

Empathic Accuracy in a Clinically Relevant Setting

Carol Marangoni
Besselaar Associates, Princeton, NJ

Stella Garcia
University of Texas at San Antonio

William Ickes
University of Texas at Arlington

Gary Teng
Besselaar Associates, Princeton, NJ

This study addressed 3 questions regarding empathic accuracy in a clinically relevant setting. First, does the empathic accuracy of a perceiver improve with increased exposure to a target individual? Second, can empathic accuracy be enhanced by providing the perceiver with feedback about the target's actual thoughts and feelings? Third, are there stable individual differences in empathic accuracy that generalize across different targets? The results indicated that although absolute performance levels varied from 1 target to another, empathic accuracy generally improved with increased exposure to the target. In addition, feedback concerning the target's actual thoughts and feelings accelerated the rate at which the perceivers' empathic accuracy improved. Finally, cross-target consistency in responding ($\alpha = .86$) revealed stable individual differences in the perceivers' empathic ability. Implications of these findings for clinical training and practice are discussed.

Can I see another's woe,
And not be in sorrow too?
Can I see another's grief,
And not seek for kind relief?
—William Blake

Empathy has long been regarded as an important phenomenon by poets, playwrights, and philosophers. Like their counterparts in the literary world, psychologists from various research disciplines have focused attention on the role of empathy in mediating culturally valued social behaviors (e.g., Deutsch & Madle, 1975; Eisenberg-Berg & Mussen, 1978; Hoffman, 1977; O'Keefe & Sypher, 1981; Rogers, 1957, 1975; Selman, 1980). Interest in the empathy construct has been most evident in the interrelated fields of clinical, counseling, and educational psy-

chology (e.g., Egan, 1976; Feldstein & Gladstein, 1980; Goldstein & Michaels, 1985; Marks & Tolsma, 1986; Rogers & Meador, 1977). As Ickes (1993) noted, however, this construct is central to the interests of personality and social psychologists as well. Because people are motivated to understand each other's psychological states as well as each other's personality traits, the traditional study of accuracy in interpersonal perception is incomplete and can benefit from the complementary insights provided by the study of empathic accuracy.

Empathic Accuracy as an Area of Interpersonal Judgment Research

Oversimplifying a bit, Ickes (1993, pp. 588–590), identified four areas of study that are relevant to the accuracy of interpersonal judgments. The first area, which has the longest history of empirical study, focuses on perceivers' accuracy in judging other people's personality traits. Research in this area relies on interrater consensus as necessary but not sufficient evidence for accuracy in trait inference (e.g., Asch, 1946; Cronbach, 1955; Estes, 1938; Funder & Colvin, 1988; McCrae, 1982; Norman & Goldberg, 1966).

The second area, which has a shorter and more recent history of study, focuses on the accuracy of dyad members' perceptions or understanding of each other's attitudes, values, and self-conceptions. Research in this area involves comparisons of the dyad members' *direct perspectives* with their partners' *metaperspectives* regarding these relatively stable dispositions (e.g., Knudson, Sommers, & Golding, 1980; Laing, Phillipson, & Lee, 1966; Newmark, Woody, & Ziff, 1977; Rogers & Dymond, 1954; Sillars, 1989; Sillars & Scott, 1983).

The third area, which has an even more recent history of study, focuses on perceivers' accuracy, or *affective sensitivity*, in inferring the emotional state(s) of one or more target persons (e.g., Costanzo & Archer, 1989; Ekman & Friesen, 1975; Hall, 1978; Kagan, 1977a, 1977b; Noller, 1980, 1981; Noller & Ven-

Carol Marangoni and Gary Teng, Besselaar Associates, Princeton, New Jersey; Stella Garcia, Department of Psychology, University of Texas at San Antonio; William Ickes, Department of Psychology, University of Texas at Arlington.

This research was conducted as part of Carol Marangoni's doctoral dissertation at the University of Texas at Arlington. Portions of this research were presented at the centennial convention of the American Psychological Association, Washington, D.C., August 1992, and at the annual meeting of the Society of Experimental Social Psychology, San Antonio, Texas, October, 1992. Carol Marangoni would like to thank Ira Bernstein, Verne Cox, Joe Jackson, and Paul Paulus, who served on the dissertation committee for the project on which this study was originally based. She gives special thanks to William Ickes, who served as her advisor and dissertation committee chairperson. She also wishes to acknowledge the help of Vic Bissonnette, her research associate, and Steve Trued for his technical expertise.

Correspondence concerning this article should be addressed to Carol Marangoni, Besselaar Associates, 210 Carnegie Center, Princeton, New Jersey 08540, or to William Ickes, Department of Psychology, University of Texas at Arlington, Arlington, Texas 76019-0528. Electronic mail may be sent via the Internet to ICKES@UTA.EDU.

ardos, 1986; Rosenthal, Hall, DiMatteo, Rogers, & Archer, 1979).

The fourth area, which is only now emerging as a field of study, focuses on perceivers' *empathic accuracy*, that is, their ability to accurately infer the specific content of another person's thoughts and feelings (e.g., Ickes, Stinson, Bissonnette, & Garcia, 1990; Simpson, Ickes, & Blackstone, 1994; Stinson & Ickes, 1992).

As an area of study relevant to interpersonal perception, the empathic accuracy research is similar to the affective sensitivity research in its focus on relatively transient dispositions (i.e., psychological states that include thoughts, feelings, and emotional states of limited duration). It is conceptually and methodologically distinct, however, from the earlier clinical work on what Carl Rogers and his colleagues called *accurate empathy* (e.g., Butler & Haigh, 1954; Rogers, 1954; Rogers & Dymond, 1954). Because Rogers and his colleagues used a Q-sort methodology to assess the convergence in therapists' and clients' perceptions of the clients' self-concept, their work is in the tradition of the second area of research, which focuses on a class of dispositions (attitudes, values, and self-conceptions) that are somewhat more stable and enduring than the thoughts and feelings that are the focus of the current research on empathic accuracy.

Ickes (1993) noted an interesting parallel between the order in which empirical work began historically in each of the four areas and the type of disposition that is the focus of study in each. The first, and oldest, research area concerns perceivers' accuracy in judging dispositions that have traditionally been viewed as the most stable and enduring—personality traits. The second, and next-oldest, area concerns accuracy in judging dispositions that generally are regarded as somewhat less stable and enduring—attitudes, values, and self-conceptions. The third, and more recent, area concerns judgments about even more unstable dispositions—emotional states. And the fourth and most recent area concerns judgments about the most transient of dispositions—thoughts and feelings.

Why did interpersonal perception researchers begin by studying the more stable and enduring dispositions and only gradually move toward studying the more unstable and transient ones? Most likely, the reasons are both theoretical and methodological (Ickes, 1993, pp. 589–590). The major theoretical reason is the assumption by many psychologists that perceivers place greater reliance on the accuracy of their trait inferences than on the accuracy of their state inferences in predicting the behavior of other people. In other words, if traits are assumed to have the greatest long-term utility in predicting behavior, it makes sense that the study of accuracy in trait inference should have taken precedence over the study of accuracy in state inference. The major methodological reason is the relative ease of conducting paper-and-pencil studies of trait inference based on the perceivers' mental representation of the target's past behavior, in contrast to the relative difficulty of conducting state inference studies, which typically require having the target person physically present or providing the perceiver with records (video, audio, or written) of the target's actual behavior.

The goal of Ickes' (1993) review was to use findings from the empathic accuracy research to illustrate how the work in this area could inform the more general study of accuracy in inter-

personal perception. For example, findings in this area suggest that interpersonal perception researchers should devote more attention to (a) the history of the perceiver–target relationship (Stinson & Ickes, 1992); (b) the perceiver's desired future relationship with the target (Ickes, Stinson, et al., 1990; Simpson, et al., 1994); (c) the perceiver's lack of metaknowledge regarding his or her own empathic accuracy (Ickes, Stinson, et al., 1990; the present article); and (d) the perceiver's motivation, in certain conditions, to make inaccurate rather than accurate inferences about another person's dispositions (Simpson et al., 1994). In contrast, the goal of the present study was to extend the methodology of the empathic accuracy research to a more clinically relevant context in order to address some questions of fundamental interest to clinical psychologists as well as personality and social psychologists. Before considering these questions, however, we consider some theoretical and methodological issues that are implicated by previous attempts to study empathic accuracy in a clinically relevant setting.

Empathic Accuracy as an Area of Clinical Research

Since the publication of Rogers's (1957) classic article describing empathy as one of the "necessary and sufficient facilitative core conditions" for therapeutic change, many books and articles have been written on the role of empathy in the context of psychotherapy (Barrett-Lennard, 1962, 1981; Feldstein & Gladstein, 1980; Rogers, 1975; Truax & Carkhuff, 1967). In this literature, empathy—more specifically, *accurate empathy*—has been assigned a role of primary theoretical importance, because it is typically viewed as a prerequisite for successful therapeutic outcomes (Carkhuff & Burstein, 1970; Rogers, 1957, 1975; Truax & Carkhuff, 1967). It is not surprising, therefore, that more research attention has been devoted to the construct of empathy than to any other single variable purported to be of relevance to the psychotherapy process (Carkhuff, 1969a, 1969b; Elliot et al., 1982; Lamber, DeJulio, & Stein, 1978; Patterson, 1984).

