amount of target variance in the data. Twenty-two percent of the variance in ratings of Extraversion was target-based. Thus, even with only a 20-s exposure to a target sitting alone in a room, and the judges having no prior interaction or communication with that target, people tended to agree on which targets were extraverted and which targets were not. Smaller consensus effects were found for the Emotional Stability, Agreeableness, and Conscientiousness factors; approximately 10% of the variation in ratings on these factors was due to the target being rated. There was no evidence of consensus for the Culture factor.

Table 2
Relative Variances for Study 1 (113 Judges and 64 Targets)

Factor	Perceiver	Target	Relationship	Error
Extraversion	.08	.22	.41	.29
Agreeableness	.07	.10	.27	.55
Conscientiousness	.07	.13	.38	.42
Emotional stability	.07	.08	.41	.44
Culture	.02	.00	.21	.77

The proportion of relationship variance is quite large, ranging from 21% to 41% of the variance for the five factors. The relationship effect measures the degree to which a judge uniquely perceives a given target. Thus, the large amount of relationship variance indicates that the judges had relatively idiosyncratic views of the targets.

Study 2

Overview

In addition to providing a replication of the consensus at zero acquaintance findings, Study 2 considers another aspect of consensus in initial impressions: stability over time. Unacquainted female subjects were put into small groups and asked to rate one another on the Norman (1963) factors. After these ratings, subjects interacted one-on-one with each member of their group. After each interaction, subjects rated their partners a second time on the Norman factors. Stability of consensus was then measured by the correspondence between the ratings at zero acquaintance and the ratings after one-on-one interaction.

Method

Subjects

Study 2 involved 108 undergraduate women enrolled in an introductory psychology class at a large, northeastern state university. Subjects participated in 4-person groups, and there were a total of 27 groups. Subjects received experimental credit in partial fulfillment of a course requirement for their participation in this study.

Procedure

Groups of subjects were assembled in the laboratory and were instructed not to interact while they awaited the arrival of additional subjects. Two female experimenters then randomly assigned subjects into 4-person groups, making certain that groups consisted of unacquainted subjects. Each subject in a group was given a name tag identifying her as Person A, B, C, or D.

The subjects were told that the purpose of the study was to try to understand how people think about each other. The procedure involved two stages. In the first stage (Wave 0), each subject rated herself

and the 3 other members of her group on the same 10 traits selected from the Norman (1963) factors that were used in Study 1. Each subject also rated her own and the others' physical attractiveness at this point.

In the second stage (Wave 1), subjects interacted one-on-one (i.e., A interacted with B and C interacted with D) with each of the other group members. Subjects were told to get to know each other. Each interaction lasted for approximately 8 min. After each interaction, each subject rated her partner on the 10 traits. These instructions were repeated two more times until all pairs of subjects within a group had interacted. The experimental session lasted approximately 50 min.

Results

The design used is a round-robin design. Because a total of 27 round-robins were gathered, group is treated as the unit of analysis, and the degrees of freedom for tests of significance are 26.

The variance partitioning for Study 2 is presented in Table 3. At zero acquaintance (Wave 0), there is substantial perceiver variance for each of the five factors. Thus, some subjects tended to rate all of their partners as quite agreeable, for example, whereas other subjects tended to rate all of their partners as somewhat disagreeable.

Evidence for consensus at zero acquaintance is found in the significant target variances for Extraversion, Conscientiousness, and Culture factors. Most noteworthy is that 26% of the variation in ratings of Extraversion was target-based: Perceivers tended to agree who within the group was extraverted. This finding replicates results found in Study 1 as well as those presented in previous zero-acquaintance studies. People apparently did not agree who in the group was agreeable or emotionally stable.

Relationship variance was significant for the Extraversion, Agreeableness, and Conscientiousness factors, indicating that the judges viewed the targets on these factors in a fairly idiosyncratic manner. There was not significant relationship variance for the Emotional Stability or Culture factors.

After interacting one-on-one (Wave 1), the level of perceiver variance did not substantially change; however, the level of target variance dropped dramatically. Note that this measure of stability of consensus assesses the changes in the level of agreement over time. It does not address whether the impressions made by targets changed over time. At Wave 1, consensus in ratings of Extraversion and Culture was no longer significant. That is, although at Wave 0 judges generally agreed who in the group was extraverted, at Wave 1, judges were no longer consistent in their ratings of who was extraverted. Consensus for the Conscientiousness factor, although still significant, was also reduced. It thus seems that the short dyadic interactions provided subjects with enough individualized information for them to no longer agree who in the group was extraverted or cultured.

