

Naturalistic Social Cognition: Empathic Accuracy in Mixed-Sex Dyads

William Ickes, Linda Stinson, Victor Bissonnette, and Stella Garcia
University of Texas at Arlington

The research paradigm that Ickes and his colleagues developed for the study of naturalistic social cognition was used to explore the phenomenon of empathic accuracy in the initial, unstructured interactions of 38 mixed-sex (male–female) dyads. The results indicated that an important aspect of empathic accuracy—content accuracy—could be measured reliably ($\alpha = .94$) with the procedure used. The results further indicated that content accuracy was, to a large extent, an emergent product of social interaction processes occurring at the level of the dyad. Although many of the findings could be explained in informational terms, some significant motivational influences were observed as well. For example, content accuracy was influenced by the partner's physical attractiveness and, more generally, by the perceiver's apparent interest in the partner (as indexed by various thought/feeling measures). The individual difference variables of grade point average and self-monitoring also predicted the subjects' levels of content accuracy; however, gender and self-report measures of empathic skills and empathic accuracy did not.

If the study of *subjective* phenomena involving or occurring within a single conscious mind is the domain of mainstream cognitive psychology, it follows logically that the study of *inter-subjective* phenomena involving or occurring between at least two conscious minds is the proper domain of cognitive social psychology (Ickes, Tooke, Stinson, Baker, & Bissonnette, 1988). The logic of this conclusion has been consensually validated by both present and past reviewers of social cognition research. For example, Markus and Zajonc (1985) ended their *Handbook* chapter on cognitive social psychology by stating that “the properties of social perception and social cognition that make them distinct are reciprocity and intersubjectivity” (p. 213). They noted that “many earlier authors, such as Mead (1934), Merleau-Ponty (1970), Asch, (1952), and Heider (1958),” have drawn essentially the same conclusion.

In general, intersubjective phenomena can be characterized as those involving some form of interdependence between the contents or processes of at least two conscious minds (cf. Wegner, Giuliano, & Hertel, 1985). Given this definition, intersubjective phenomena are clearly not the most frequently studied phenomena in cognitive social psychology (Ickes et al., 1988). In most studies of human social cognition, researchers have not inquired how the contents and processes of one mind are interdependent with those of another. Instead, using as their models the studies conducted in more traditional areas of psychology, they have inquired how the contents and processes of

individual minds tested “one at a time,” are related to “social” stimulus materials whose features and contents have been predetermined by the experimenter.

There are, however, some important and noteworthy exceptions to this rule. Group dynamics researchers have for some time studied intersubjective phenomena under headings such as brainstorming (Diehl & Stroebe, 1987; Street, 1974), group decision making (Janis & Mann, 1977; Miller, 1989; Stasser, Kerr, & Davis, 1989) groupthink (Janis, 1972; McCauley, 1989), group polarization (Moscovici & Zavalloni, 1969; Myers, 1982; Myers & Lamm, 1976), group socialization (Moreland & Levine, 1982; 1989) and majority and minority influence (Latané & Wolfe, 1981; Maass, West, & Cialdini, 1987; Moscovici & Mugny, 1983; Nemeth, 1986). More recently, social cognition researchers have begun to address intersubjective phenomena such as consensus and meta-accuracy in person perception (Funder, 1980, 1987; Funder & Harris, 1986; Kenny & Albright, 1987; Malloy & Albright, 1990), dyadic intersubjectivity (Ickes et al., 1988; Krauss & Fussell, 1989), and transactive memory (Wegner et al., 1985).

We believe that intersubjective phenomena are of central importance to the study of social cognition. To facilitate research on these phenomena, we recently extended the original dyadic interaction paradigm (Ickes, 1982, 1983) to permit the study of both the overt behavior and the covert thoughts and feelings of dyad members during a period of unstructured interaction (Ickes, Bissonnette, Garcia, & Stinson, 1990; Ickes, Robertson, Tooke, & Teng, 1986; Ickes & Tooke, 1988). By adding thought/feeling assessment procedures of the type developed by researchers at Ohio State University (e.g., Brock, 1967; Cacioppo, Glass, & Merluzzi, 1979; Cacioppo & Petty, 1981; Greenwald, 1968; Petty & Cacioppo, 1979), we are now able to study a wide range of phenomena that can be classed under the general heading of *naturalistic social cognition*.

Empathic Accuracy

In the present study, the dyadic interaction paradigm was further extended to permit the study of another important in-

We would like to express our appreciation to Dorothy Anderson, Kristine Arnold, Ray Bristow, Susan Ferry, Jenny Gatlin, Andrea DeBolt, Mujahid Khan, Laurie Luce, Carol Marangoni, Marianne Mickan, Keith Milstead, and Jacquetta Talley for their assistance in coding and analyzing the data from this study. We also wish to thank David Kenny for his suggestions regarding data analysis and David Funder, Tom Malloy, Carol Marangoni, Paul Paulus, and Brian Spitzberg for their constructive comments on a draft of this article.

Correspondence concerning this article should be addressed to William Ickes, Department of Psychology, University of Texas, Arlington, Texas 76019.

tersubjective phenomenon that our research had not previously addressed: *empathic accuracy*.

Empathic accuracy is an intersubjective phenomenon that falls within the more general domain of interpersonal perception and has perhaps the longest history of interdisciplinary study. In clinical and counseling psychology, empathic accuracy has been widely regarded as one of the most important nonspecific factors common to all successful therapeutic outcomes—see, for example, Rogers (1957), Barrett-Lennard (1962), Truax and Carkhuff (1967), Gladstein (1977), Goldstein and Michaels (1985), and Marks and Tolsma (1986). In personality psychology, a valid self-report measure of empathic accuracy has been the elusive goal of several generations of researchers—see, for example, Chapin (1942), Dymond (1949), Kerr and Speroff (1954), Crow and Hammond (1957), Cline (1964), Hogan (1969), Campbell, Kagan, and Krathwohl (1971), Danish and Kagan (1971), Mehrabian and Epstein (1972), Greif and Hogan (1973), Davis (1983), and Funder and Harris (1986). And in communication studies, empathic accuracy has been viewed as an element that is essential to effective communication—see, for example, Mehrabian and Reed (1968), Pearce (1976), O'Keefe and Sypher (1981), and Honeycutt, Knapp, and Powers (1983).

Several terms have been used in these literatures as synonyms or near synonyms of empathic accuracy (Marangoni, 1989). These include accurate empathy (Rogers, 1957), discrimination (Carkhuff, 1969), perceptual accuracy (Goldstein & Michaels, 1985), social acuity (Funder & Harris, 1986), and social intelligence (Sternberg, 1985; Walker & Foley, 1973). As one would expect, conceptual definitions of the construct reveal the inevitable variations across writers on what appears to be the same common theme. The common elements in these definitions suggest that empathic accuracy is (a) a skill, ability, or facility (b) to understand, apprehend, infer, interpret (c) with accuracy (d) the private, covert, subjective (e) phenomenological reality, mental experience, thoughts and feelings (f) of some other person(s).

In a social interaction context, empathic accuracy can be defined more simply as the degree to which one interactant is able to accurately infer the specific content of another interactant's thoughts and feelings. Although there are a number of different aspects of empathic accuracy that one might choose to study, the two aspects we regard as most important are content accuracy and valence accuracy. *Content accuracy* refers to the degree to which one interactant's description of the inferred content of an interaction partner's thoughts or feelings matches the actual content of the specific thoughts or feelings reported by the partner. *Valence accuracy* refers to the degree to which one interactant's inferences about the emotional tone (positive, neutral, negative) of an interaction partner's thoughts or feelings match the actual valence of the specific thoughts and feelings reported by the partner.

Overview of the Present Study

To study these aspects of empathic accuracy in the context of naturally occurring social interaction, it was necessary to further modify the dyadic interaction paradigm so that dyad members could report not only their own thoughts and feelings,

but their inferences about their partner's thoughts and feelings as well. This procedure, which is described in detail in the Method section of this article, can be summarized briefly as follows. The members of each dyad were led into a waiting room and left there together in the experimenter's absence. During this interval in which the subjects were ostensibly waiting for the experiment to begin, their verbal and nonverbal behaviors were unobtrusively audio- and videotaped. When the experimenter returned at the end of the observation period, the subjects were partially debriefed. They then were asked for their signed consent that the tape could be released for use as data and that they would participate in the second part of the study, which concerned their thoughts and feelings during the interaction.

If consent was given, the subjects were seated in separate but identical cubicles where they were each instructed to view a videotaped copy of the interaction. By stopping the videotape with a remote start/pause control at those points at which they remembered having had a specific thought or feeling, each subject made a written, time-logged listing of those *actual* thought/feeling entries. The subjects were then instructed to view the videotape a second time, during which the tape was stopped for them at each of those points at which their interaction partner reported a thought or feeling. The subject's task during this viewing of the tape was to infer the content of their partner's thoughts and feelings and provide a written, time-logged listing of those *inferred* thought/feeling entries. When both subjects had completed this task and filled out a short posttest questionnaire, they were debriefed more completely and then thanked and released.