As a consequence of the theoretical importance attributed to the empathy construct, graduate programs in clinical, counseling, and educational psychology have placed substantial emphasis on Rogerian models of human relations training (cf. Carkhuff & Anthony, 1979), with a particular focus on the training of empathic understanding and communication skills. However, investigators have been plagued by various problems in their attempts to assess and train this elusive ability. One of these problems is the different conceptual definitions of empathy (e.g., unidimensional versus multidimensional approaches, and approaches that distinguish between cognitive and emotional empathy, and between global and component subskills). With regard to the last of these distinctions, many researchers now advocate that empathy be conceptualized and assessed as a collection of individual, sequentially related subskills rather than as a global construct. Proponents of this *components approach* agree that researchers must recognize, clearly distinguish between, and devise training techniques suitable for each of three individual skills: empathic understanding, expression, and communication (Barrett-Lennard, 1981; Elliott et al., 1982; Goldstein & Michaels, 1985; Kagan, 1972, 1977b; Keefe, 1976, 1979).

When empathic accuracy is defined in strictly theoretical terms, this ability to accurately infer the content of other people's thoughts and feelings is clearly most synonymous with the first of the three components—empathic understanding. However, when empathic accuracy is operationally defined for the purpose of empirical study, it is typically necessary to expand its definition to include the second component—empathic expression—as well. The reason, of course, is that it is only when a perceiver's empathic understanding is expressed in some form that we are able to assess its accuracy. This reasoning led Ickes (1993) to propose that “the most straightforward way to measure empathic accuracy is to compare the content of a target person's actual thoughts and feelings with the content of the corresponding inferred thoughts and feelings reported by the perceiver” (p. 591).

If empathy research is to tap into the construct as conceptualized by Rogers (1975), then the methodology used must allow the perceiver to generate empathic inferences and to do so in dynamic or process terms rather than in a static fashion. For Rogers, “[empathy] involves being sensitive, *moment-to-moment*, [italics added] to the changing felt meanings which flow in this other person” (p. 4). It is important to note, however, that the vast majority of empathy measures and training procedures in both clinical and social-personality research fail to sample empathy at enough points to track its development over time (Carkhuff, 1969a, 1969b; Danish & Kagan, 1971; Davis, 1980; Hogan, 1969; Kagan, Krathwohl, & Farquhar, 1965; Kagan et al., 1967; Mehrabian & Epstein, 1972; Truax & Carkhuff, 1967). Finally, although the importance of accuracy in empathic inference is well recognized, an objective and reliable measure of empathic accuracy has been slow to emerge. With one notable exception (Kagan et al., 1967; Kagan, 1972), many of the commonly used training methodologies use as their accuracy criterion the empathic inferences generated by trainer-supervisor “experts,” and they do so despite the lack of research documenting the objective veracity of such inferences (Carkhuff & Burstein, 1970; Kurtz & Grummon, 1972; Truax & Carkhuff, 1967; Wilson & Griswold, 1985).

Ideally, a method could be developed that would enable researchers to (a) measure individual differences in empathic accuracy using an objective accuracy criterion, (b) assess these differences in the context of ongoing client-therapist interaction, and (c) monitor the changes in empathic accuracy that occur across time. The basis for such a method can be found in studies reported by Ickes, Stinson, et al. (1990) and by Stinson and Ickes (1992). In these studies, a videotape cued-recall procedure was developed to assess not only the actual thoughts and feelings reported by each participant in an unstructured interaction, but also the participant's inferences about each of the thoughts and feelings reported by his or her interaction partner. By having trained raters subsequently judge the similarity between the actual and the inferred thought-feeling entries, Ickes and his colleagues were able to derive summary measures of empathic accuracy that were highly reliable (α s = .94 and .95) in both studies.

Available data not only support the reliability of this procedure for measuring empathic accuracy but also provide converging evidence for the construct validity of the empathic accuracy measure in its network of relations with other variables.

For example, Ickes, Stinson, et al. (1990) found that this measure of empathic accuracy was significantly correlated with various indexes of the perceivers' degree of interest in their target interaction partners. These indexes of partner interest included the percentage of the perceivers' thoughts and feelings that concerned the partner, the percentage of partner attributions, and the number of times the perceivers looked at their partners during their interaction. Stinson and Ickes (1992) found that the empathic accuracy of male friends was significantly better than that of male strangers, a finding later replicated and extended by Graham (1994). Stinson and Ickes further demonstrated that the friends' advantage in this regard was largely attributable to their mutual knowledge about events occurring outside the experimental context. In addition, Stinson and Ickes found that the empathic accuracy scores of male friends were significantly correlated, whereas those of male strangers were not—a finding consistent with the notion that empathic accuracy is based on mutual self-disclosure and increased intersubjectivity. Indeed, a failure to achieve mutual self-disclosure and increased intersubjectivity may underlie Simpson et al.'s (1994) conceptually related finding that empathic accuracy scores were uncorrelated in unstable dating couples (those who reported that their relationship had ended 4 months following an initial assessment) but were significantly correlated in stable dating couples.

The Present Investigation

In the present investigation we adapted the method developed by Ickes and his colleagues in order to explore some issues of fundamental theoretical and practical importance to the study of empathic ability. Specifically, we addressed three primary questions concerning the development of empathic understanding in a clinically relevant context. First, does the empathic accuracy of a perceiver improve with increasing exposure to a target individual? Second, does the provision of immediate, veridical feedback about the target's actual thoughts and feelings affect the accuracy of subsequent empathic inferences? Third, are there stable individual differences in empathic accuracy that will generalize to subsequent target individuals?

Although the importance of these three questions has long been recognized within the clinical literature (see Goldstein & Michaels, 1985, for a review), the first question has received preliminary support in only two studies (Bernstein & Davis, 1982; Neimeyer, Neimeyer, & Landfield, 1983). The second question has, to the best of our knowledge, not been addressed in previous research on empathy or interpersonal judgment. Although the third question has stimulated at least a few studies—conducted primarily by personality and social psychologists in nonclinical settings (see Funder, 1994, for a review)—definitive evidence for the stable inferential ability that characterizes what Funder (1994) referred to as the “good judge” has yet to be provided. “Although, historically, the good judge is the first potential moderator [of accuracy effects] to have been addressed by the research, it remains the one upon which, to date, the accuracy literature has the sparsest data and fewest firm findings to report” (Funder, 1994, p. 23).

In the current study we also attempted to address some deficiencies in the existing research by (a) specifically focusing on one component of the empathy process (the assessment and

training of empathic understanding); (b) using a new methodology that provides an objective accuracy criterion (i.e., the clients' own recorded thoughts and feelings); (c) exploring the empathy process over a relatively extended time frame, thereby allowing sufficient time for the perceiver to become familiar with the target individuals; and (d) examining empathic understanding as an ongoing process, as posited by Rogers (1975). It is interesting to note that virtually all of these elements are also found in an innovative technique that Levenson and Ruef (1992) developed to relate perceivers' accuracy in judging targets' emotional states with the degree of physiological "linkage" between the perceiver and the target.

For various reasons, we conducted an analogue study in which participants inferred the thoughts and feelings of "normal" clients in simulated psychotherapy sessions. Because the clients in actual psychotherapy sessions represent a range of clinical populations and psychological disorders, they might be expected to exhibit less insight and more confusion about their thoughts and feelings than "normal" individuals would. In addition, clients suffering from psychological disorders might be less willing to reveal their thoughts and feelings, because such revelations might make them feel ashamed, deviant, unworthy, or vulnerable. Other considerations included the legal and ethical problems associated with recruiting clinically disturbed individuals to participate in a self-disclosure experience whose potential impact on their own lives and therapeutic outcomes was unclear. Accordingly, before trying to demonstrate the utility of providing perceivers with feedback about the thoughts and feelings of "disturbed" clients whose ability and motivation to report accurately in this regard might be impaired, we first sought to demonstrate the utility of providing perceivers with feedback about the thoughts and feelings of "normal" clients who were both willing and able to report these experiences.

Analogue studies of this type are common in other applied areas; for examples, see the elegant courtroom simulation studies conducted by Borgida and his colleagues (e.g., Borgida, 1979, 1980). Because such studies typically involve a trade-off between internal and external validity, however, reactions to them can vary greatly depending on which side of the trade-off one values most. Accordingly, we viewed the present study as a first-approximation attempt to explore some issues that clinical psychologists as well as personality and social psychologists have long sought to address, with the clinical relevance of this work suggested but in no way guaranteed.

Method

Participants and Design

The participants were 80 undergraduate students (40 male and 40 female) who took part in exchange for experimental credit in their introductory psychology course. At the start of the semester, all participants completed a pretest questionnaire containing the Empathic Accuracy Scale (Ickes, 1988), the Interpersonal Reactivity Index (Davis, 1980), and the Crowne-Marlowe Social Desirability Scale (Crowne & Marlowe, 1960). These measures proved to have only minimal relevance to the findings of this study and are not discussed further. Their failure to predict empathic accuracy in this and other studies is discussed extensively by Davis (1994, pp. 85-93) and by Ickes (1993, pp. 603-606), who suggested several reasons why this failure could be at-

tributed to participants' lack of metaknowledge regarding their own empathic ability.