One might wonder why at Wave 1 the 12% target variance for Extraversion is not significant, but the 7% for Conscientiousness is. Group is used as the unit of analysis and so if a large estimate varies by group a great deal, it may not be significant. However, if a small estimate consistently replicates across groups, it will be significant.

Whereas target variance declined, relationship variance increased for four of the five factors. This increase in dyadic-level variance further substantiates the notion that the brief interactions provided subjects with unique information that guided their ratings.

The degree to which the judges' first impressions related to the judges' impressions after the interactions is discussed in conjunction with stabilities in Study 3.

Table 3
Relative Variances for Study 2

Factor	Perceiver	Target	Relationship	Error
Wave 0				
Extraversion	.13*	.26*	.26*	.35
Agreeableness	.21*	.01	.23*	.54
Conscientiousness	.36*	.12*	.10*	.42
Emotional stability	.20*	.00	.07	.73
Culture	.31*	.06*	.00	.63
Wave 1				
Extraversion	.10	.12	.47*	.31
Agreeableness	.21*	.02	.12*	.66
Conscientiousness	.23*	.07*	.14*	.55
Emotional stability	.22*	.00	.24*	.54
Culture	.25*	.00	.14*	.61

* $p < .05$, $df = 26$.

Study 3

Overview

Study 3 was undertaken to provide both a replication of the consensus results as well as to examine the stability of consensus in a context differing from that of Study 2. In Study 3, students in two classes were divided into small groups on the first day of class and were asked to rate each other, as well as themselves, on the five Norman (1963) factors. During the semester, people within their original groups worked together on a series of tasks, and at the end of the semester, ratings on the Norman factors were retaken.

Method

Subjects

Eighty-three students from two different social psychology classes at an eastern university served as subjects for this study.

Procedure

The procedure of this study was as in Albright et al. (1988). Students on the first day of class were placed into 4- or 5-person groups in which all members were unacquainted (Wave 0). Once the groups had been formed, each member of the group was assigned an identification letter (A–E). They then rated themselves and each other on the 10 traits chosen to represent the five Norman (1963) factors. (The same traits from Studies 1 and 2 were used, except *good-natured* was used instead of *cheerful* for the Agreeableness factor.) Unlike Studies 1 and 2, the rating scales used numbers and not checkpoints. The subjects also rated their own and each others' physical attractiveness.

During the semester, the same groups worked together on a series of problem-solving tasks and subsequently became acquainted with one another. On the last or next to last day of class (Wave 1), the group members rated one another on the same scales. Complete data at both

times were available for a total of 70 subjects and 17 groups. Of the groups, 4 were 3-person groups, 7 were 4-person groups, and the remaining 6 groups were 5-person groups. Inclusion of the 3-person groups in the analysis requires the assumption that dyadic reciprocity is zero. (Dyadic reciprocity implies that if one person, say Dan, sees his partner, say Marilyn, as especially sociable, then Marilyn will see Dan as especially sociable.) The analysis excluding the 3-person groups indicated that reciprocity was not present in the data and so the assumption of no reciprocity is reasonable.

Results

As in Study 2, the design is a round-robin, and group is the unit of analysis. Because group size varies, we weighted by that factor. The degrees of freedom for significance testing are 16.

The variance partitioning for Study 3 is presented in Table 4. At zero acquaintance, perceiver variances are generally small and nonsignificant. This lack of response set is most likely due to the fact that at Wave 0, subjects were encouraged to differentiate the targets. That is, subjects were asked to use the entire scale range when rating members of their group. Significant perceiver variance was found only for the Emotional Stability factor.

Only the Extraversion factor showed significant consensus effects. At zero acquaintance, 40% of the variance in ratings of Extraversion was target-based. People do seem to arrive at similar first impressions of a person's degree of sociability.

Finally, for Wave 0, approximately 25% of the variance in ratings was relationship variance. For all but the Culture factor, the relationship variances were significant.

At the end of the semester, Wave 1, significant perceiver variance emerges. It seems apparent that the message to differentiate targets and use the entire range of the scale was not heeded at Wave 1, and therefore, response set played a more substantial role in the ratings. Unlike Study 2, the consensus effect for Extraversion remained virtually stable over time and actually increased for the Agreeableness and Culture factors. Changes in relationship variance did not show any consistent pattern over time.