Consistent with the exploratory nature of the study, we deliberately assessed a wide range of variables that potentially could serve as predictors of content and valence accuracy. These included various measures of the dyad members' (a) verbal and nonverbal interaction behavior, (b) the content of their own reported thoughts and feelings, (d) their posttest perceptions of the interaction they experienced, and (e) their responses on personality measures of shyness (Cheek & Buss, 1981), self-monitoring (Snyder, 1974), empathic ability (Davis, 1983) and empathic accuracy (Ickes, 1988). Current grade point average (GPA) scores and ratings of physical attractiveness were also obtained for each of the subjects in the study, because empathic accuracy could plausibly be determined in part by both the intelligence of the perceivers and the attractiveness of their "target" partner.

Balanced against the substantial costs involved in such an ambitious data collection effort was the possibility of achieving some fundamental insights about the processes underlying empathic accuracy in naturally occurring social cognition. With respect to the initial unstructured interactions between college-age strangers of the opposite sex that were the focus of the present investigation, we sought to address a range of questions that included the following.

1. Can the major aspects of empathic accuracy—content accuracy and valence accuracy—be measured reliably?
2. Is empathic accuracy related to attributional inference? In other words, is the accuracy of one's inferences about the specific content of an interaction partner's thoughts and feelings

related to one's tendency to make more global inferences about the partner's traits and dispositions?

3. To what extent does empathic accuracy depend on processes occurring at the individual level of analysis, as opposed to those occurring at the dyad level of analysis (cf. Kenny & La Voie, 1985)?

4. Is empathic accuracy influenced by motivational, as well as by informational, factors?

5. Is empathic accuracy reliably predicted by individual differences in gender or by self-reported differences in empathic skills and dispositions?

Method

Subjects

The subjects were 38 female and 38 male undergraduates enrolled in introductory psychology classes at the University of Texas at Arlington. These subjects had completed Davis's (1983) measure of empathic abilities and Ickes's (1988) measure of empathic accuracy during a general pretesting at the start of the semester. Separate sign-up sheets were later used to schedule the nonpreselected male and female volunteers, whose pairing into mixed-sex dyads was essentially random. Mixed-sex dyads were used for reasons of economy; they allowed us to study the responses of both men and women using only a single dyad type.

Setting and Equipment

The observation room in which the dyad members' interaction was recorded was furnished with a long couch and an accompanying coffee table that concealed an FM wireless microphone. A video camera focused on the area of the couch and coffee table was concealed in a darkened storage room across a hallway from the observation room. This arrangement permitted the dyad members' interaction to be recorded through the open doorways of the storage and observation rooms with minimal likelihood of the camera being detected.

Adjacent to the observation room was a control room used to house the video and audio equipment operated by the experimenter and a research assistant. Along one wall of the control room was the experimenter's control station. Here the experimenter could sit in front of a single table that supported two identical videocassette recorders (VCRs) that were both connected (by means of a Y-adaptor) to the video camera in the storage room across the hall. Each VCR was also connected, respectively, to one of two 25 inch color TV monitors that were oriented to face through one-way mirrors into identical test cubicles. The test cubicles were each equipped with a help button that allowed subjects to signal the experimenter (by means of a red light) that they needed additional information or assistance of some other type. Each test cubicle also contained (a) a remote start/pause switch connected to the VCR and TV monitor system unique to that cubicle and (b) a supply of thought/feeling coding forms (to be described). For a schematic diagram of the laboratory and its arrangement, see Ickes et al. (1986, p. 68).

Procedure

Following the directions given on their respective sign-up sheets, the two subjects who were scheduled for each session reported to different waiting areas in the psychology building. This separation helped to ensure (a) that two friends or previously acquainted people would not sign up for the same session, and (b) that each scheduled pair of subjects would not meet and interact before their session began.

Collection of the videotape data. After meeting the subjects and be-

ing satisfied that they were indeed strangers, the experimenter escorted them into the observation room and directed them to be seated on the couch.¹ The experimenter then pretended to discover that she had only one copy of the informed consent form. Promising to return in a few minutes, she left the observation room to get additional copies. At this point, a research assistant in the control room activated the video equipment to begin videotaping. Exactly 6 min later, the experimenter returned to the observation room, and the videotaping was terminated.

After probing for any evidence of suspicion, the experimenter conducted a partial debriefing and informed the subjects that their written consent was required for the tape to be used as data.² She further explained that if one or both of them exercised the right to have the tape erased immediately, instead of releasing it for use as data, the content of the interaction would remain their own private concern. (None of the subjects requested to have the tape erased.)

Collection of the thought/feeling data. Once the subjects' written consent had been obtained, the experimenter described the next part of the study. In this part, the subjects were asked to view the videotape and make a written record of the thoughts and feelings they experienced during the 6-min interaction period. When the subjects had been seated in their individual cubicles, the experimenter used video instruction tapes to give them detailed instructions regarding the thought/feeling data collection procedure. The two VCRs were then activated, and the subjects used their start/pause controls to independently start and stop their respective copies of the videotape.

Consistent with the instructions they received, the subjects stopped the videotape at each point during the interaction that they recalled having had a specific thought or feeling. They recorded each thought or feeling on a thought/feeling coding form by entering (a) the time the thought or feeling occurred (available as a digital readout on the upper-left part of the video image), (b) whether the entry was a thought or a feeling (coded "I was thinking:" or "I was feeling:" on the coding sheet), (c) the specific content of the thought/feeling entry (written in sentence form to complete the initial phrase), and (d) whether the entry was presumed to be positive, negative, or neutral in its overall affective tone (coded +, -, or 0 on the coding sheet).

The instructions the subjects received explicitly encouraged them to report all of the thoughts and feelings they remembered having had as accurately and honestly as possible. The importance of being completely candid was emphasized, and the subjects were assured that the thoughts and feelings they recorded would not be seen by their interaction partners. On the other hand, the instructions cautioned them to report only those thoughts and feelings they distinctly remembered having had during the observation period. They were not to report any

¹ If the subjects in a given session displayed any sign that they were already acquainted with each other, the experimenter cited equipment problems and gave each subject an experimental point credit without requiring any further participation in the study. If the subjects were strangers, the experimenter continued the procedure without introducing them to each other or providing any instruction or encouragement for them to interact or get to know each other.

² Four dyads were dropped from the study after the probe for suspicion revealed that one or both of the dyad members had located the video camera while being taped. This level of suspicion was so unusual in our research that we were motivated to investigate it further. In doing so, we found that a common element in three of the four cases was attendance at a lecture given by an introductory psychology instructor who discussed the unobtrusive measurement of behavior by means of a concealed video camera. With the instructor's cooperation, this lecture is no longer given during the first several weeks of the semester.

thoughts and feelings that occurred to them for the first time while viewing the videotape.

Collection of the empathic accuracy data. When both subjects had completed their thought/feeling listings, the experimenter collected these data and then initiated a new set of videotaped instructions, which informed the subjects that they would now view the videotape a second time. On this second viewing, however, each subject would no longer use his or her remote start/pause control to start and stop the tape. Instead, the experimenter (aided by a research assistant) would start and stop each subject's copy of the videotape at the appropriate times.

The experimenter and the research assistant independently presented their respective copies of the tape to the two dyad members, stopping them at each of those points in time at which the partner's thoughts and feelings were recorded. The subjects' instructions were in most respects analogous to those of the previous thought-listing task. Each time the tape was stopped, the subjects were required to make an inference about the nature of the specific thought or feeling reported by their partner at that point. Then, using the same type of coding form on which they had previously recorded their own thoughts and feelings, the subjects coded the inferred content of their partner's thoughts and feelings by entering (a) the time the thought or feeling occurred, (b) whether the entry was presumed to be a thought or a feeling (coded "he/she was thinking:" or "he/she was feeling:" on the coding sheet), (c) the specific content of the thought or feeling (again recorded in sentence form), and (d) whether the entry was presumed to be positive, neutral, or negative in its overall affective tone (again coded as +, 0, or -).

Collection of the final self-report measures. In the final part of the study, the subjects were asked to remain in their respective cubicles long enough to complete a posttest questionnaire. This questionnaire included personality measures of shyness (Cheek & Buss, 1981) and of self-monitoring (Snyder, 1974). It also included items that assessed the subjects' perceptions of (a) the quality of their interaction and (b) the degree to which they liked each other. When each subject had completed the questionnaire and returned it to the experimenter, the subjects were then debriefed more fully and were asked not to discuss the study with potential future subjects. They were then thanked, given point credit for their participation, and released.