The experimental design was a 2×2 factorial in which two levels of feedback (target-generated feedback or no feedback) were crossed with two orders of videotape presentation (Target 1-Target 2-Target 3 or Target 1-Target 3-Target 2). Sex of participant was treated as a blocking variable. An equal number of male and female students were randomly assigned to the four experimental conditions, with 20 students (10 male and 10 female) per cell.

Preparation of Stimulus Tapes

Four female volunteers (all acquaintances of the experimenter) independently participated in simulated psychotherapy sessions ranging from 30 to 55 min in length. The 4 volunteers were informed that their session would be conducted by a licensed male therapist trained in the Rogerian (nondirective) tradition, and that they should come prepared to discuss one or more real, ongoing problems in their lives as openly and directly as possible. The sessions were videotaped "live" from beginning to end without any prior rehearsal, and the genuineness and spontaneity of the sessions were evident in the volunteers' range of emotional expression (e.g., one woman wept openly as she discussed her recent divorce).

Immediately following the videotaped session, each target was escorted to a private cubicle to view the videotape. Targets were instructed to make a complete log of all the thoughts and feelings that they experienced during the therapy session by recording the time (displayed by a digital timer on the videotape) as well as the content and the valence of their recollected thought or feeling each time they paused the tape. The targets' thought-feeling entries were recorded on standardized forms available in the cubicle. Following the thought-feeling recording task, all targets agreed to sign a release form allowing their videotape and thought-feeling data to be used as experimental stimulus materials in the present study.

The thought-feeling data obtained from three of these videotargets served as the objective criterion against which the empathic accuracy of the participants in the present study was subsequently assessed. Consequently, care was taken to minimize the possible biasing influence of self-censorship and selective omissions. Both the verbal and the written instructions to the targets emphasized the importance of recording each and every recollected thought and feeling as accurately, honestly, and completely as possible. Targets were further told that immediately following the task (or anytime thereafter) they could delete any recorded entries and portions of videotape that they would prefer remain private. These efforts to eliminate potential reporting biases resulting from self-censorship were considered successful in that three of the four videotargets later chose to delete some of the personal material they had disclosed. Consistent with their wishes, these videotaped and written thought-feeling data were not included in the stimulus materials.¹

Three of the four target videotapes were subsequently selected for use

¹ Another potential source of bias involves the possibility that the targets would report "new" thoughts and feelings that they had first experienced while *viewing* the videotapes. To help minimize this bias, the therapist was instructed to deliberately structure the first few minutes of each therapeutic session as an informal "acclimation period." The casual conversation that occurred during this period was intended to increase the targets' comfort and allow them to become accustomed to their appearance on videotape before they began the recall task. The rationale for the acclimation period was explained to the targets, who were told not to begin recording their recollected thoughts and feelings until the point in the tape at which they had begun discussing their personal problems. In addition, the importance of reporting recollected

as experimental stimulus materials. The criteria for selection included (a) consensual agreement (among groups of 5–12 graduate and undergraduate psychology students) regarding the inherent interest value of the discussion and the degree of an observer's emotional involvement with the target, and (b) a minimum of 30 thought–feeling entries generated by the target during the thought–feeling listing task. These criteria were used to ensure the selection of interesting and “affectively laden” stimulus materials that would effectively simulate the psychotherapy milieu and minimize the chances that the participants in the present study would become bored during the empathic accuracy task.

Three videotargets were used so that we could test the generalizability of participants' empathic accuracy across different targets. Practical constraints (i.e., the lengthiness of the procedure—4 hr to run each participant) mandated that the number of targets be limited to no more than 3. Female rather than male targets were chosen, on the assumption that women would more willingly disclose personally meaningful, intimate concerns, and that they would do so in a more expressive fashion.

Stimulus Materials

The stimulus materials consisted of pre- and postexperimental instructional videotapes and videotapes of the 3 volunteer clients individually participating in simulated psychotherapy sessions. The videotargets were all white, college-educated, 24- to 32-year-old women from middle- to upper-middle-class backgrounds. Two of the targets chose to discuss the topic of divorce; their tapes are referred to as Divorce 1 and Divorce 2. The third target chose to discuss the conflicts associated with the dual social roles of wife and career woman; her tape is referred to as Role Conflict. The unedited videotapes ranged from 30 to 50 min in duration and contained between 30 and 46 thought–feeling entries. These original tapes were edited to create final versions that ranged from 26 to 36 min in duration and contained 30 thought–feeling entries each.

Sample thought–feeling entries reported by the target in the Divorce 1 tape are: “I was thinking about when I first started to resent him,” and “I was feeling sad about how long I denied things.” Sample entries reported by the target in the Divorce 2 tape are: “I was thinking that we really didn't have *that* much fun—at least not enough to be happy,” and “I was thinking about how hard it would be to face him and admit it is over.” Sample entries reported by the target in the Role Conflict tape are: “I was feeling that although my husband tries outwardly not to impose on me, he does in fact impose by actions or insinuations,” and “I was thinking if my husband could only understand me or switch places with me for awhile.”

Each stimulus tape could be divided into three phases without disrupting the natural flow of events on tape. Phases 1 and 3 contained 8 entries each, whereas Phase 2 contained 14 entries. The entries were relatively evenly distributed within each phase, with a mean interentry interval of 1 min. Consistent with the nondirective psychotherapeutic nature of the interaction, all targets generated a greater number of feeling entries than of thought entries, most of which were negatively valenced (rather than neutral or positive). There were, however, no significant differences *among* targets in terms of (a) the number of thought

versus feeling entries, or (b) the number of positively, negatively, and neutrally valenced entries generated (see Table 1).

Phase 1 was designated the baseline phase, Phase 2 was designated the training phase, and Phase 3 was designated the test phase. Phase 2 contained the largest number of entries to maximize the number of feedback training trials that the participants received. It seemed reasonable to expect that if 42 training trials (14 feedback trials \times 3 tapes) were adequate for an overall feedback effect to occur, then the viability of this procedure as an empathy training technique could be assessed within the confines of a short-term laboratory experiment. On the other hand, if more than 42 feedback trials were required, there would seem to be little point in trying to demonstrate an overall feedback effect in a laboratory experiment as opposed to a more extensive training intervention.

The experimental instructions administered to all of the participants in the study were standardized through the use of an instructional videotape. This videotape contained three 30-s videoclips of the target individuals, and the students were asked to press a signal button if they were acquainted with any of the videotargets (no cases of prior acquaintance with a target were reported). The experimenter on the instructional tape then proceeded to describe the inferential task to be performed. For the students in the experimental (i.e., feedback) condition, instructions explaining the feedback procedure were edited onto this instructional tape.

All participants were informed that a computer message would appear on their video monitor twice during each target tape. These messages were presumably provided as a procedural check to ensure that the participants were working on the appropriate entry on the thought–feeling inference form. The first message would appear after the participants completed their 8th inference (i.e., at the end of Phase 1), and the second message would appear after they completed their 22nd inference (i.e., at the end of Phase 2). The participants in the feedback condition were further informed that the first message would also serve to signal the onset of the feedback trials, whereas the second message would serve to signal the termination of feedback.

The postexperimental instructional tape requested that the students fill out a short questionnaire that was enclosed in a manila envelope located on the back of their cubicle door. Before they completed the questionnaire, the students were shown brief videoclips of the 3 target individuals (in the order in which they appeared during the experiment) and were required to answer specific questions regarding the degree to which they could relate to each of the targets and their presenting problems.

Procedure

All participants were contacted by phone the day before they were scheduled to participate, to remind them of their appointment and give them directions to the experimental laboratory. In each session, 1 or 2 participants were met by the experimenter at the prearranged meeting area and were taken individually to adjacent experimental cubicles.² They were informed that they would be watching three videotapes on

instead of novel thoughts and feelings was stressed repeatedly throughout the targets' instructions.

As a final precaution, while targets were engaging in the thought–feeling recording task, the experimenter simultaneously viewed the videotape on a monitor in an adjoining control room, to record the exact time of each pause in the videotape. The times recorded by the experimenter were then compared with the times recorded by the targets themselves as an additional check on the accuracy of the target's reports and their freedom from self-censorship.

² Of the 80 participants whose data were included in the study, 22 took part individually and 58 took part in pairs (i.e., 29 paired observations). Although the members of each pair were seated in separate cubicles immediately after being met by the experimenter, a check on the independence of their empathic accuracy scores was desirable. Intraclass correlations computed at each level of target and phase revealed no evidence of interdependence in the empathic accuracy scores of the paired participants. These six correlations were all nonsignificant; they ranged from $-.15$ to $.25$, with a mean of $.06$. Accordingly, all findings reported in this study were based on tests that used the individual participant as the unit of analysis (cf. Kenny & La Voie, 1985).

Table 1
*Chi-Square Tests of Possible Differences Between Targets in the
 Distribution of Entries by State and Valence*

Type of entry	Target			χ^2	df	$\chi^2_{.05}$
	Divorce 1	Divorce 2	Role conflict			
State						
Thought	9	13	16	3.39 ^a	2	5.99
Feeling	21	17	14			
Valence						
Positive	5	5	9	5.83 ^a	4	9.49
Negative	20	20	12			
Neutral	5	5	9			

^a *ns.*

the television monitor and that all instructions regarding the task would be presented on instructional videotapes. The participants were then given an opportunity to read and sign the informed-consent form.