Physical and Behavioral Cue Results

There are two ways in which the studies in this article examine possible cues that lead to consensus in first impressions. In Study 1, 4 independent raters viewed the stimulus tapes and rated each target on a set of behavioral variables, such as smiling and body movement. For these behavior ratings, 10 had sufficient interrater reliability (alpha at least .70) to be used in subsequent analysis. The 6 variables that were dropped were frowning, downward gaze, information-seeking gaze, slow movement, hair groomed, and fashionable dress. (We are not considering attractiveness as it is discussed later in conjunction with attractiveness ratings for Studies 2 and 3.) These ratings were then correlated with the target effect estimates on the five Norman (1963) factors. The results of these analyses are presented in Table 5.

Table 4
Relative Variances for Study 3

Factor	Perceiver	Target	Relationship	Error
Wave 0				
Extraversion	.06	.40*	.27*	.28
Agreeableness	.06	.01	.20*	.73
Conscientiousness	.00	.09	.24*	.67
Emotional stability	.17*	.00	.22*	.61
Culture	.05	.02	.05	.88
Wave 1				
Extraversion	.18*	.37*	.11*	.33
Agreeableness	.28*	.13*	.19*	.40
Conscientiousness	.35*	.11	.07	.47
Emotional stability	.14	.06	.46*	.34
Culture	.30*	.07*	.08	.56

* $p < .05$, $df = 16$.

Table 5
Mediators of Target Effects in Study 1

Factor	Mediator	r
Extraversion	Smiling	.49
	Rapid body movement	.47
Agreeableness	Smiling	.66
Conscientiousness	Formal dress	.44
	Rapid body movement	−.35
Emotional stability	Rapid body movement	−.57
	Primping	−.40
Culture	Smiling	.37

Note. For all correlations, $p < .005$, $df = 62$.

Only correlations statistically significant at the .005 level are presented to correct for inflation of the Type I error rate due to performing multiple tests. If the 10 correlations for each factor were independent, using an alpha of .005 for each test, the experimentwise error rate would be approximately .05.

Significant effects were found for only four of the behavioral variables. These variables are rapid body movement (interrater reliability = .81), smiling (reliability = .86), formal dress (reliability = .72), and primping (reliability = .88). The two variables that correlated with ratings of Extraversion were rapid body movement and smiling; judges tended to rate targets who moved about frequently and who smiled as extraverts. Smiling was also positively related to ratings of Agreeableness and Culture. A negative relationship between rapid body movement with both Emotional Stability and Conscientiousness emerged such that targets who were relatively stationary were rated as more emotionally stable and conscientious than those who were more mobile. Primping was also found to correlate negatively with Emotional Stability. Formal dress related positively to Conscientiousness. Targets who were formally dressed were rated as more conscientious.

Studies 1, 2, and 3 all addressed the role of physical attractiveness in ratings on the Extraversion factor. Albright et al. (1988) found that much of the variation in ratings of Extraversion can be explained by the physical attractiveness of the target. The physical attractiveness correlations are presented in Table 6. Because these results are replications, we use the conventional .05 level of significance.

For Study 1, the ratings of physical attractiveness were made by the 4 independent raters (interrater reliability = .68). In Studies 2 and 3, the physical attractiveness ratings were made by the participating subjects themselves at the time that they made

Table 6
Correlations of Target Effects in Extraversion With Physical Attractiveness and Self-Judgments in Extraversion

Study by wave	Physical attractiveness	Self
Study 1	.36*	.03
Study 2		
Wave 0	.39**	.33**
Wave 1	.42	.57**
Study 3		
Wave 0	.59*	.62*
Wave 1[a]	.41**	.66*

[a] Physical attractiveness ratings and self-ratings were made at Wave 1.
* $p < .05$. ** $p < .10$.

their ratings on the Norman (1963) factors. In these two studies, subjects rated every other person in the group (as well as themselves) on the Norman factors as well as on physical attractiveness. Across the three studies, the zero-acquaintance correlations between Extraversion and physical attractiveness average to .45, a value somewhat lower than the .74 value observed by Albright et al. (1988). However, Study 3, which is most similar to the Albright et al. study, has the largest correlation. At Wave 1 in Studies 2 and 3, the attractiveness–extraversion correlations are not significant, but they are still moderate to large in size. Despite their lack of statistical significance, there is little or no decline in the correlation between attractiveness and Extraversion from Wave 0 to Wave 1.