Behavioral Measures

The behavioral measures relevant to this study were coded from the videotapes by independent judges who were kept unaware, as much as possible, of any predictor variables being studied. These behaviors included the total frequency and duration of verbalizations (i.e., speaking turns), directed gazes, mutual gazes, expressive gestures, and expressions of positive affect (i.e., smiles and laughter). The interrater reliabilities of these measures ranged from .75 to .99, with a mean of .90. (For operational definitions of the measures, see Ickes & Turner, 1983, pp. 214-215.)

Personal pronoun usage was also coded from the videotapes by two independent judges who used tally marks to record the number of first-, second-, and third-person singular and plural pronouns occurring in each dyad member's conversation. These values were then divided by the total number of personal pronouns used by the dyad member to create percentage scores that previously have been shown to be free of any confound with individual differences in the amount of talking (Ickes, Reidhead, & Patterson, 1986). These percentage score measures provide a common metric for assessing the degree to which the content of the dyad members' conversations reflect a focus on self (first-person pronouns), on one's current interaction partner (second-person pronouns), or on some other person(s) not involved in the

current interaction (third-person pronouns). The interrater agreement rates for these measures ranged from .92 to .97.

Self-Report Measures

Four types of self-report data were obtained that are relevant to this investigation: (a) the subjects' responses on personality measures of shyness (Cheek & Buss, 1981), self-monitoring (Snyder, 1974), empathic ability (Davis, 1983) and empathic accuracy (Ickes, 1988), (b) the dyad members' own reported thoughts and feelings, (c) their inferences about their partner's reported thoughts and feelings, and (d) their responses to the previously described items on the posttest questionnaire.

Thought/feeling measures. The subjects' own thought/feeling data were later coded by two independent judges who determined both the *perspective* and the *target* of each thought/feeling entry. The judges first decided whether the entry was reported as a direct perspective (subject's own perspective) or as a meta-perspective (subject's representation of his or her partner's perspective; cf. Laing, Phillipson, & Lee, 1966). They then decided on the target or object of the perception expressed by the entry (i.e., whether the thought or feeling primarily concerned self, partner, other person(s), or an environmental object, event, or circumstance). The rates of interrater agreement for these categorical classifications were .99 for perspective and .87 for target, values quite consistent with those previously reported by Ickes et al., 1986.

The coded thought/feeling data for the subjects within each dyad were subsequently entered into a microcomputer as the input to a software program called COLLECT YOUR THOUGHTS. This program produces as output various summary indexes of thought/feeling content. In the present study, these measures included the percentages of each subject's reported thought/feeling entries that were coded as thoughts, feelings, positive entries, negative entries, direct perspective entries, and meta-perspective entries. They also included the percentages of self entries, partner entries, other(s) entries, and environment entries, as determined by the judges' ratings of the primary target of each thought or feeling.

A few more fine-grained measures that were of special theoretical interest (e.g., the percentages of negative self entries and positive partner entries) were also computed. Among these was a measure of the percentage of partner attribution entries. These entries represented the dyad members' attributions about the relatively stable and enduring characteristics of their interaction partners (e.g., "this guy is pretty creative," "she is snobbish," "what a weirdo!").

Empathic accuracy measures. Two measures of empathic accuracy were computed from comparisons of the actual and inferred thought/feeling entries. The first measure, valence accuracy, was an index of the percentage of instances in which a dyad member's inference about the overall emotional tone of his or her partner's entry (positive, neutral, or negative) matched the actual valence label assigned to the entry by the partner. The second measure, content accuracy, was an index of the degree to which a dyad member's written description of the inferred content of his or her partner's entry matched the actual content of the sentence that the partner had written to describe the entry.

The valence accuracy measure was computed with the aid of a specialized utility program in the COLLECT YOUR THOUGHTS software package. This program accessed the relevant raw data from a master data file and computed the number of matches and mismatches between the rated valence of the actual entries and the corresponding inferred entries for the members of each dyad. It then computed the percentage of correct matches for each dyad member.

Because the content accuracy measure required subjective judgments to be made about the degree of similarity in the written content of the actual and inferred thought/feeling entries, a different proce-

ture was used to derive this measure. First, a transcript of all the thought/feeling entries reported in the study, both actual and inferred, was created, so that each dyad member's actual thought/feeling entries appeared on the left side of the page with the corresponding inferred entries appearing directly beside them on the right. Then, separate copies of the transcript were given to six independent judges. Their task was to compare the written content of each actual entry with that of the corresponding inferred entry and rate the degree of similarity (i.e., content accuracy), using a 3-point scale ranging from *essentially different content* (0) through *somewhat similar, but not the same, content* (1) to *essentially the same content* (2).

In the present study, the internal consistency of the six judges' content accuracy ratings was .94. The high reliability of these ratings indicated that the six raters were using essentially the same judgment criteria, and it justified the decision to use the mean of the ratings as our empirical measure of content accuracy.

Physical Attractiveness Measure

The physical attractiveness of each subject was assessed by 11 raters who viewed a clip of the first 20 s of each dyadic interaction with the sound turned off. Each dyad member's attractiveness was rated on a 10-point scale ranging from *extremely unattractive* (1) to *extremely attractive* (10). The internal consistency of these ratings was .93, a value that justified using the mean of the ratings as the measure of physical attractiveness.

Estimation of Baseline Accuracy Levels

The logic underlying component analyses of accuracy in person perception—see, for example, Cronbach (1955, 1958), Gage and Cronbach (1955), Funder (1980, 1987), Harackiewicz and DePaulo (1982), Kenny (1986, 1988), Kenny and Albright (1987), and Schneider, Hastorf, and Ellsworth (1979)—was applied, with some necessary modifications, to the present study of empathic accuracy. This logic suggested that our subjects could accurately infer the content of a given thought/feeling entry for any of three reasons. First, they could correctly infer the specific content of that specific entry (differential accuracy). Second, they could correctly infer the general theme or tenor of their partner's thought/feeling entries and, by basing their educated guess on that general stereotype, make a correct inference about a given entry with a probability greater than chance (stereotype accuracy). Third, they could simply make a lucky, uneducated guess about the content of the entry and be correct on a few rare occasions because of chance alone (chance accuracy).

Of these three accuracy components, the first two reflect some type of inferential ability, whereas the third does not. Although in some studies an estimate of both the second and third components might be required, for our present purposes it was necessary to estimate and correct for only the third component in order to distinguish accuracy that was based on some form of inferential ability from a false accuracy component that was due simply to chance.³

D. A. Kenny (personal communication, 1988) proposed that the simplest and most straightforward way to estimate the baseline accuracy levels that reflect the false accuracy component was to randomly pair each set of actual thought/feeling entries with the corresponding set of partner inferences and then determine the levels of content and valence accuracy obtained with these random pairings. These estimated baseline accuracy scores could then be subtracted from the original accuracy scores to create refined accuracy scores from which the effects of false accuracy have been removed. The logic here is analogous to the logic of making a correction for guessing in the scores obtained on an objective multiple-choice test in which subjects vary in the number of items they complete.

Obtaining estimates of the baseline accuracy levels for our measure of valence accuracy was relatively easy. Because the valence accuracy measure was already computed with the aid of a specialized utility program in the COLLECT YOUR THOUGHTS software package, Victor Bissonnette modified the software so that the lab computer could generate valence accuracy scores after randomly pairing each set of actual valence ratings with the corresponding set of partner inferences. Ten of these random pairings were run for each of the 76 cases in which a subject made inferences about the set of valence accuracy ratings provided by his or her partner. The valence accuracy scores across the 10 pairings were then averaged to provide an estimate of each subject's baseline valence accuracy score.⁴

Obtaining baseline accuracy estimates for our measure of content accuracy was considerably more difficult, because it again required a group of raters to make subjective judgments about the degree of similarity in the written content of the actual and inferred thought/feeling entries. The time demands of this task were such that we could afford to have 11 raters (8 experienced raters and 3 novices) rate only one random pairing of the 76 sets of actual and inferred content entries. Fortunately, however, the internal consistency of the content accuracy judgments provided by the 11 raters was very high ($\alpha = .91$).

Results and Discussion

Validity Tests of the Thought/Feeling Assessment

As checks on the validity of the thought/feeling assessment, two tests of concurrent validity were conducted to replicate those reported by Ickes et al. (1986). The results of these tests are reported in Table 1 in the form of dyad-level and adjusted individual-level correlations. Because the dyad members' responses were significantly correlated (i.e., nonindependent) for the posttest measures of liking and discomfort, it is appropriate to report only the dyad-level correlations and the adjusted individual-level correlations for which the dyad-level covariation has been statistically controlled (cf. Kenny, 1988; Kenny & La Voie, 1985).