After obtaining the informed consent, the experimenter went to an adjoining control room to initiate the instructional tape. After the instructional tape was shown, the experimenter activated the first of the three videotarget stimulus tapes. The participants in all conditions individually viewed each of the stimulus tapes, with the order of tape presentation counterbalanced among participants. The two presentation orders were designated as Order 1 (Divorce 1–Role Conflict–Divorce 2) and Order 2 (Divorce 1–Divorce 2–Role Conflict).³

Using the digital timer that appeared on each videotape to keep track of the time, the experimenter (or her assistant) manually paused the stimulus tapes for the participant at each of the 30 points at which the target had previously reported having had a thought or feeling. These tape stops were the same for each participant *within* each tape, but the temporal spacings of course differed *between* tapes on the basis of the targets' reporting of when their thoughts and feelings had actually occurred. The participants' task was to record their inferences regarding the state (i.e., thought or feeling), the content, and the valence (i.e., positive, negative, or neutral) of each particular thought–feeling entry, using standardized thought–feeling inference forms of the type described by Ickes, Bissonnette, Garcia, and Stinson (1990, pp. 29–32).⁴

The verbatim instruction for the content inference task was "Every time I pause the tape, you are to write down what you think [the target person] was thinking or feeling at that moment by filling in one of these slots [on the thought–feeling coding form]." In other words, just as the targets had been instructed to write down what they were actually thinking and feeling at each of the stop points they selected, the perceivers were asked to infer the targets' actual thoughts and feelings at each of those points. It is important to note that the perceivers' task was not complicated by a requirement that they try to infer possible feelings and motives of which the targets themselves were unaware, or that they attempt to distinguish between what the target was actually thinking or feeling and what she might report she was thinking or feeling. Instead, the perceivers' task was to make a straightforward inference about what each target was actually thinking or feeling at each of the stop points on the tape.

When they finished recording their responses for a given thought–feeling entry, the participants restarted the tape by means of a remote control device located in each cubicle. The tape then continued playing until it reached the point at which the target had reported her next thought or feeling, when it was again manually paused by the experimenter, and the process was repeated. This basic procedure was in effect throughout the entire session for the control group and throughout the baseline and test phases for the experimental group. It was modified during Phase 2 training for the experimental group as follows.

After recording their written inferences for a given thought–feeling entry, the experimental (i.e., feedback condition) participants restarted the tape in the standard fashion. The subsequent 20 s of videotape then displayed the actual target-generated feedback for that entry, allowing participants time to read the feedback and compare it with their own written inference. Each feedback entry had previously been typed on a computer terminal that was connected to the VCR and edited onto the stimulus tapes. Following each 20-s feedback interval, the tape resumed playing until the designated stop point for the next target-reported thought or feeling. Experimental participants received feedback in this fashion following each of the 14 thought–feeling inferences made during Phase 2. Feedback was not provided during Phase 3 (i.e., the test phase), when the procedure for all participants was identical to that instituted during Phase 1 (the baseline phase).⁵

When each participant had completed all three stimulus tapes and had filled out the postexperimental questionnaire, the experimenter returned to provide a complete debriefing and an opportunity for the participant to ask questions about the experiment.

³ Although fully counterbalancing all three tapes would have been desirable, it was not tenable because of (a) the lack of a sufficiently large number of eligible participants, and (b) the formidable time investment that was required by the experimental procedure (i.e., 40 additional participants at 4 hr per participant). It was therefore decided to begin both sequences with the Divorce 1 tape, an ordering that permitted a comparison to determine whether there might be a greater transfer of feedback training to a subsequent tape that portrayed a similar (Divorce 2) versus a dissimilar (Role Conflict) content topic.

⁴ Consistent with our experience in other studies, subsequent analyses of the participants' accuracy in inferring whether the target reported a thought or a feeling, or whether the valence of the thought–feeling entry was positive, negative, or neutral, did not yield any significant effects that were relevant to the research questions we addressed. Accordingly, the data pertaining to these types of inference are not discussed, and the results focus instead on the participants' accuracy in inferring the content of the specific thoughts and feelings that the targets reported.

⁵ Alternate modes of feedback delivery were considered but ultimately rejected. Auditory feedback was unacceptable because of potentially confounding fluctuation in vocal intonation or other paralinguistic cues (cf. Mill, 1984). Another problem was the potential contamination introduced through eavesdropping by participants in the adjacent cubicle. Although the use of headphones might have solved this problem, such a procedure would also have introduced a confounding element unique to the experimental group.

Task Interest and Motivation

Although precautions were taken to standardize the experience of all participants as much as possible, the fact remains that control group participants did not experience comparable post-inference conditions during the Phase 2 training trials. Although there were no a priori expectations that differences in interest or motivation would occur, two questions were included on the postexperimental questionnaire that asked participants to rate (a) their degree of motivation on the task and (b) how interesting or boring they found the task to be. These questionnaire data permitted the assumption that there was no interest or motivational difference to be tested directly.

Computation of Empathic Accuracy Scores

On the basis of the logic and procedures developed by Ickes and his colleagues (Ickes & Tooke, 1988; Ickes, Stinson, et al., 1990; Stinson & Ickes, 1992), we computed empathic accuracy scores by comparing each participant-generated inference with the corresponding thought-feeling entry obtained from the videotargets. Both the target-generated thought-feeling entries and the participant-generated inferences were typed into word processor files and stored on diskettes. A custom software program was developed to access the diskette files and present the entry-inference pairs on a monitor screen to four raters who individually rated the empathic accuracy of each inference.

Separate copies of the program and corresponding data disks were distributed to each of the four raters. Their task was to compare the written content of each target entry with that of the participant's corresponding inference and rate the degree of similarity (i.e., empathic accuracy) using a 3-point scale ranging from 0 (*essentially different content*) through 1 (*somewhat similar, but not the same, content*) to 2 (*essentially the same content*). In the present study, the internal consistency (Cronbach's alpha) of the four judges' empathic accuracy ratings was .85. The raters recorded their empathic accuracy ratings directly on the data disks, and the software program then computed average accuracy scores (i.e., the mean of the accuracy scores assigned by the four raters) for each individual inference.

The same software program then computed composite (i.e., aggregated) accuracy scores from the averaged scores for each inference. First, the mean ratings were summed across the 8 thought-feeling inferences within each phase (baseline, test) for each target (Divorce 1, Divorce 2, Role Conflict). These summed values were divided by 16, the maximum number of accuracy points that could be obtained in each phase (i.e., 8 inferences \times 2 points possible per inference) to derive composite accuracy scores that represented the percentage of "accuracy points" obtained by the participant within each phase and target. These percentages are easy to interpret because they have a possible range of 0 (*total inaccuracy*) to 100 (*total accuracy*); thus, a score of 50 means that the participant obtained half of all the accuracy points that were possible to obtain for a given set of thought-feeling inferences. Both intraphase and global accuracy scores (i.e., scores composited across phase for a given target) were used as the dependent measures of empathic accuracy.

Results

The full statistical model consisted of three between-subjects variables (feedback vs. no feedback, Order 1 vs. Order 2, male vs. female participants) and two within-subjects repeated-measures variables (three target tapes and three phases per tape).

No main or interaction effects were found for the order of tape presentation; therefore, the data were collapsed across the order variable and reanalyzed. We did not include the data from the feedback training phase (Phase 2) in the analyses because

Table 2
Main Effects on Empathic Accuracy of the Phase, Target, and Feedback Variables

Variable	M	SD
Phase		
1 (baseline)	25.29	9
3 (test)	35.37	11
Target		
Divorce 1	34.13	12
Divorce 2	23.62	7
Role conflict	33.25	10
Feedback		
Feedback	32.33	7
No feedback	28.33	10

Note. The empathic accuracy data are reported as mean percentages of the total accuracy points possible.

this phase represented an empathy training intervention (the major independent variable), which, if successful, should have enhanced the feedback participants' post-training performance at Phase 3 (the major dependent variable). For this reason, the results reported below are based on the analyses of the baseline (Phase 1) and test (Phase 3) data. Unless noted otherwise, the modified model is a $2 \times 2 \times 3 \times 2$ design, consisting of two between-group variables (two levels each of feedback and gender) and two within-group variables (three targets and two levels of phase within each target).

Empathic Accuracy: Analysis of Variance

Main effects. The analysis revealed significant main effects for phase, $F(1, 76) = 101.91, p < .0001$; feedback, $F(1, 76) = 4.49, p < .03$; and target, $F(2, 152) = 60.11, p < .0001$. The descriptive statistics for these findings are summarized in Table 2. In general, the participants more accurately inferred the content of the targets' thoughts and feelings during the test phases (Phase 3) than during the baseline phases (Phase 1) of the three tapes. In addition, participants who received target-generated feedback during the second phase of each tape were generally more accurate than control participants, who received no feedback. Finally, participants were generally more accurate in inferring the thoughts and feelings of the targets in the Divorce 1 and Role Conflict tapes than those of the target in the Divorce 2 tape.⁶

As in four other studies using the same thought-feeling inference measure of empathic accuracy (Hancock & Ickes, 1994; Ickes, Stinson, et al., 1990; Simpson et al., 1994; Stinson & Ickes, 1992), there was no hint of a main effect for the gender of the perceivers in this study. All of the other studies yielded $F_s < 1$ for the test of the gender main effect, and the same was true in

⁶ Such differences were not unexpected, in that different reactions to the targets were exhibited by the participants during debriefing. Specifically, the Divorce 1 target seemed to elicit the greatest degree of admiration, whereas the Divorce 2 target seemed to elicit a generally negative reaction from participants because of her inability, after 3 years, to terminate a marital relationship that was unhappy but that offered great financial security.

the present case, $F(1, 76) = 0.29, ns$. Although women appear to display a slight but statistically significant advantage over men in their ability to correctly identify others' emotional expressions (see Hall, 1987, for a review), they do not display a corresponding advantage in their ability to infer the specific content of other people's thoughts and feelings.