The three studies also examined the relationship between self-ratings on Extraversion and target effects on Extraversion. Table 6 presents these correlations. The self-correlations provide some evidence that people who consider themselves to be extraverted are rated as such. However, the correlations vary quite a bit from study to study. Of particular note, the correlation between self-ratings and target effects on Extraversion is essentially zero for Study 1. In addition, in Study 1, we asked targets how easy it was for them to make friends and how active their social lives were. Neither of these variables correlated with the target effect in Extraversion.

We also correlated target effects with gender in Studies 1 and 3. (Recall that in Study 2, all the subjects were women.) The factor with the strongest sex correlation in both studies was Conscientiousness (Study 1: $r = .36$; Study 3: $r = .29$). Female targets were seen as more conscientious than were men.

Finally, we correlated target effects in Culture with self-reported grade point average and math and verbal Standardized Achievement Test scores in Study 1. None of these correlations were statistically significant.

Stability Results for Studies 2 and 3

There are two ways in which stability can be addressed in these studies. The first way was discussed in the *Results* sections for Studies 2 and 3. There the issue of stability in the level of consensus concerns whether consensus effects occurred at both zero acquaintance and then after interaction.

The second measure of stability, stability in impressions, is presented in Table 7. Table 7 contains the correlations between Wave 0 and Wave 1 of perceiver, target, and relationship effects. To avoid anomalous correlations, both variables in the correlation had to explain at least 10% of the total variance. The target stability correlation for the Extraversion factor in Study 2 ($r = .89$) indicates that people who were originally rated as extraverted at zero acquaintance were also rated as extraverted after one-on-one interaction. In Study 3, the target Extraversion correlation is also high ($r = .72$), but is somewhat smaller than for Study 2.

Study 2 also indicates a high degree of stability in perceiver effects such that a perceiver who tended to rate most targets as conscientious at Wave 0 also tended to rate these subjects as conscientious at Wave 1. Because there were not large amounts of perceiver variance at Wave 0 in Study 3, we could not examine that stability.

Little stability was found for the relationship component. Only the Emotional Stability factor showed statistically significant stability in Study 3. Evidently the idiosyncratic perceptions that judges have of a target at zero acquaintance do not persist over time.

General Discussion

Level of Consensus

The three studies presented have provided a variety of information concerning consensus at zero acquaintance. Study 1 clearly indicates that consensus at zero acquaintance is not an artifact of prior acquaintance or interaction during the rating session. The level of consensus using videotaped subjects as targets showed the same pattern of consensus results as in face-to-face classroom studies. The amount of target variance was somewhat smaller in the videotape study relative to the face-to-face studies but not appreciably so.

The profile of consensus across the five Norman (1963) factors is remarkably consistent. The consensus profile from these three studies is very close to the pattern from the four previous studies summarized in Table 1. Data have now been collected

Table 7
Stabilities of Components for Studies 2 and 3

Factor by study	Perceiver	Target	Relationship
Study 2			
Extraversion	—	.89*	.23
Agreeableness	.42*	—	.54
Conscientiousness	1.00*	—	−.01
Emotional stability	.52*	—	—
Culture	.61*	—	—
Study 3			
Extraversion	—	.72*	.19
Agreeableness	—	—	.38
Conscientiousness	—	—	—
Emotional stability	—	—	.56*
Culture	—	—	—

Note. The dash (—) indicates less than 10% of the variance.
* $p < .05$.

from about 900 subjects, and to a large extent, the relative level of consensus across the five factors can be predicted.

In seven different studies, Extraversion has been shown to exhibit the highest level of agreement. Rapid body movements were found to correlate with judgments of Extraversion. Allport and Vernon (1933) speculated that expansiveness or emphasis of movement is an important cue in personality judgments, particularly extraversion. Eysenck (1967) argued that extraverts need higher levels of stimulation than introverts. Research in nonverbal communication has consistently shown that extraverts are more active and expressive than introverts (e.g., Riggio & Friedman, 1983). People seem to know implicitly that body movement is correlated with extraversion. In this way, they can infer extraversion from observing a person sitting alone in a room.

Our replication of the physical attractiveness–extraversion correlation supports a stereotype explanation of consensus. A recent meta-analysis by Eagly, Ashmore, Makhijani, and Longo (1991) showed that the physical attractiveness stereotype primarily affects judgments concerning social behaviors. Physical attractiveness is not a global stereotype.