Consistent with the findings reported by Ickes et al. (1986), the percentage of negative self entries (percentage of negative thoughts and feelings about self) was positively correlated at the dyad level with the posttest measure of discomfort during the

³ If this point is not clear, consider whether a student should be given credit for spelling the word *perceive* correctly (a) because he or she evidences the ability to spell that specific word (differential accuracy) or (b) because he or she evidences the ability to apply the appropriate stereotypic rule, "i before e except after c" (stereotype accuracy). Pragmatic teachers would probably credit a correct response in either case and might deny credit only if they had reason to believe that the student had simply made a lucky guess.

⁴ When the number of actual thoughts and feelings reported by a subject was small (e.g., as typified by cases in which only three or four entries were reported), the estimated levels of baseline valence accuracy were likely to be slightly inflated because of the greater probability that the random pairing of the actual valence ratings and the inferred valence ratings would result in a larger percentage of the original pairings being preserved. To correct for this potential bias, the computer program allowed no original pairings to be preserved in the matching procedure.

interaction.⁵ In addition, the percentage of positive partner entries (percentage of positive thoughts and feelings about partner) was positively correlated at the dyad level with the posttest rating of liking for partner. The corresponding adjusted individual-level correlations were not significant, indicating that the significant dyad-level relationships reported in Table 1 were emergent consequences of the dyad members' interactions that cannot be accounted for by processes occurring only at the individual level of analysis.

Comparisons of Original and Revised Accuracy Scores

Table 2 contains data that provide comparisons of (a) the original content and valence accuracy scores, (b) the baseline levels for each of these measures, and (c) the revised scores for each measure (original score minus baseline). Because there were no significant gender differences in the means or the patterns of intercorrelation of these measures, only the data for the combined sample are reported.

Content accuracy. As the data in the top portion of Table 2 indicate, the means for the original, the baseline, and the revised measures of content accuracy were all significantly greater than zero. These data indicate that our original content accuracy scores contained a small baseline component in addition to a much larger component that was presumably a refined measure of differential content accuracy. Each of these components was highly reliable (all $ps < .0001$) when tested against the null hypothesis.

Further analysis revealed that the original and the revised content accuracy scores were correlated .91 for the women in the sample, .92 for the men, and .91 for all subjects combined. Thus, although the baseline component of content accuracy was statistically reliable for both the men and the women in the sample, correcting for its influence had relatively little impact on the rank ordering of the subjects' content accuracy scores.

Valence accuracy. As the data in the bottom portion of Table 2 indicate, the means for the original and baseline measures of valence accuracy were significantly greater than zero ($ps < .0001$), but the mean for the revised (i.e., difference-score) measure was not. In sharp contrast to the content accuracy data, these valence accuracy data revealed that our original valence accuracy scores contained a very large baseline component in addition to a second, much smaller component. The fact that this second component was not significant when tested against

Table 2

Components of the Content and Valence Accuracy Measures

Empathic accuracy measure	<i>M</i>	<i>SD</i>	<i>F</i> (1, 75)	<i>p</i>
Content accuracy	21.7	12.1	242.4	<.0001
Baseline content accuracy	5.8	5.2	94.7	<.0001
Revised content accuracy	15.9	12.6	121.0	<.0001
Valence accuracy	40.1	17.1	419.0	<.0001
Baseline valence accuracy	36.8	12.6	646.7	<.0001
Revised valence accuracy	3.3	21.8	1.8	<.1885

Note. The *F* ratios test the significance of the mean for each component against the null hypothesis. The method of computation makes each component analogous to a percentage of a perfect accuracy score of 100. Because the scores of the male dyad members were not significantly correlated with those of their female partner for any of these measures, the data were collapsed across gender and tested using the individual as the unit of analysis.

the null hypothesis suggests that it was not a reliable component of empathic accuracy but instead represented chance deviations from the valence accuracy baseline.

Given this finding, further discussion of the valence accuracy measure is not warranted, because the subjects' scores on this measure provided no indication of differential valence accuracy. Accordingly, the results reported next concern only the content accuracy measure, for which a highly reliable level of differential accuracy was found.

Correlates of Content Accuracy

The revised content accuracy scores of the male dyad members were not significantly correlated with those of their female partners ($r = -.13$, *ns*). This finding indicates that the level of content accuracy achieved by the male dyad members was essentially independent of that achieved by their female partners. On the other hand, the analogous correlations for the remaining variables revealed that many, though not all, of the predictor variables with which content accuracy proved to be correlated were ones for which the responses of the male and female dyad members were significantly correlated (i.e., were nonindependent). For this reason, it was necessary to apply Kenny and La Voie's (1985) guidelines for determining whether a given correlation exists and should be reported at the individual level, at the dyad level, or at both levels of analysis.⁶

Table 1
Thought/Feeling Correlates of Self-Reported Discomfort and Liking of Partner

Thought/feeling measure	Self-reported discomfort		Liking for partner	
	Dyad	ADJ IND	Dyad	ADJ IND
% negative self entries	.49**	.15	-.15	-.14
% positive partner entries	-.43*	-.07	.51**	.16

Note. Dyad = dyad-level correlations ($df = 36$); ADJ IND = adjusted individual-level correlations ($df = 37$).

* $p < .01$. ** $p < .005$.

⁵ When the substantial correlation between the percentage of positive partner entries and self-reported discomfort was statistically controlled, the estimated partial correlation between the percentage of negative self entries and self-reported discomfort remained highly significant, $r_p(35) = .44$, $p < .005$.

⁶ Although it would be logical and desirable at this point to present a thorough discussion of the levels-of-analysis issue, space limitations preclude such discussion. Instead see Kenny and La Voie (1985). Suffice it to say here that dyad-level correlations are based on the mean (i.e., average) of the two dyad members' scores on each variable, whereas the adjusted individual-level correlations are based on the deviations of the two dyad members' scores on each variable from its respective dyad-level mean.

Table 3
*Correlates of Revised Content Accuracy in
 Initial Mixed-Sex Interactions*

Measures	UNADJ	Dyad	ADJ IND
Verbal behavior			
Verbalizations (frequency)	.17	.34**	-.16
% questions asked	.19	.03	.29*
% first-person singular pronouns	.14	.35**	-.11
% third-person pronouns	-.21	-.35**	-.04
Nonverbal behavior			
Directed gazes (frequency)	.23**	.38**	.03
Positive affect (frequency)	.35****	.29	.45****
Thought/feeling			
% partner attributions	.39*****	.47****	.31*
% partner entries	.24**	.24	.24
% positive partner entries	.35****	.19	.53*****
Individual difference			
Grade point average	.28**	.41***	.15
Self-monitoring	.24**	.12	.33**
Partner's attractiveness	.24**	.31*	.16

Note. UNADJ = unadjusted individual-level correlations ($df = 74$); Dyad = dyad-level correlations ($df = 36$); ADJ IND = adjusted individual-level correlations ($df = 37$). The correlations that can be meaningfully interpreted in this table, according to Kenny & La Voie's (1985) criteria, have been placed in boxes and set in boldface type.

* $p < .07$. ** $p < .05$. *** $p < .01$. **** $p < .005$. ***** $p < .001$.

By these guidelines, for cases in which the responses of the male dyad members were statistically independent of those of their female partners for both the criterion variable (content accuracy) and the predictor variable, the significant correlates of content accuracy characterized relationships that occurred and should be interpreted *only at the (unadjusted) individual level of analysis*. In contrast, for cases in which the dyad members' responses were independent for the criterion variable but were interdependent (i.e., correlated) for the predictor variable, the significant correlates of content accuracy characterized relationships that occurred and should be interpreted *only at the dyad and adjusted individual levels of analysis* (D. A. Kenny, personal communication, 1988). We applied these guidelines to the data reported in Table 3 by using boldface type and surrounding boxes to indicate at what level(s) of analysis the correlations should be reported and interpreted.

Partner attributions and content accuracy. An initial and very important question we wished to ask of the data concerned the relationship between attributional inference and empathic inference. Specifically, did the degree to which dyad members made attributions about each other's more stable and enduring characteristics correlate with the degree to which they were able to accurately infer each other's more transient thoughts and feelings? An affirmative answer to this question would suggest that separate theories to account for attribution making and empathic accuracy may not be required, because both phenomena may be outcomes or products of a more general epistemic attempt to "understand" another person (cf. Kruglanski, 1989).

To answer this question, the revised content accuracy measure was correlated with the measure of the percentage of the

dyad members' thought/feeling entries that concerned relatively enduring dispositions of their interaction partners. The results, reported under the heading of Thought/feeling measures in Table 3, indicated that the percentage of partner attribution entries was significantly correlated with revised content accuracy at both the dyad ($r = .47$) and the adjusted individual ($r = .31$) levels of analysis. In other words, even when the significant dyad-level effect was statistically controlled, those individual dyad members who reported the largest percentages of partner attributions were also likely to make the most accurate inferences about the content of their partner's thoughts and feelings.