Phase × Feedback interaction. The main effects for phase and feedback were qualified, however, by a significant Phase × Feedback interaction, $F(1, 76) = 4.32, p < .04$. This interaction, represented in Figure 1, reveals that there was no significant difference in overall performance between the participants in the feedback ($M = 0.26$) and the no-feedback ($M = 0.24$) conditions during Phase 1, whereas a significant post-training difference was evident during Phase 3 ($M_s = 0.38$ and 0.32 , respectively, $t[76] = 2.21, p < .03$).

In general, the data in Figure 1 indicate that although empathic accuracy improved for *all* participants from Phase 1 to Phase 3 (an apparent "practice" or "acquaintanceship" effect), the experimental participants who received a relatively modest number of feedback training trials showed significantly better performance during Phase 3 than did control participants, who received no target-generated feedback. In other words, the provision of target-generated feedback appeared to accelerate the rate at which the participants' empathic accuracy improved across trials.⁷

Phase × Target interaction. The main effects for phase and target were qualified by a significant Phase × Target interaction,

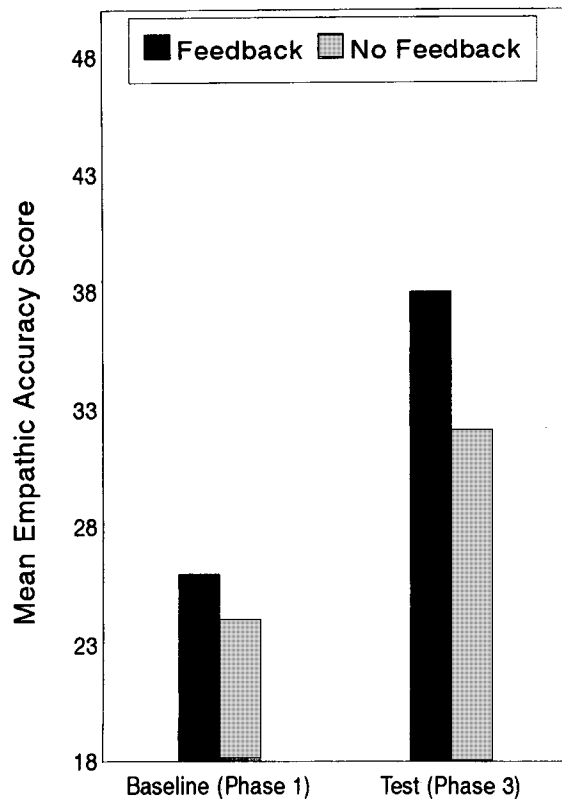


Figure 1. Phase × Feedback interaction. Empathic accuracy scores for feedback versus no-feedback participants.

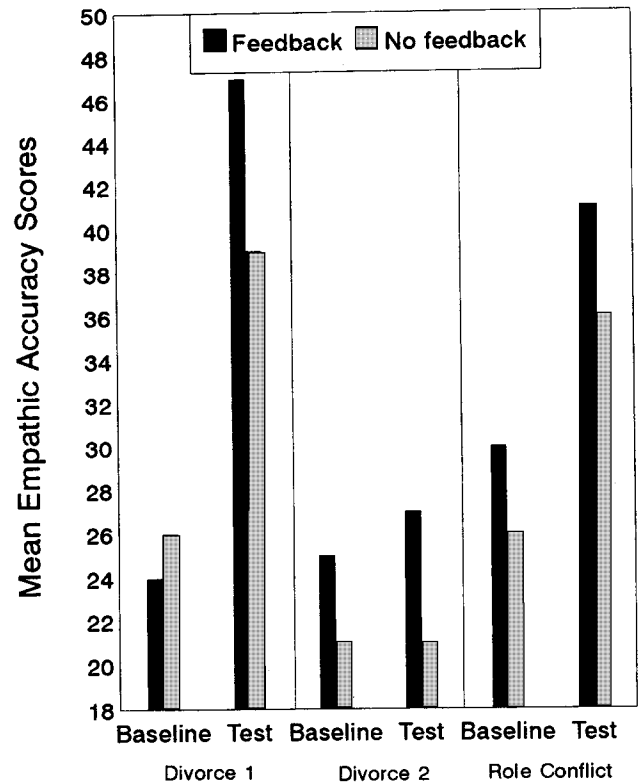


Figure 2. Individual target analysis for feedback and no-feedback conditions.

$F(2, 152) = 26.85, p < .0001$. As can be seen in Figure 2, this interaction was based on the fact that the participants' performance improved significantly from the baseline to the test phase for the Divorce 1 and Role Conflict tapes (both $p_s < .001$), whereas it did not improve significantly for the Divorce 2 tape.

The reason for the difficulty participants had in "getting to know" the woman in the Divorce 2 tape was clearly evident in the content of the tape itself. The woman in Divorce 2 was highly ambivalent about whether to divorce or not, having been in an indecisive holding pattern for the previous 3 years. Her extreme ambivalence, which remained unresolved throughout the tape, led her to vacillate in expressing first positive, then negative, then positive, and so on, thoughts and feelings about her marriage. Apparently, as Figure 2 suggests, the control (no-feedback) participants understood her no better at the end of the tape than at the beginning, whereas the general advantage of the experimental (feedback) participants was evident at both points. In contrast, the women in the Divorce 1 and Role Con-

⁷ The posttest data suggested that the effect of feedback training was not attributable to differential self-reported task interest or motivation between experimental and control group participants. No significant between-group differences were obtained for the mean level of task interest, $t(78) = -0.09, ns$, or the mean level of task motivation, $t(78) = 1.25, ns$. Thus, the enhanced Phase 3 accuracy on the part of feedback participants did not appear to be a function of differential task interest or motivation.

flict tapes both gave internally consistent accounts of their problems: The first woman described the personal difficulties that led up to and followed her divorce, and the second woman elaborated in sequence the various problems resulting from the competing demands of her roles in the workplace and at home. As time went by, all participants seemed to gain a better understanding of the nature of these two women's thoughts and feelings.

Comparison of effects within each target. To further explore the differences in the participants' reactions to the different targets, we conducted analyses of variance (ANOVAs) on the data for each target to see how the pattern of results may have differed from one tape to the next. These analyses revealed a different pattern of effects for the Divorce 1 tape, which all participants saw first, than for the Divorce 2 and Role Conflict tapes, which all participants saw subsequently (with their order counterbalanced, as noted above).

For the Divorce 1 tape, which was the first stimulus tape in the series, there was a significant main effect for phase, $F(1, 76) = 93.08, p < .0001$, and a significant Phase \times Feedback interaction, $F(1, 76) = 7.15, p < .01$. As the left panel of Figure 2 reveals, although there was no initial (Phase 1) difference between the feedback and the no-feedback groups, the facilitative effect of the feedback training was already evident during the test phase of Divorce 1, $t(76) = 2.09, p < .04$. This finding indicates that the feedback training significantly enhanced empathic accuracy in the first training tape the participants viewed, after only 14 feedback training trials.

Although the absolute performance level was higher for Role Conflict than for Divorce 2, the results for both of these tapes were similar in revealing a general (facilitative) main effect for feedback, $F_s(1, 76) = 6.79$ and $4.36, p_s < .01$ and $.04$, without the qualifying Phase \times Feedback interaction found for Divorce 1.

Serial order comparison. To explore how the findings for all three tapes might reflect the cumulative effect of the feedback training across time, we conducted another set of analyses to provide a serial order comparison of the tape participants saw first (Divorce 1) with the tapes they saw second and third (either Divorce 2 followed by Role Conflict, or vice versa). For this serial order comparison, the second and third tapes were collapsed across targets, a procedure that was justified given the lack of any significant effects for the order in which the second and third tapes were presented.

Comparisons of the means depicted in Figure 3 (and reported in Table 3) revealed that significant differences between the feedback and the no-feedback participants occurred in the test phase of Tape 1, $t(76) = 2.09, p < .04$; the baseline phase of Tape 2, $t(76) = 2.38, p < .02$; and the test phase of Tape 3, $t(76) = 1.99, p < .05$. These findings indicate that (a) no prior differences existed between the feedback and no-feedback groups (i.e., no significant differences were evident in the baseline phase of Tape 1); (b) significant differences between the feedback and no-feedback groups emerged as early as the first test phase (Tape 1, Phase 3); and (c) participant fatigue did not hamper performance over time, in that significant feedback training effects were evident during the test phase of the last tape that the participants viewed.