Why is there greater consensus at zero acquaintance in Extraversion than there is on the other four factors? First, judges, in perceiving others, seek out extraversion information because it is useful in guiding the course of interaction. Second, the cues that are used in determining extraversion appear to be directly observable and thus do not need to be inferred. Third, the judges' rules for combining cues to form a judgment about the target's extraversion are shared. Fourth, the cues that the target emits to the judges are invariant across time.

Consensus on Agreeableness is most likely largely based on facial cues. Using photographs, Berry (1990) showed that people judged as baby-faced are considered to be warm. Our Study 1, which carefully controlled the presentation of target stimulus, revealed relatively high levels of consensus on Agreeableness. Perhaps the reason why consensus is relatively low in the prior studies is that for the most part, targets' faces were not directly visible to the judges because the targets themselves were busy filling out rating forms. In addition, in Study 1, all judges viewed the same stimulus, whereas in the face-to-face studies, judges rated the same target at different times.

In Study 1, there was a considerable correlation between Agreeableness and Extraversion ($r = .58$, $p < .001$). To some extent this correlation is due to the common cue of smiling, which was correlated with both factors. One might expect that if only photographs were presented to subjects, the Extraversion and Agreeableness factors would collapse into one factor. Certainly, the rapid body movement cue that predicts Extraversion and not Agreeableness would be lost by such a presentation.

Consensus for Conscientiousness appears to be due to consensus on the cue of neatness in dress and grooming. Study 1 replicates the correlation between conscientiousness and dress that was previously found by Albright et al. (1988). However, as pointed out by Watson (1989), consensus at zero acquaintance varies quite a bit for the different Conscientiousness indicators. For honest–dishonest, there is virtually no agreement, whereas for fussy–careless (used in Study 3 of Albright et al., 1988), agreement is much higher. Presumably, elements of Conscientiousness that are tied to dress and grooming will show consensus at zero acquaintance, and those not tied to dress will not.

The amount of consensus for Emotional Stability is variable from study to study. The factor that may bring about this variation is the degree to which the situation creates relatively high levels of anxiety. Situations that are relatively anxiety-provoking (e.g., waiting in a room alone for some unknown experimenter to appear) might bring about varying levels of anxiety. However, if the situation is quite relaxed (e.g., a graduate student handing out clearly anonymous forms), the subjects will show relatively little anxiety.

What cues signal the Emotional Stability factor? The two cues suggested by Study 1 were rapid body movement and primping. Targets who made more body movements and primped more were perceived to be low on the Emotional Stability factor. Recent work by Zuckerman, Hodgins, and Miyake (1990) seems to indicate that Emotional Stability is communicated through the vocal channel by what they called *vocal attractiveness.* Interestingly, the one zero-acquaintance study that required subjects to verbalize (Watson, 1989) has the greatest consensus for the Emotional Stability factor.

No consistent consensus results were found for the Culture factor in the different studies. The studies in this article as well as those of Albright et al. (1988) showed low levels of consensus on this factor. These studies used *intelligent* and *imaginative* as indicators of the Culture factor, and these two traits may not be measuring precisely the same factor. If the two indicators of a factor do not both measure the same factor, a social relations analysis underestimates the judge, target, and relationship variances. Note also that this factor shows low levels of relationship variance, a result consistent with the view that the two indicators measure different factors.

There is some indication that the level of consensus for Culture is higher for groups attending more competitive, achievement-oriented universities. The greatest level of consensus that has been found for this factor is in the two studies that used students at private colleges (Watson's [1989] study and Study 1 of Albright et al., 1988). It may be the case that consensus on the Culture factor is higher in groups that place a great value on intelligence. Finally, as in Park and Judd (1989), in Study 1, ratings of Conscientiousness and Culture were positively correlated ($r = .55$).

Study 2 showed that the level of consensus after zero acquaintance declines when the interactions are one-on-one, but as demonstrated by Study 3, the level of consensus remains relatively stable (and in some cases even increases) when the interactions are in a group. Although there are several differences between Studies 2 and 3 (e.g., the length of time between the measurements), the decline in consensus in Study 2 may be due to the subjects basing their ratings on the individualized information learned through the one-on-one interactions. In Study 3, interactions were in a group, and so all perceivers had access to approximately similar information and therefore were rating a common stimulus (Kenny, 1991).