Verbal behavior and content accuracy. The verbal behavior that occurred in these initial, mixed-sex interactions had a pronounced effect on the dyad members' content accuracy, as indicated in the first four rows of Table 3. Of the verbal behavior measures reported, content accuracy was found to be positively correlated at the dyad level with a global measure of the total number of verbalizations (i.e., speaking turns) as well as with the more specific measure of the percentage of first-person singular pronouns (I, me, my, mine, myself). On the other hand, content accuracy was negatively correlated at the dyad level with the percentage of third-person pronouns (he, she, they, them, etc.). Because the percentages of first-person singular and third-person pronouns were themselves negatively correlated at the dyad level, $r(36) = -.71$, $p < .001$, these effects should not be viewed as independent of each other. Rather, it was the *relative* use of first-person singular versus third-person pronouns that contributed to content accuracy.

It makes sense that these relationships with verbal behavior would be found at the dyad level of analysis, for that was obviously the level at which the conversation occurred. Note also that these conversational influences on empathic accuracy lend themselves well to an informational interpretation. On the one hand, the more information that dyad members conveyed to each other through self-disclosing, self-referenced conversation, the better able they were to accurately infer the specific content of each other's thoughts and feelings. On the other hand, the more that these strangers talked to each other about third parties, many of whom were known by only one of the dyad members, the less able they were to accurately infer the specific content of each other's thoughts and feelings.

The remaining verbal behavior measure—the percentage of verbalizations that were coded as questions—was positively correlated with revised content accuracy, but only at the adjusted individual level of analysis. Because the data at this level represented the dyad member's scores as deviations from the dyad's mean on each variable, the obtained positive correlation of .29 indicates that the higher the percentage of questions asked by one dyad member relative to his or her partner, the greater was that same dyad member's content accuracy as compared with his or her partner's. This finding suggests that relative differences in the amount of information that dyad members elicit through questioning each other contribute to relative differences in content accuracy within the dyads.

Nonverbal behavior and content accuracy. The correlation of revised content accuracy with the measure of how frequently the dyad members looked at each other can also be interpreted

in informational terms. Frequent gazing at one's interaction partner should facilitate content accuracy, because the visual monitoring of relatively rapid and subtle changes in body posture, gestures, facial expression, and so on enables the perceiver to further qualify the meaning of the other's words as well as to acquire additional information about the other's emotional state.

How should we explain the adjusted individual-level correlation of revised content accuracy with the measure of how frequently the dyad members smiled at each other (i.e., the positive affect measure)? Because the data at this level represent the dyad member's scores as deviations from the dyad's mean on each variable, the obtained positive correlation of .45 indicates that the more often one dyad member smiled relative to his or her partner, the greater was that same dyad member's content accuracy relative to his or her partner's.

We explored three possible mediational interpretations. First, we tested and ruled out the possibility that this relationship was an indirect and socially mediated effect of the perceiver's smiling on his or her partner's motivation to be more open and self-disclosing. We found that partialing out the effect of the partner's talking did not eliminate the correlation between the perceiver's smiling and his or her content accuracy ($r_p = .44, p < .01$). Second, we tested and ruled out the possibility that either the partner's physical attractiveness or the perceiver's own rated liking for the partner mediated the relationship between smiling and content accuracy. With these variables partialled out, the residual correlations were still significant, being .43 and .45, respectively ($ps < .01$).

Third, we tested the possibility that the dyad members who smiled more frequently were more visually attentive or more verbally responsive to their partners than those who smiled less. The data revealed that the measure of smiling was indeed positively correlated at the adjusted individual level with the measure of gazing, $r(36) = .32, p < .05$. However, the correlation at this level between smiling and content accuracy was not substantially attenuated when the effects of gazing, talking, or use of verbal reinforcers were partialled out, $r_p s > .43, ps < .01$.

In summary, the data revealed no obvious mediation for the adjusted individual-level relationship between smiling and content accuracy. Our best guess is that smiling in mixed-sex dyads is an expression of an underlying interest in the partner—an interest that also leads, by a route more subtle than the sheer amount of looking, talking, and so on, to a higher level of content accuracy. This interpretation, though speculative, is supported by the pattern of measures that significantly correlate with revised content accuracy at the adjusted individual level. All three measures—percentage of questions, frequency of smiling, and percentage of partner attributions—appear to be manifestations of an underlying interest in the partner. Note that this interpretation has a motivational/attentional flavor that distinguishes it from the more strictly informational interpretation of the verbal behavior data.

Self-report measures and content accuracy. Consistent with this speculation are the two remaining correlations reported in Table 3 between content accuracy and measures of thought/feeling content. These (unadjusted) individual-level correlations also suggest that the individual dyad members who mani-

Table 4
Nonsignificant Personality Correlates of Revised Content Accuracy

Personality measures	Revised content accuracy		
	Men	Women	Combined
Perspective taking (PT)	-.15	-.15	-.14
Empathic concern (EC)	.09	-.09	.04
Fantasy identification (F)	-.13	-.16	-.11
Personal distress (PD)	.16	-.12	.06
Empathic accuracy (EA)	.10	-.21	-.03

Note. Because the scores of the male dyad members were not significantly correlated with those of their female partners for any of these individual difference measures, all of the correlations in this table are reported at the individual level of analysis ($df = 36$ for the men and women separately and $df = 74$ for the combined sample).

festated the greatest interest in and positive sentiment for their interaction partners (i.e., those reporting the largest percentages of positive, partner-relevant thoughts and feelings) were the most accurate in inferring the content of their partner's thoughts and feelings.

The significant individual difference correlates of content accuracy are reported in the last three rows of Table 3. These effects, all at the (unadjusted) individual level, lend themselves to fairly straightforward interpretations. First, the combination of intelligence and motivation that is reflected in high GPA scores may also reflect the individual dyad members' more specific ability and motivation to accurately "read" their interaction partners.

Second, the well-documented "other orientation" (e.g., Gabrenya & Arkin, 1980; Briggs & Cheek, 1986; Briggs, Cheek, & Buss, 1980; Ickes, Reidhead, & Patterson, 1986) and social perceptiveness (e.g., Geizer, Rarick, & Soldow, 1977; Mill, 1984) that are reflected in high self-monitoring scores may constitute a second, *independent* source of motivation and ability to achieve relatively high levels of content accuracy. Analyses of the self-monitoring factor scores (Briggs & Cheek, 1986) provided direct support for this speculation: Content accuracy was found to be correlated with the Other-Directedness factor, $r(74) = .31, p < .01$, but not with the Acting or Extraversion factors, $rs = .09$ and $.00, ns$. GPA and self-monitoring scores were uncorrelated in this sample, $r(74) = .03, ns$.

Third, the physical attractiveness of their opposite-sex interaction partner is a highly salient external factor that also could plausibly increase the motivation of individual dyad members to accurately read their partner's thoughts and feelings.

Self-reported empathic dispositions and content accuracy. The self-report personality measures of empathic skills and empathic accuracy proved to be disappointingly poor predictors of revised content accuracy. The relevant correlations, reported in Table 4, indicated that none of the subscales of Davis's (1983) empathy measure nor the empathic accuracy measure developed by Ickes (1988) was a significant predictor of content accuracy for either the male or the female dyad members in this study. Worse yet, most of the correlations reported in Table 4

were negative, rather than positive, in sign. Apparently, subjects' perceptions of their own empathic abilities do not predict their actual empathic accuracy, at least not in the context of an initial interaction with a stranger of the opposite sex.

Gender and content accuracy. Finally, note that although gender is stereotypically regarded as an individual difference variable that is related to empathic accuracy, the women's revised content accuracy scores were not significantly greater than those of their male partners, $M_s = 17.3$ and 14.5 , respectively, $t(37) < 1$, *ns*. This type of finding is by no means unique. As a recent review by Eisenberg and Lennon (1983) suggests, significant gender differences are commonly found for fairly obvious measures of *self-reported* empathy but are rarely found for *behavioral* measures of empathy.

The most general interpretation of this null finding would be that the widely assumed relationship between gender and empathic accuracy is a cultural myth. It is simply not warranted, however, to draw so general a conclusion at this point. First, a difference between women and men in empathic accuracy would probably be most evident in long-term, established relationships, not in initial interactions. Second, because the target's gender is perfectly confounded with the perceiver's gender in this study, it is possible that women may be more accurate perceivers than men but that men may be harder to read. Third, although Riggio, Tucker, and Coffaro (1989) also found no difference between women and men in their performance on a cognitive empathy task similar to ours (see also Archer & Akert, 1977), they did replicate the common finding that women are better decoders of facial expressions of emotion (Friedman, Prince, Riggio, & DiMatteo, 1980; Hall, 1979; Rosenthal, Hall, DiMatteo, Rogers, & Archer, 1979). This common finding suggests that a reliable measure of *valence* accuracy (which ours was not) might have revealed a gender difference favoring women. For all of these reasons, the generalizability of the null sex difference obtained in the present study must remain open to question.