The significant difference that emerged between the feedback

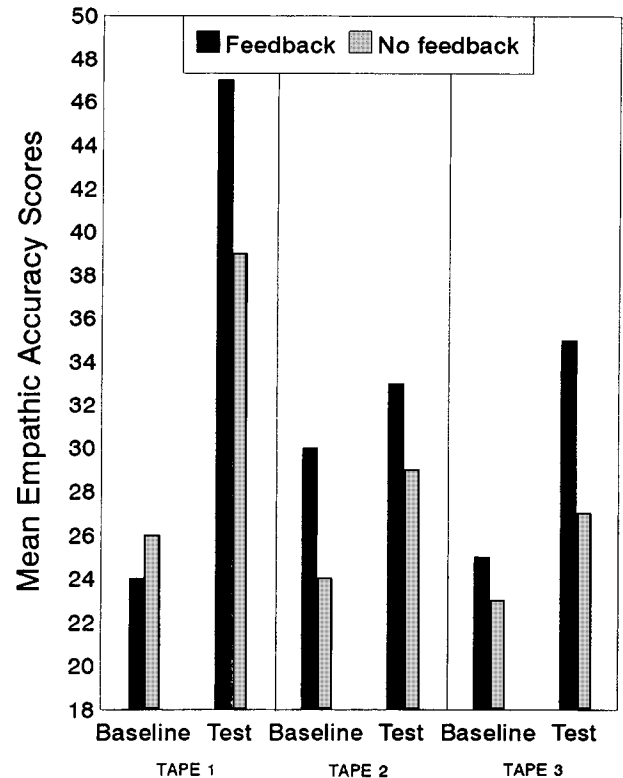


Figure 3. Serial order analysis: Mean empathic accuracy scores for feedback and control group participants.

and the no-feedback groups during the baseline phase of Tape 2 may reflect a feedback training carryover effect from Tape 1. By extension, the significant feedback main effects found for both the Divorce 2 and Role Conflict tapes suggest that once the experimental participants had benefited from the feedback they received on Tape 1, their performance on Tapes 2 and 3 was generally enhanced relative to that of the control participants. This enhanced performance, which was evident even during the baseline phase of Tape 2, was again strong enough to be reliable during the final test phase of Tape 3. Although the same difference did not attain a conventional level of significance for the initial, baseline phase of Tape 3, a trend in this direction was still evident. In addition, inspection of Figure 2 (individual target analysis) reveals enhanced baseline performance for feedback compared with no-feedback participants on the Divorce 2 and Role Conflict tapes, whereas this difference did not emerge on the initial tape that the participants viewed (Divorce 1).

Stability of Performance Across Targets

Another way to conceptualize the participants' performance across the three tapes is in terms of an underlying ability component of empathic accuracy that generalizes to different target individuals. To test the possibility that there might be cross-target consistency in responding, we computed Cronbach's alpha for the participants' empathic accuracy scores for each of the three videotargets. The results revealed an intertarget reliability coefficient of

Table 3
Mean Empathic Accuracy Scores for the Feedback and No-Feedback Participants (Serial Order Analysis)

Phase	Tape 1			Tape 2			Tape 3		
	Feedback	Control	<i>p</i>	Feedback	Control	<i>p</i>	Feedback	Control	<i>p</i>
Baseline									
<i>M</i>	24.10	26.32	<i>ns</i>	29.53	24.10	< .02	25.11	22.57	<i>ns</i>
<i>SD</i>	11	14		9	11		9	12	
Test									
<i>M</i>	46.87	39.21	< .04	33.28	29.45	<i>ns</i>	35.07	28.35	< .05
<i>SD</i>	14	18		13	15		15	13	
Combined									
<i>M</i>	35.48	32.77	<i>ns</i>	31.40	26.77	< .05	30.09	25.46	< .05
<i>SD</i>	9	14		11	13		12	13	

Note. The empathic accuracy data are reported as mean percentages of the total accuracy points possible.

.86, indicating a high degree of performance stability across the three targets.

If an ability component does indeed underlie the cross-target stability of the participants' empathic accuracy scores, then the effects of the feedback training might be expected to alter and therefore disrupt this stability. Thus, cross-target stability should be more evident for the no-feedback than for the feedback participants. In a test of this reasoning, we examined the intercorrelations of the participants' accuracy scores for each of the three

targets and in all three phases separately by feedback condition. As can be seen in Table 4, the cross-target stability of individual performance was indeed disrupted by the provision of feedback information. When feedback was given, only 3 out of 15 (20%) of the correlations among the individual phase scores were significant. In contrast, when no feedback was given, 15 out of 15 (100%) of these correlations were significant.

Table 4 also reveals that for both the feedback and the no-feedback participants, highly significant intercorrelations were ob-

Table 4
Significant Correlations Between the Target × Phase Empathic Accuracy Scores as a Function of Feedback Condition

Target/Phase	1	2	3	4	5	6	7	8	9
Feedback condition									
1. Div1-P1	—								
2. Div1-P3		—	.41**			.47***			
3. Div2-P1			—						
4. Div2-P3				—		.43**			
5. RC-P1					—				
6. RC-P3						—			
7. Div 1							—	.52***	.50***
8. Div 2								—	.52***
9. RC									—
No-feedback condition									
1. Div1-P1	—	.56***	.57***	.31*	.60***	.35*			
2. Div1-P3		—	.48***	.36*	.47***	.46***			
3. Div2-P1			—	.39**	.55***	.43***			
4. Div2-P3				—	.47***	.35*			
5. RC-P1					—	.43**			
6. RC-P3						—			
7. Div 1							—	.59***	.61***
8. Div 2								—	.62***
9. RC									—

Note. Div1-P1 = Divorce 1, Phase 1 (baseline); Div1-P3 = Divorce 1, Phase 3 (test); Div2-P1 = Divorce 2, Phase 1 (baseline); Div2-P3 = Divorce 2, Phase 3 (test); RC-P1 = Role Conflict, Phase 1 (baseline); RC-P3 = Role Conflict, Phase 3 (test).

p* < .05. *p* < .01. ****p* < .005.

tained among the global empathic accuracy scores for each of the three targets. These correlations tended to be higher, however, in the no-feedback condition (.59–.62) than in the feedback condition (.50–.52), again suggesting that cross-target stability in individual performance is most clearly evident in the absence of a strong situational moderator of empathic accuracy.

Subjective Perceptions of Empathic Accuracy

To examine the relationship between the participants' perceptions of their empathic accuracy and their tested empathic accuracy, the following posttest question was included for each target: "How well do you think you inferred [the specific videotarget's] thoughts and feelings?" Following a brief videoclip of each target on the posttest tape, the participants responded to this question on a 5-point scale anchored by *not at all* and *extremely well*.

Interestingly, the results revealed virtually no relationship between perceptions of empathic accuracy and actual empathic accuracy (see Table 5). For the sample as a whole, subjective perceptions of empathic accuracy were uncorrelated with actual accuracy. Similarly, when the data were examined separately by gender and feedback condition, perceptions of accuracy and actual accuracy were again found to be unrelated. Consistent with similar null findings reported by Ickes, Stinson, et al. (1990), these null findings suggest that most perceivers have little insight regarding their actual level of empathic ability.

Topic Similarity and the Generality of Feedback Training Effects

The method used to counterbalance the order of tape presentation was specifically designed to enable us to test whether any feedback training effects would be more likely to generalize to a second target tape that was similar versus dissimilar in content to the first tape the participants viewed. In other words, after receiving feedback during Phase 2 of the Divorce 1 tape, would participants exhibit a greater carryover effect of the training if the second tape they viewed had a similar topic (Divorce 2) than if it had a dissimilar topic (Role Conflict)?

In this case at least, the answer was no. Despite our a priori

judgment regarding the greater topical similarity of the Divorce 1 and Divorce 2 tapes, the Divorce 1 target elicited the highest overall empathic accuracy scores, whereas the Divorce 2 target elicited the lowest. Indeed, absolute performance levels were more similar on the Divorce 1 and Role Conflict tapes ($M_s = 0.34$ and 0.33 , respectively) than they were on the two presumably similar tapes ($M = 0.34$ and 0.24 for Divorce 1 and Divorce 2, respectively). As noted previously, the fact that the Divorce 2 tape was unique among the three in failing to exhibit a main effect for phase suggests that participants (in particular, the no-feedback group) failed to benefit from increased familiarity with this particular target over time.

Reanalysis Adjusting for Immediate Context Influences on Inferential Ease Versus Difficulty

One of the reviewers of this study raised the question of whether the significant overall main effect for phase was really attributable to the perceivers' increased understanding of the targets based on greater familiarity with their problems and perspectives, or whether this phase effect was instead an artifact of the Phase 3 thoughts and feelings being coincidentally easier to infer than those in Phase 1. To address this issue, we had seven independent raters view each of the three videotapes and rate the difficulty of inferring each thought or feeling based on the information available in the words and actions of the videotarget and therapist in the 10-s period immediately before the specific thought or feeling was reported. The raters were provided with copies of the target's actual thoughts and feelings, to which they could refer at each tape stop before making their difficulty rating on a 3-point scale ranging from 1 (*very difficult to infer given the immediate context*) to 2 (*somewhat difficult to infer given the immediate context*) to 3 (*easy to infer given the immediate context*).