Validity of Judgment at Zero Acquaintance

We have avoided stating the extent to which we feel that consensus at zero acquaintance reflects accuracy. To date, the most

impressive evidence that seems to show that consensus reflects accuracy is the correlation between stranger ratings and self-ratings. Norman and Goldberg (1966), Albright et al. (1988), and Watson (1989) have all shown correlations between ratings by strangers and self-ratings for both Extraversion and Conscientiousness. We replicate these results in two of the three studies. However, Study 1 fails to replicate the other studies. One might try to explain the failure to replicate as a chance event, but because Study 1 carefully controlled the presentation of stimulus information, one should not be quick to dismiss it.

A second source of evidence concerning accuracy is the stability of judgments at zero acquaintance, particularly the stability of the target effect. If target effects are stable, then it seems more reasonable to argue that they are accurate. In both Studies 2 and 3, very strong correlations were found between the target effect in Extraversion at zero acquaintance and the target effect after interaction. Such a result seems very surprising. If judgments at zero acquaintance are based only on stereotypes and if these stereotypes are not valid, then the correlation should be near zero. How can this correlation be explained?

Consider the model in Figure 1. This figure is an adaptation of a model recently proposed by Jussim (1991) to explain expectancy effects. A target is judged at Time 0 (zero acquaintance) and at Time 1 (after interaction). The Time 0 judgment is based on a set of cues (e.g., gender, age, grooming, body movement) with which there are associated stereotypes. At Time 1, the judgment is based on the same set of cues as well as the actual behavior of the target. (To simplify matters, we have not drawn the self-fulfilling prophecy link from Time 0 judgment to behavior.) Stability can occur for one of two reasons. First, because target judgments are determined by cues at both times, the stereotype operating at both times could bring about stability. Second, given the correlation between the cues and the behavior, it is possible that "a kernel of truth" leads to stability. Our data do not allow us to separate these two different explanations.

Relationship Effects

Although the primary focus of these studies was to examine target effects, all three studies do provide important data about relationship effects. There is significant relationship variance at zero acquaintance. Besides error, the relationship component is the largest source of variance, accounting for about 25% of the total variation. However, relational first impressions are not lasting impressions. If you see a person across the room and you presume that the person has a certain personality, your unique impression of that target will radically change once you interact with the target. So first impressions, at least the idiosyncratic component, change over time. Counterintuitively, consensual impressions of extraversion change relatively little over time.

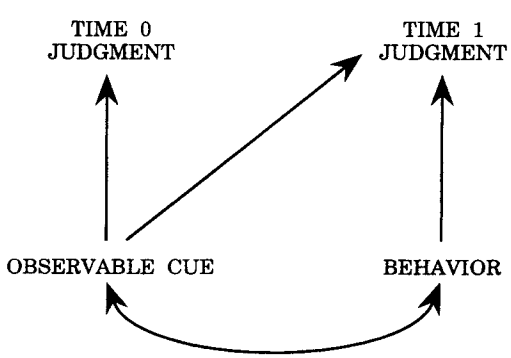

Figure 1. Over-time effects of stereotypes and behavior on the judgments at zero acquaintance (Time 0) and after interaction (Time 1).

Concluding Comments

About 25 years ago, Norman and Goldberg (1966) presented the surprising fact that judges agree in their assessments of a target's personality even if the judges have not interacted with that target. Even more surprising these snap judgments may indeed be accurate. In the past few years, we have begun to probe the Norman and Goldberg finding. We now know which traits consistently show consensus (Extraversion and to a lesser extent Conscientiousness), and we are beginning to understand the cues that are used to bring about this consensus. We have learned in this study that a consensual part of the judgment made at zero acquaintance is highly stable after the judge and target have interacted.

Although much has been learned, a number of important questions remain. First, to what extent will these results replicate in other populations? In particular, will the results be the same for children and older adults and in other cultures? Second, we are now positioned to examine more carefully the issue of accuracy. That is, does the consensus that observers show at zero acquaintance actually predict the behavior of targets in social interaction?

Future work in the study of zero acquaintance will sometimes require restricting the interaction between judge and target interaction, much as was done in Study 1. We would hope—following the suggestion of Kenny and Albright (1987)—that the study of interpersonal perception will also include studies involving 2 interacting people. It is only by moving back and forth between controlled laboratory ratings and naturalistic interactions between people in relationships that the process of interpersonal perception will be elaborated.