Summary and Implications

There were two goals of this research. The first was to establish the utility of our method for assessing empathic accuracy in the context of initial, mixed-sex interactions; the second was to use the method to explore some of the factors that determine the level of empathic accuracy in this same interactional context.

Assessment of Empathic Accuracy

The first goal was realized with respect to our original measure of content accuracy, which proved to be highly reliable ($\alpha = .94$). Moreover, the revised measure of content accuracy was so highly correlated with the original measure (.91) that our correction for guessing appeared to have had only a negligible effect on the correlations of the content accuracy scores with other variables.

On the other hand, the same correction for guessing had serious implications for the valence accuracy measure. Because this measure failed to indicate a degree of differential accuracy that significantly exceeded baseline levels, we are doubtful that

it will have much utility in future studies of the initial, short-term interactions of strangers. However, it might still prove to be reliable in longer term interactions or in relationships in which stronger and more volatile emotions are expressed.

Correlates of the (Revised) Content Accuracy Measure

The second goal of this study was also realized, in that the data patterns offered a coherent yet highly detailed view of content accuracy's relations with a range of other variables. These data patterns, which should not be generalized beyond the initial, unstructured interactions of mixed-sex dyads, were as follows:

First, the degree to which the dyad members made attributions about each other's more stable and enduring characteristics was positively correlated with the degree to which they were able to accurately infer the content of each other's more transient thoughts and feelings. This novel, though not counterintuitive, finding establishes a clear-cut link between the processes underlying attributional inference and the processes underlying content accuracy as a form of empathic inference. Such a link may be important in developing a theoretical model that is adequate to account not only for those social inferences that are relatively global and enduring (i.e., attributed traits and other stable dispositions) but also for those that are relatively specific and fleeting (i.e., attributed thoughts and feelings; cf. Kruglanski, 1989).

Second, a pattern of dyad-level correlations revealed that the more information the dyad members elicited from each other (i.e., through directed gazes and by asking questions) and conveyed to each other (i.e., through self-referenced conversation), the better able they were to accurately infer the specific content of each other's thoughts and feelings. On the other hand, the more these strangers talked to each other about third parties, many of whom were known by only one of the dyad members, the less able they were to accurately infer the specific content of each other's thoughts and feelings.

These findings underscore what is clearly a major lesson of the data in Table 3: Content accuracy in initial, mixed-sex interactions is, to a large extent, the emergent product of social interaction processes occurring at the dyad level of analysis. But what, exactly, does this mean? Does it mean that two interacting dyad members should be more accurate in inferring each other's thoughts and feelings than should a third person who is present only as a passive (or even a hidden) observer of their interaction? Not at all. It simply means that the information that any of the parties (interacting dyad members or passive observers) derive from the dyad members' behavior and use as the basis of their inferences about the content of the dyad members' thoughts and feelings *emerges during the interaction as a socially constructed meaning context that exists only at the dyad level*. Because this dyad-level meaning context is the joint, interdependent construction of the two dyad members and is available not only to them but to nonparticipating observers as well, the empathic accuracy of all parties present is, in large measure, a product of this emergent, dyad-level influence.

Third, a pattern of individual-level correlations indicated the importance to their content accuracy of the perceivers' degree of interest in their partner. The greater the dyad member's inter-

est in his or her partner (as defined by the frequency of smiling and the percentage of positive, partner-relevant thoughts and feelings), the higher was his or her level of content accuracy. This pattern of correlations, like those for gazing and smiling at the partner, suggests that it is the subject's interest in the partner *during the interaction* that is crucial in mediating these effects. Whereas all subjects were later encouraged by the empathy task instructions to accurately infer their partner's thoughts and feelings, those subjects who were most attentive to and interested in their partners *during the interaction* still obtained higher accuracy scores than those who were not. Apparently, the impairment in empathic accuracy resulting from a lack of interest in and attentiveness to one's partner during an interaction cannot be easily compensated for by a later opportunity to view the interaction on videotape. If so, these results may have important, and somewhat counterintuitive, implications regarding the use of videotape review procedures for training empathic skills in clinical and counseling psychology programs.

Fourth, another pattern of individual-level correlations indicated that content accuracy was related to various individual differences variables in ways that sometimes did but sometimes did not support commonsense intuition. The specific ability and motivational components that are reflected in GPA and in self-monitoring scores appeared to contribute independently to the dyad members' ability and motivation to read each other accurately. In addition, the level of their partners' physical attractiveness appeared to increase the motivation of the individual dyad members to accurately read their partner's thoughts and feelings.

On the other hand, the subjects' gender and measures of their self-reported empathic skills proved to be disappointingly poor (i.e., nonsignificant) predictors of content accuracy. The generality of these null findings must remain open to question, particularly given Bernstein and Davis's (1982) finding that the relation between perspective-taking scores and accuracy was moderated by the length of exposure to the target individual. However, the fact that we found significant effects for other individual difference predictors (e.g., GPA and self-monitoring) suggests that the counterintuitive null findings for gender and self-rated empathic skills should not readily be dismissed.

Implications for Theory and Research

These findings have a number of important implications for the study of empathic accuracy and intersubjective phenomena in general. For one thing, they clearly portray empathic accuracy as the outcome of processes that are psychological (i.e., inferential) in nature. In the same vein, these findings suggest a link between the processes that determine both the accuracy of one's inferences about the content of an interaction partner's specific thoughts and feelings and the extent to which one makes more global, dispositional attributions about the partner. And, most reassuring perhaps for the more rigorous adherents of information processing approaches, our data indicate that considerable weight should be assigned to informational factors as determinants of empathic accuracy in naturally occurring social interaction.

On the other hand, the present findings also point to the

complexity of empathic accuracy and the need to address this complexity in both theoretical and methodological terms. This complexity is evident in at least three respects. First, there is evidence that although informational determinants of empathic accuracy are clearly important, they by no means tell the whole story. Content accuracy was also positively correlated with individual-level motivational factors such as the partner's physical attractiveness and the perceiver's degree of interest in the partner during the interaction itself. We expect that these extra-informational influences on empathic accuracy will prove to be among the most theoretically provocative, with important clinical/applied implications as well (cf. Barrett-Lennard, 1962; Gladstein, 1977; Goldstein & Michaels, 1985; Kagan, 1972; Truax & Carkuff, 1967).

Second, there is evidence that empathic accuracy is affected by individual difference variables in ways that do not always accord with conventional wisdom or cultural stereotype. Conventional wisdom would probably have little difficulty in accepting our findings that, in initial mixed-sex interactions, high GPA scores were associated with greater empathic accuracy. It would have somewhat more difficulty, however, in accepting our findings that gender and self-report measures of empathic skills and empathic accuracy were disappointingly poor predictors of content accuracy. The lack of correspondence between people's perceived and actual levels of ability to accurately infer the thoughts and feelings of others was quite striking in this study, and we regard this flawed estimation of one's own empathic accuracy as one of the most fascinating and compelling foci for future theory and research. It, too, should be a topic of major clinical/applied interest.

Third, there is evidence that empathic accuracy is, to a large extent, the emergent product of interaction processes occurring at the dyad (or group) level of analysis. From the standpoint of a variety of theoretical perspectives ranging from symbolic interactionism (Stryker & Statham, 1985) to communications theory (Fisher, 1978; Krippendorff, 1984; Penman, 1980; Watzlawick, Beavin, & Jackson, 1967) to social systems theory (Bateson, 1972; Buckley, 1967), it is certainly not surprising that such evidence should be found. On the other hand, the strength and consistency of such emergent, dyad-level influences in our data clearly underscore the intersubjective nature of empathic accuracy and related phenomena. By implication, they also underscore the limitations of more traditional research paradigms that fail to address many of the theoretical and methodological issues that arise when social cognition is studied in the context of naturally occurring social interaction.

It was, of course, our awareness of these limitations that led us to conduct this most recent of our studies of naturalistic social cognition. In this study, one of our goals was to demonstrate that it is possible to operationally define and measure empathic accuracy in a meaningful, objective, and reliable way. The findings we obtained lead us to conclude that our measure of content accuracy is not just a distant cousin of the corresponding empathic understanding that naturally emerges in ongoing social interaction. It is, instead, a reasonably faithful representation of the degree to which each dyad member can accurately infer the specific content of his or her partner's thoughts and feelings. We believe that this method for assessing content accuracy in the context of ongoing social interaction is

a viable one that should be applicable to work in other social science disciplines as well as in other areas of psychology.