For the resulting set of 48 difficulty ratings (three tapes \times two phases [baseline and test] \times eight thought-feeling entries per phase), the interrater reliability (computed as Cronbach's alpha) for the seven raters was .80. An ANOVA of these difficulty ratings revealed that the entries in Phase 3 were indeed rated as easier to infer, $M = 2.32$, than those in Phase 1, $M = 2.01$, $F = 13.63$, $p < .01$, consistent with the reviewer's speculation. It is possible, of course, that our attempt to get the raters to use only the immediate context as the basis for these ratings was not entirely successful, and that their rating of the Phase 3 entries as easier to infer also reflected their own greater familiarity with the targets at this stage of the therapy session. If this were the case, however, there is even more reason for believing that if the empathic accuracy data were reanalyzed in a manner that "corrected" for these differences in perceived difficulty, the significant main effect for phase would be substantially attenuated. Moreover, this reanalysis also could be used to address the concern of a different reviewer, who wondered what the results would look like if it were possible to control for instances in which it was easy to infer the targets' thoughts and feelings because they had just expressed them a moment or two before to the therapist.

In this proposed reanalysis of the empathic accuracy data, the difficulty ratings could not be used as a conventional covariate, because they did not vary from one participant (i.e., perceiver)

Table 5
Correlations of Actual Empathic Accuracy and Subjective Perception of Empathic Accuracy for Each Videotarget

Group	Videotarget		
	Divorce 1	Divorce 2	Role conflict
All	.08	.00	-.18
Women	.07	.04	-.10
Men	.08	-.05	-.22
Feedback	.21	.04	-.19
No feedback	.06	.10	-.12

Note. None of the correlations in this table was statistically significant. To attain a significance level of .05 with $df = 72$, a correlation of .23 would be required.

to the next but were instead constant attributes of the stimulus materials themselves. Accordingly, we used the difficulty ratings to derive a set of constant weights that could be used to adjust each participant's empathic accuracy scores at each level of Target \times Phase (i.e., by weighting the accuracy scores for easy-to-infer entries less heavily than the accuracy scores for difficult-to-infer entries). We then conducted the same ANOVA used in the original analysis on these adjusted scores to address the questions raised by the two reviewers.

Consistent with the first reviewer's expectation, the main effect for phase was indeed substantially attenuated in this reanalysis, with the F value dropping from 101.91 to 66.16. It is important to note, however, that the revised F value is still highly significant ($p < .0001$), suggesting that the Phase 3 entries were easier to infer not only because more cues were available in the immediate context but also because the perceivers had "gotten to know" the clients through the more expanded context of meaning that had developed over the entire course of the clients' preceding interaction with the therapist.

With regard to the second reviewer's interest in what other effects from the original ANOVA might be altered in a reanalysis of this type, the data revealed relatively little change. The main effect for feedback, $F = 4.52$, was virtually the same as in the original analysis, $F = 4.49$, as was the main effect for target, $F_s = 59.59$ versus 60.11. The Phase \times Feedback interaction was slightly attenuated, $F_s = 3.47$ vs. 4.32, but the Target \times Phase interaction was somewhat stronger, $F_s = 35.32$ vs. 26.85. The only apparent surprise in these comparisons is that the target effect, unlike the phase effect, was not substantially attenuated by adjusting the scores for inferential difficulty. This lack of attenuation is most likely an artifact of the rating procedure itself, in that this procedure undoubtedly resulted in the raters being much more sensitive to differences in the difficulty of inference of entries within each of the tapes than across the tapes. However, it also attests to the strength of the difference in how difficult it was to "get to know" each target based on the overall coherence of the account she gave to the therapist: The overall lack of coherence in the Divorce 2 account apparently made this target particularly difficult to understand even when we attempted to control for the differential weighting of immediate contextual cues.

Discussion

This study addressed three questions regarding empathic accuracy in a clinically relevant setting. First, does the empathic accuracy of a perceiver improve with increased exposure to a target individual? Second, can empathic accuracy be enhanced by providing the perceiver with objective feedback about the target's actual thoughts and feelings? Third, are there stable individual differences in empathic accuracy that generalize across different targets?

The results indicated that although absolute performance levels varied across individual targets, empathic accuracy generally improved with increased exposure to the target. In addition, feedback about the target's actual thoughts and feelings accelerated the rate at which the perceivers' empathic accuracy improved. Finally, cross-target consistency in responding revealed stable and highly significant individual differences in the

perceivers' empathic ability. We discuss each of these findings below and consider their implications for clinical training and practice.

Target Effects

The large main effect for target indicated that the thoughts and feelings of two of the targets (Divorce 1 and Role Conflict) were generally inferred with greater accuracy than those of the remaining target (Divorce 2). Although the present study was not specifically designed to explore the basis of such target differences, our findings clearly indicate that closer attention should be paid to this issue in future research. In this study, target differences were important not only because of their size but also because they moderated a number of other effects in the data (e.g., the variable absolute performance levels depicted in Figure 2 and the failure to obtain an overall phase effect for the target in the Divorce 2 tape).

Practice Effects and Target Familiarity

The large main effect for phase indicated that empathic accuracy improved as a function of increased familiarity with the target individuals. This finding is consistent with those of previous studies that have also reported improvements in empathic accuracy as a function of target observation time (Bernstein & Davis, 1982; Neimeyer et al., 1983). Collectively, these findings suggest that empathic accuracy in a clinical setting may require time to develop, regardless of initial differences in the ability levels of the perceivers. These same findings also highlight the methodological importance in empathy assessment studies of providing the perceiver with a reasonable opportunity to come to "know" the target individual (cf. Levenson & Ruef, 1992).

Feedback Effects

Among the most interesting findings in this study were the significant feedback main effect and the significant Phase \times Feedback interaction. The main effect revealed that perceivers who received feedback about the targets' actual thoughts and feelings during Phase 2 of each tape were generally more accurate in inferring the targets' thoughts and feelings in the baseline and test phases (Phases 1 and 3) than were perceivers who received no feedback during Phase 2. The interaction revealed more specifically that the provision of feedback increased the rate at which the perceiver's empathic accuracy improved over time (i.e., from Phase 1 to Phase 3), a finding that is impressive because as few as 42 training (i.e., target-generated feedback) trials were sufficient to produce this effect. To the best of our knowledge, these findings are the first of their type to be reported.

The individual target analyses further revealed that the Phase \times Feedback interaction occurred on the first tape (Divorce 1) that participants viewed. Specifically, the feedback and no-feedback groups did not differ in their performance during the initial, baseline phase of the Divorce 1 tape, but they did differ in the predicted direction during the final test phase that followed the feedback manipulation. These data indicate that as few as

14 training trials were sufficient to enhance Phase 3 empathic accuracy on the Divorce 1 tape.

The efficacy of the feedback training was not attributable to differential levels of task motivation and task interest on the part of feedback versus control-group participants. Moreover, the combined feedback training effect was a general one that was not further qualified by a higher order (Phase \times Feedback \times Target) interaction. The overall feedback effect (for the combined Phase 1 and Phase 3 data) was significant even for the target in the Divorce 2 tape, who was particularly difficult to "read" because of her highly ambivalent and unresolved feelings about the desirability of seeking a divorce.

The finding that feedback training effectively enhanced empathic accuracy suggests that this procedure may constitute a promising method for the training of empathic understanding skills. Of particular interest are trends in the data suggesting that, to some extent, the effects of the feedback training generalized to subsequent targets. Recall that on the first target tape, the feedback versus control-group baseline scores were not significantly different. Indeed, the mean for the control group was actually slightly higher than the mean for the feedback group. In contrast, the data for the subsequent tapes revealed that, compared with control group participants, participants who received prior feedback training obtained higher mean baseline as well as test phase scores. Although these between-group differences on the subsequent tapes did not always reach significance, the baseline scores of feedback participants were consistently higher than those of the no-feedback controls. More important, in the original analysis (which did not pool the data for different tapes), the combined baseline-test-phase scores of feedback participants were significantly higher than those of their no-feedback counterparts for each of the three tapes (Divorce 1, Divorce 2, and Role Conflict). Hence, the data suggest that feedback training not only enhanced empathic accuracy in the test phase for each respective training target but also enhanced overall empathic accuracy (pooled across baseline and test phases) for subsequent training targets.

These data support the belief that the capacity for empathic understanding is a trainable skill, and that through the provision of immediate, target-generated feedback a "generalized" or global improvement in this skill may be obtainable. Although caveats about the external validity of analogue studies are always in order, the present findings might eventually prove to have important implications for graduate training programs in clinical psychology. They suggest that, under the specific training conditions established in this study, trainees can accelerate their learning of this particular component of the empathy process (i.e., accurate empathic understanding). On the other hand, the current study did not attempt to address the intriguing issue of whether trainees may also develop more accurate empathic inference skills under typical graduate training conditions (i.e., following the provision of delayed, clinician-generated feedback).

Despite the promising results obtained with feedback training in the current experiment, several more specific caveats are in order. First, the efficacy of this training approach must be replicated in future research, using targets who vary in their sex, race, age, socioeconomic status, topic content, and so forth. Second, the present results do not guarantee that the rapid man-

ifestation of feedback effects observed with regard to the Divorce 1 tape in the present study (e.g., after 14 training trials) would emerge with the use of different stimulus tapes. Although a dramatic increase in empathic accuracy was observed following feedback administration on the Divorce 1 tape, absolute performance levels on the subsequent target tapes were somewhat less compelling. The impact of target characteristics on empathic accuracy is obviously one important factor to consider in this regard. Third, although the feedback training in the current study resulted in a statistically significant increase in empathic accuracy, whether such an increase would be of clinical significance in an applied context remains open to question.