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.

Berry, D. S. (1990). Taking people at face value: Evidence for the kernel of truth hypothesis. *Social Cognition, 8,* 343–361.

Crowne, D. P., & Marlowe, D. (1964). *The approval motive: Studies in evaluative dependency.* New York: Wiley.

DePaulo, B., Kenny, D. A., Hoover, C., Webb, W., & Oliver, P. V. (1987). Accuracy of person perception: Do people know what kind of impressions they convey? *Journal of Personality and Social Psychology, 52,* 303–315.

DiPilato, M. (1989). *Person perception in initial acquaintanceship: Changes over time and task-trait match* (Doctoral Dissertation, Arizona State University, 1989). *Dissertation Abstracts International, 51,* 1038.

Dornbusch, S. M., Hastorf, A. H., Richardson, S. A., Muzzy, R. E., & Vreeland, R. S. (1965). The perceiver and the perceived: Their relative influence on categories of interpersonal perception. *Journal of Personality and Social Psychology, 1,* 434–440.

Eagly, A. H., Ashmore, R. D., Makhijani, M. G., & Longo, L. C. (1991). What is beautiful is good, but . . . : A meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin, 110,* 109–128.

Eysenck, H. J. (1967). *The biological basis of personality.* Springfield, IL: Thomas.

Funder, D. C., & Dobroth, K. M. (1987). Differences between traits: Properties associated with interjudge agreement. *Journal of Personality and Social Psychology, 52,* 409–418.

Jussim, L. (1991). Social perception and social reality: A reflection–construction model. *Psychological Review, 98,* 54–73.

Kenny, D. A. (1988). Interpersonal perception: A social relations analysis. *Journal of Social and Personal Relationships, 5,* 247–261.

Kenny, D. A. (1990). Design and analysis issues in dyadic research. In C. Hendrick & M. S. Clark (Eds.), *Review of personality and social psychology: Research methods in personality and social psychology* (Vol. 11, pp. 164–184). Newbury Park, CA: Sage.

Kenny, D. A. (1991). A general model of consensus and accuracy in interpersonal perception. *Psychological Review, 98,* 155–163.

Kenny, D. A., & Albright, L. (1987). Accuracy in interpersonal perception: A social relations analysis. *Psychological Bulletin, 102,* 390–402.

Kenny, D. A., & La Voie, L. (1984). The social relations model. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 18, pp. 142–182). San Diego, CA: Academic Press.

Kenny, D. A., & Malloy, T. E. (1988). Partner effects in social interaction. *Journal of Nonverbal Behavior, 12,* 34–57.

Kenrick, D. T., & Funder, D. C. (1988). Profiting from controversy: Lessons from the person–situation debate. *American Psychologist, 43,* 23–34.

Malloy, T. E., & Albright, L. (1990). Interpersonal perception in a social context. *Journal of Personality and Social Psychology, 58,* 419–428.

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.

Park, B., & Judd, C. (1989). Agreement on initial impressions: Difference due to perceivers, trait dimensions, and target behaviors. *Journal of Personality and Social Psychology, 56,* 493–505.

Riggio, R. E., & Friedman, H. S. (1983). Individual differences and cues to deception. *Journal of Personality and Social Psychology, 45,* 899–915.

Snyder, M. (1974). The self-monitoring of expressive behavior. *Journal of Personality and Social Psychology, 30,* 526–547.

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., Hodgins, H., & Miyake, K. (1990). The vocal attractiveness stereotype: Replication and elaboration. *Journal of Nonverbal Behavior, 14,* 97–112.

Received September 18, 1990
Revision received August 19, 1991
Accepted August 21, 1991 ■

Zahn-Waxler Appointed New Editor, 1993–1998

The Publications and Communications Board of the American Psychological Association announces the appointment of Carolyn Zahn-Waxler as editor of *Developmental Psychology.* Zahn-Waxler is associated with the National Institute of Mental Health. As of January 1, 1992, manuscripts should be directed to

Carolyn Zahn-Waxler
4305 Dresden Street
Kensington, Maryland 20895

Manuscript submission patterns make the precise date of completion of the 1992 volume uncertain. The current editor will receive and consider manuscripts through December 1991. Should the 1992 volume be completed before that date, manuscripts will be redirected to the incoming editor for consideration in the 1993 volume.