As a final note, we acknowledge that our own research has some major limitations as well. Some of these limitations concern the methodology we have used. Despite its many advantages, the method is so expensive, labor intensive, and time consuming that we would be reluctant to use it in the absence of the custom software we developed. Moreover, although the method is well suited to the study of short-term, dyadic interactions, it cannot be easily applied to study longer-term interactions; nor does it adapt easily to the study of interaction in which subjects participate in more than one dyad or in larger groups (cf. Ickes, 1982, 1983; Ickes et al., 1990).

Other limitations concern the preliminary, exploratory nature of this research. We have obviously not yet taken full advantage of the method's potential in allowing researchers to relate content accuracy to a wide range of behavioral predictors. In particular, our measures of personal pronoun usage were only a token acknowledgment of the importance of conversational structure and content in determining levels of empathic accuracy. Although empathic accuracy is an inferential phenomenon that is deeply rooted in cognitive social psychology, we must remember that it implicates communications, discourse analysis, kinesics, linguistics, paralinguistics, proxemics, systems theory, and many other approaches as well. Having scratched the surface of this phenomenon, there is still much we have to do.

References

- Archer, D., & Akert, R. M. (1977). Words and everything else: Verbal and nonverbal cues in social interpretation. *Journal of Personality and Social Psychology, 35*, 443-449.
- Asch, S. E. (1952). *Social psychology*. Englewood Cliffs, NJ: Prentice-Hall.
- Barrett-Lennard, G. T. (1962). Dimensions of therapist response as causal factors in therapeutic change. *Psychological Monographs, 76*, 1-33.
- Bateson, G. (1972). *Steps to an ecology of mind*. New York: Ballantine Books.
- Bernstein, W. M., & Davis, M. H. (1982). Perspective-taking, self-consciousness, and accuracy in person perception. *Basic and Applied Social Psychology, 3*, 1-19.
- Briggs, S. R., & Cheek, J. M. (1986). The role of factor analysis in the development and evaluation of personality scales. *Journal of Personality, 54*, 106-148.
- Briggs, S. R., Cheek, J. M., & Buss, A. H. (1980). An analysis of the Self-Monitoring Scale. *Journal of Personality and Social Psychology, 38*, 679-686.
- Brock, T. C. (1967). Communication discrepancy and intent to persuade as determinants of counterargument production. *Journal of Experimental Social Psychology, 3*, 269-309.
- Buckley, W. (1967). *Sociology and systems theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Cacioppo, J. T., Glass, C. R., & Merluzzi, T. V. (1979). Self-statements and self-evaluations: A cognitive response analysis of heterosexual anxiety. *Cognitive Therapy and Research, 3*, 249-262.
- Cacioppo, J. T., & Petty, R. E. (1981). Social psychological procedures for cognitive response assessment: The thought-listing technique. In T. Merluzzi, C. Glass, & M. Genest (Eds.), *Cognitive assessment* (pp. 309-342). New York: Guilford Press.
- Campbell, R. J., Kagan, N., & Krathwohl, D. R. (1971). The development and validation of a scale to measure affective sensitivity (empathy). *Journal of Counseling Psychology, 18*, 407-412.
- Carkhuff, R. R. (1969). *Helping and human relations: A primer for lay and professional helpers* (Vols. 1-2). New York: Holt, Rinehart & Winston.
- Chapin, F. S. (1942). Preliminary standardization of a social insight scale. *American Sociological Review, 7*, 214-225.
- Cheek, J. M., & Buss, A. H. (1981). Shyness and sociability. *Journal of Personality and Social Psychology, 41*, 330-339.
- Cline, V. B. (1964). Interpersonal perception. In B. Maher (Ed.), *Progress in experimental personality research* (Vol. 1, pp. 221-284). San Diego, CA: Academic Press.
- Cronbach, L. J. (1955). Processes affecting scores on "understanding of" others and "assumed similarity." *Psychological Bulletin, 52*, 177-193.
- Cronbach, L. J. (1958). Proposals leading to analytic treatment of social perception scores. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior* (pp. 353-379). Stanford, CA: Stanford University Press.
- Crow, W. J., & Hammond, K. R. (1957). The generality of accuracy and response set in interpersonal perception. *Journal of Abnormal and Social Psychology, 54*, 384-390.
- Danish, S. J., & Kagan, N. (1971). Measurement of affective sensitivity: Toward a valid measure of interpersonal perception. *Journal of Counseling Psychology, 18*, 51-54.
- Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44*, 113-126.
- Diehl, M., & Stroebe, W. (1987). Productivity loss in brainstorming groups: Toward the solution of a riddle. *Journal of Personality and Social Psychology, 53*, 497-509.
- Dymond, R. F. (1949). A scale for the measurement of empathic ability. *Journal of Consulting Psychology, 13*, 127-133.
- Eisenberg, N., & Lennon, R. (1983). Sex differences in empathy and related capacities. *Psychological Bulletin, 94*, 100-131.
- Fisher, B. A. (1978). *Perspectives on human communication*. New York: Macmillan.
- Friedman, H. S., Prince, L. M., Riggio, R. E., & DiMatteo, M. R. (1980). Understanding and assessing nonverbal expressiveness: The Affective Communication Test. *Journal of Personality and Social Psychology, 39*, 333-351.
- Funder, D. C. (1980). On seeing ourselves as others see us: Self-other agreement and discrepancy in personality ratings. *Journal of Personality, 48*, 473-493.
- Funder, D. C. (1987). Errors and mistakes: Evaluating the accuracy of social judgment. *Psychological Bulletin, 101*, 75-90.
- Funder, D. C., & Harris, M. J. (1986). On the several facets of personality assessment: The case of social acuity. *Journal of Personality, 54*, 528-550.
- Gabrenya, W. K., & Arkin, R. M. (1980). Self-Monitoring Scale: Factor structure and correlates. *Personality and Social Psychology Bulletin, 6*, 13-22.
- Gage, N. L., & Cronbach, L. J. (1955). Conceptual and methodological problems in interpersonal perception. *Psychological Bulletin, 53*, 258-266.
- Geizer, R. S., Rarick, D. L., & Soldow, G. F. (1977). Deception and judgment accuracy: A study in person perception. *Personality and Social Psychology Bulletin, 3*, 446-449.
- Gladstein, G. A. (1977). Empathy and counseling outcome: An empirical and conceptual review. *The Counseling Psychologist, 6*, 70-79.
- Goldstein, A. P., & Michaels, G. Y. (1985). *Empathy: Development, training, and consequences*. Hillsdale, NJ: Erlbaum.
- Greenwald, A. G. (1968). Cognitive learning, cognitive response to persuasion, and attitude change. In A. G. Greenwald, T. C. Brock, &

- T. M. Ostrom (Eds.), *Psychological foundations of attitudes* (pp. 147–170). San Diego, CA: Academic Press.
- Greif, E. B., & Hogan, R. (1973). The theory and measurement of empathy. *Journal of Counseling Psychology, 20*, 280–284.
- Hall, J. A. (1979). Gender, gender roles, and nonverbal communication skills. In R. Rosenthal (Ed.), *Skill in nonverbal communication* (pp. 32–67). Cambridge, MA: Oelgeschlager, Gunn, & Hain.
- Harackiewicz, J. M., & DePaulo, B. M. (1982). Accuracy of person perception: A component analysis according to Cronbach. *Personality and Social Psychology Bulletin, 8*, 247–256.
- Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.
- Hogan, R. (1969). Development of an empathy scale. *Journal of Consulting and Clinical Psychology, 33*, 307–316.
- Honeycutt, J. M., Knapp, M. L., & Powers, W. G. (1983). On knowing others and predicting what they say. *The Western Journal of Speech Communication, 47*, 157–174.
- Ickes, W. (1982). A basic paradigm for the study of personality, roles, and social behavior. In W. Ickes & E. S. Knowles (Eds.), *Personality, roles, and social behavior* (pp. 305–341). New York: Springer-Verlag.
- Ickes, W. (1983). A basic paradigm for the study of unstructured dyadic interaction. In H. Reis (Ed.), *New directions for methodology of social and behavioral science* (pp. 5–21). San Francisco: Jossey-Bass.
- Ickes, W. (1988). *An 18-item measure of empathic accuracy*. Unpublished scale and associated psychometric data, University of Texas at Arlington.
- Ickes, W., Bissonnette, V., Garcia, S., & Stinson, L. (1990). Implementing and using the dyadic interaction paradigm. In C. Hendrick & M. Clark (Eds.), *Review of personality and social psychology: Vol. 11. Research methods in personality and social psychology* (pp. 16–44). Newbury Park, CA: Sage.
- Ickes, W., Reidhead, S., & Patterson, M. L. (1986). Machiavellianism and self-monitoring: As different as “me” and “you.” *Social Cognition, 4*, 58–74.
- Ickes, W., Robertson, E., Tooke, W., & Teng, G. (1986). Naturalistic social cognition: Methodology, assessment, and validation. *Journal of Personality and Social Psychology, 51*, 66–82.
- Ickes, W., & Tooke, W. (1988). The observational method: Studying the interaction of minds and bodies. In S. Duck (Ed.), *The handbook of personal relationships: Theory, research, and interventions* (pp. 79–97). Chichester, England: Wiley.
- Ickes, W., Tooke, W., Stinson, L., Baker, V. L., & Bissonnette, V. (1988). Naturalistic social cognition: Intersubjectivity in same-sex dyads. *Journal of Nonverbal Behavior, 12*, 58–84.
- Ickes, W., & Turner, M. (1983). On the social advantages of having an older, opposite-sex sibling: Birth-order influences in mixed-sex dyads. *Journal of Personality and Social Psychology, 45*, 210–222.
- Janis, I. L. (1972). *Victims of groupthink*. Boston: Houghton-Mifflin.
- Janis, I. L., & Mann, L. (1977). *Decision making: A psychological analysis of conflict, choice, and commitment*. New York: Free Press.
- Kagan, N. (1972). *Influencing human interaction*. East Lansing: Michigan State University Press.
- Kenny, D. A. (1986). Methods for measuring dyads and groups. In W. Crano & M. B. Brewer (Eds.), *Principles and methods of social research* (pp. 288–309). Boston: Allyn & Bacon.
- Kenny, D. A. (1988). The analysis of data from two-person relationships. In S. Duck (Ed.), *The handbook of personal relationships: Theory, research, and interventions* (pp. 57–77). Chichester, England: Wiley.
- 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. (1985). Separating individual and group effects. *Journal of Personality and Social Psychology, 48*, 339–348.
- Kerr, W. A., & Spieroff, B. J. (1954). Validation and evaluation of the empathy test. *The Journal of General Psychology, 50*, 269–276.
- Krauss, R. M., & Fussell, S. R. (1989). *Implementing architectural designs: On the carpentry of intersubjectivity*. Paper prepared for a festschrift to honor Ragnar Rommetveit.
- Krippendorff, K. (1984). An epistemological foundation of communication. *Journal of Communication, 34*, 21–36.
- Kruglanski, A. W. (1989). *Lay epistemics and human knowledge*. New York: Plenum Press.
- Laing, R. D., Phillipson, H., & Lee, A. R. (1966). *Interpersonal perception: A theory and method of research*. New York: Harper & Row.
- Latané, B., & Wolfe, S. (1981). The social impact of majorities and minorities. *Psychological Review, 88*, 438–453.
- Maass, A., West, S. G., & Cialdini, R. B. (1987). Minority influence and conversion. In C. Hendrick (Ed.), *Group processes: Review of personality and social psychology* (Vol. 8, pp. 55–79). Newbury Park, CA: Sage.
- Malloy, T. E., & Albright, L. (1990). Interpersonal perception in a social context. *Journal of Personality and Social Psychology, 58*, 419–428.
- Marangoni, C. (1989). *The facilitative effect of feedback training on objectively assessed empathic accuracy in a clinically relevant context*. Unpublished doctoral dissertation, University of Texas at Arlington.
- Marks, S. E., & Tolsma, R. J. (1986). Empathy research: Some methodological considerations. *Psychotherapy, 23*, 4–20.
- Markus, H., & Zajonc, R. B. (1985). The cognitive perspective in social psychology. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (3rd ed., pp. 137–230). New York: Random House.
- McCauley, C. (1989). The nature of social influence in groupthink: Compliance and internalization. *Journal of Personality and Social Psychology, 57*, 250–260.
- Mead, G. H. (1934). *Mind, self and society*. Chicago: University of Chicago Press.
- Mehrabian, A., & Epstein, N. (1972). A measure of emotional empathy. *Journal of Personality, 40*, 525–543.
- Mehrabian, A., & Reed, H. (1968). Some determinants of communication accuracy. *Psychological Bulletin, 70*, 365–384.
- Merleau-Ponty, M. (1970). *Phenomenology of perception*. (Colin Smith, Trans.). New York: Humanities Press. (Original work published 1962)
- Mill, J. (1984). High and low self-monitoring individuals: Their decoding skills and empathic expression. *Journal of Personality, 52*, 372–388.
- Miller, C. E. (1989). The social psychological effects of group decision rules. In P. B. Paulus (Ed.), *Psychology of group influence* (2nd ed., pp. 327–355). Hillsdale, NJ: Erlbaum.
- Moreland, R. L., & Levine, J. M. (1982). Group socialization in small groups: Temporal changes in individual–group relations. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 15, pp. 137–192). San Diego, CA: Academic Press.
- Moreland, R. L., & Levine, J. M. (1989). Newcomers and oldtimers in small groups. In P. B. Paulus (Ed.), *Psychology of group influence* (2nd ed., pp. 143–186). Hillsdale, NJ: Erlbaum.
- Moscovici, S., & Mugny, G. (1983). Minority influence. In P. B. Paulus (Ed.), *Basic group processes* (pp. 41–64). New York: Springer-Verlag.
- Moscovici, S., & Zavalloni, M. (1969). The group as a polarizer of attitudes. *Journal of Personality and Social Psychology, 12*, 125–135.
- Myers, D. G. (1982). Polarizing effects of social interaction. In H. Brandstatter, J. H. Davis, & G. Stocker-Kreichgauer (Eds.), *Group decision making* (pp. 125–161). San Diego, CA: Academic Press.
- Myers, D. G., & Lamm, H. (1976). The group polarization phenomenon. *Psychological Bulletin, 83*, 602–627.
- Nemeth, C. (1986). Differential contributions of majority and minority influence. *Psychological Review, 93*, 23–32.

- O'Keefe, D. J., & Sypher, H. E. (1981). Cognitive complexity measures and the relationship of cognitive complexity to communication. *Human Communication Research*, 8, 72-92.
- Pearce, W. B. (1976). The coordinated management of meaning: A rules-based theory of interpersonal communication. In G. R. Miller (Ed.), *Explorations in interpersonal communication* (pp. 17-36). Beverly Hills, CA: Sage.
- Penman, R. (1980). *Communication processes and relationships*. San Diego, CA: Academic Press.
- Petty, R. E., & Cacioppo, J. T. (1979). Effects of forewarning and persuasive intent and involvement on cognitive responses and persuasion. *Personality and Social Psychology Bulletin*, 5, 173-176.
- Riggio, R. E., Tucker, J., & Coffaro, D. (1989). Social skills and empathy. *Personality and Individual Differences*, 10, 93-99.
- Rogers, C. R. (1957). The necessary and sufficient conditions of therapeutic personality change. *Journal of Counseling Psychology*, 21, 95-103.
- Rosenthal, R., Hall, J. A., DiMatteo, M. R., Rogers, P. L., & Archer, D. (1979). *Sensitivity to nonverbal communication: The PONS Test*. Baltimore: Johns Hopkins University Press.
- Schneider, D. J., Hastorf, A. H., & Ellsworth, P. C. (1979). *Person perception* (2nd ed.). Reading, MA: Addison-Wesley.
- Snyder, M. (1974). Self-monitoring of expressive behavior. *Journal of Personality and Social Psychology*, 30, 526-537.
- Stasser, G., Kerr, N. L., & Davis, J. H. (1989). Influence processes and consensus models in decision-making groups. In P. B. Paulus (Ed.), *Psychology of group influence* (2nd ed., pp. 279-326). Hillsdale, NJ: Erlbaum.
- Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. New York: Cambridge University Press.
- Street, W. R. (1974). Brainstorming by individuals, coacting and interacting groups. *Journal of Applied Psychology*, 59, 433-436.
- Stryker, S., & Statham, A. (1985). Symbolic interaction and role theory. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (3rd ed., pp. 311-378). New York: Random House.
- Truax, C. B., & Carkuff, R. R. (1967). *Toward effective counseling and psychotherapy: Training and practice*. Chicago: Aldine.
- Walker, R. E., & Foley, J. M. (1973). Social intelligence: Its history and measurement. *Psychological Reports*, 33, 839-864.
- Watzlawick, P., Beavin, J., & Jackson, D. (1967). *Pragmatics of human communication: A study of interaction patterns, pathologies, and paradoxes*. New York: Norton.
- Wegner, D. M., Giuliano, T., & Hertel, P. T. (1985). Cognitive interdependence in close relationships. In W. Ickes (Ed.), *Compatible and incompatible relationships* (pp. 253-276). New York: Springer-Verlag.

Received June 15, 1989

Revision received May 1, 1990

Accepted May 16, 1990 ■