Fourth, a remaining question is the precise function that the feedback administration serves. Recall that the ratings of empathic accuracy were based on a similarity, or "matching," criterion. To the degree that the participants' inferences were congruent with the written thought-feeling entries provided by the videotargets, the raters judged them to be more accurate. But did the feedback actually teach the participants to become more accurate in their subsequent empathic inferences, as has been suggested, or did it merely train them to mimic the "language" (i.e., the vocabulary and characteristic expressions) used by the target individuals?⁸

This alternative interpretation seems unlikely, in that all of the participants—those in both the feedback and the no-feedback conditions—had the opportunity to learn to mimic the "language" of the target individuals from observing their extensive verbal interactions with the therapist. For this alternative interpretation to be valid, then, one would have to assume that the 14 thought-feeling statements that appeared on the monitor for the feedback participants provided substantial new knowledge about the "language" of the target individuals that was unavailable in the approximately 30-min sample of each target's verbal interaction with the therapist.

Although we find such an assumption implausible, the alternative interpretation it suggests may nonetheless deserve some attention in future research. One approach to testing this possibility would be to instruct the videotargets to systematically change the topic of discussion, such that any overlap in content and verbalizations from the early to the late portions of the tape would be minimized. Research of this type could more precisely delineate the function served by feedback administration.

Dispositional Empathic Accuracy

In general, there was a striking degree of cross-target stability in the levels of empathic accuracy displayed by the participants in this study. The cross-target reliability coefficient of .86 indicated that stable, individual differences in empathic accuracy were reliably measured by the videotape assessment procedure we used. This finding assumes particular importance in light of the fact that the 3 targets differed substantially according to the ease with which perceivers could accurately infer their thoughts

⁸ This effect—teaching the therapist to more readily use the client's own "vocabulary or experience"—could still have major therapeutic value (i.e., in increasing the client's feeling of being understood by the therapist).

and feelings. Despite the large target main effect that emerged in this study, however, the participants' level of performance was highly consistent across targets. This is encouragingly strong evidence that perceivers differ reliably in their empathic ability—evidence of the type that Funder (1994) called for to support the notion of the “good judge.”

These data suggest that it may be profitable to use the present methodology for purposes of scale development, in much the same fashion that the Minnesota Multiphasic Personality Inventory (MMPI; Hathaway & McKinley, 1967) was developed through the criterion-keying method of scale construction. In other words, our measure of empathic accuracy could function as both a predictor of dispositional empathic accuracy per se and as a criterion measure for developing a more predictively valid paper-and-pencil measure of this construct. Although a predictively valid self-report measure of empathic accuracy has remained elusive, the present method may assist in the identification of individuals who are high versus low in dispositional empathic accuracy. This work would have obvious application in the selection of applicants for graduate study in clinical and counseling psychology.

Finally, it is noteworthy that the feedback manipulation reduced the range of scores on the criterion measure for the experimental-group participants and thus attenuated the cross-target correlations. The implication—that the significant feedback effect occurred by disrupting (i.e., enhancing) the perceivers' characteristic levels of empathic accuracy—clearly points to the contributions of both dispositional and situational factors in the empathy process.

Subjective Perceptions of Empathic Accuracy

In general, perceptions of accuracy on the experimental task were found to be unrelated to actual empathic accuracy scores. The inability of participants to provide reliable self-reports of their own empathic ability is particularly striking given that they were asked to make specific postexperimental estimates about their empathic accuracy with respect to each of the videotargets. These data are consistent with those reported by previous investigators (e.g., Ickes, Stinson, et al., 1990), who have also found that participants are poor and unreliable judges of their own empathic ability. Moreover, our data revealed no association between perceptions of accuracy and actual empathic accuracy as a function of feedback condition (see Table 5).

It is interesting that the provision of information regarding the relative accuracy or inaccuracy of empathic inferences, although sufficient to enhance *actual* accuracy, was insufficient to alter *perceptions* of accuracy among feedback participants. Although most people can provide good estimates of their performance on IQ tests or in the bowling alley, they appear to have relatively poor insight into their ability to accurately infer the thoughts and feelings of others. Ickes (1993) proposed that this lack of metaknowledge may be largely responsible for the failure of researchers during the last half-century to develop a predictively valid test of self-rated empathic accuracy (for reviews, see Chlopan, McCain, Carbonell, & Hagen, 1985; Funder, 1994; Funder & Harris, 1986; Goldstein & Michaels, 1985; Kenny & Albright, 1987). As Funder (1994, p. 23) said, “While a good judge is likely to know that he or she is good, there is little stop-

ping a poor judge from also believing himself or herself to be a good judge. His or her very lack of insight may create an assessment conundrum, making it difficult to detect this lack through self-report (Fletcher, 1993; Ickes, 1993).”

Conclusion

In summary, the present findings indicate that although empathic accuracy may indeed be a trainable skill, it is a skill whose level is often unclear to the one who possesses it. The dynamic nature of the empathy process mandates that both sender and perceiver characteristics influence the degree to which training enhances empathic understanding. Both the ability of the perceiver and the “readability” of the target must be taken into account.

The findings also indicate that the present methodology represents a promising approach for the assessment and training of accurate empathic understanding. Its advantages relative to more commonly used approaches include (a) a reliable, objective criterion for the assessment of empathic accuracy; (b) time and cost-effectiveness (i.e., the effects of training emerge rapidly and independently of trained professionals); and (c) its ability to capture the spirit of the person-centered tradition by requiring the observer to generate inferences rather than choose from a set of prefabricated responses, and providing the opportunity to train and assess empathy as an ongoing process, as recommended by Rogers (1975).

If future research confirms that the present method is a viable research and training technique, there may be several applications for this methodology. For example, future studies could attempt to relate our measure of empathic accuracy with the physiological linkage between perceiver and target that was the focus of Levenson and Ruef's (1992) promising research. In terms of practical application, a “library” of stimulus tapes such as those used in the present study could be developed to serve as a programmed, self-instructional adjunct to more traditional training methods. In addition, the present methodology could be used to assess improvement in empathic understanding following long-term interventions specifically designed to enhance this skill.

Perhaps the most compelling implication of the results of the present study concern the serendipitous finding of cross-target consistency in responding. The cross-target reliability coefficient of .86 suggests that individual differences in empathic ability can be assessed reliably with even a limited number of target tapes. Thus, the present procedure may also constitute a psychometrically viable tool for the selection or recruitment of individuals high versus low in dispositional empathic accuracy.

In summary, the results suggest that the empathy training and assessment technique used in the present study represents a promising avenue for future research and application, enabling (a) further exploration of the benefits of feedback training; (b) further delineation of perceiver, target, and relationship characteristics that affect the perceiver's empathic accuracy; and (c) the development of a predictively valid paper-and-pencil measure of empathic ability to complement the reliable but time-intensive performance measure used in the present investigation.

References

- Asch, S. E. (1946). Forming impressions of personality. *Journal of Abnormal and Social Psychology*, 41, 258-290.
- Barrett-Lennard, G. T. (1962). Dimensions of therapist response as causal factors in therapeutic change. *Psychological Monographs*, 76, 1-33.
- Barrett-Lennard, G. T. (1981). The empathy cycle: Refinement of a nuclear concept. *Journal of Counseling Psychology*, 28, 91-100.
- Bernstein, W. M., & Davis, M. H. (1982). Perspective-taking, self-consciousness, and accuracy in person perception. *Basic and Applied Social Psychology*, 3, 1-19.
- Borgida, E. (1979). Character proof and the fireside induction. *Law and Human Behavior*, 3, 189-202.
- Borgida, E. (1980). Evidentiary reform of rape laws: A psycholegal approach. In P. D. Lipsitt and B. D. Sales (Eds.), *New directions in psychological research* (pp. 171-197). New York: Van Nostrand Reinhold.
- Butler, J. M., & Haigh, G. V. (1954). Changes in the relationship between self-concepts and ideal concepts consequent upon client-centered counseling. In C. R. Rogers & R. F. Dymond (Eds.), *Psychotherapy and personality change* (pp. 55-75). Chicago: University of Chicago Press.
- Carkhuff, R. R. (1969a). *Helping and human relations: A primer for lay and professional helpers: Vol. 1. Selection and training*. New York: Holt, Rinehart & Winston.
- Carkhuff, R. R. (1969b). *Helping and human relations: A primer for lay and professional helpers: Vol. 2. Practice and research*. New York: Holt, Rinehart & Winston.
- Carkhuff, R. R. & Anthony, W. A. (1979). *The skills of helping: An introduction to counseling skills*. Amherst, MA: Human Resources Development Press.
- Carkhuff, R. R., & Burstein, J. (1970). Objective therapist and client ratings of therapist-offered facilitative conditions of moderate to low functioning therapists. *Journal of Clinical Psychology*, 26, 394-395.
- Chlopan, B. E., McCain, M. L., Carbonell, J. L., & Hagen, R. L. (1985). Empathy: Review of available measures. *Journal of Personality and Social Psychology*, 48, 635-653.
- Costanzo, M., & Archer, D. (1989). Interpreting the expressive behavior of others: The Interpersonal Perception Task. *Journal of Nonverbal Behavior*, 13, 225-245.
- Cronbach, L. J. (1955). Processes affecting scores on "understanding of others" and "assumed similarity." *Psychological Bulletin*, 52, 177-193.
- Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 24, 349-354.
- Danish, S., & Kagan, N. (1971). Measurement of affective sensitivity: Toward a valid measure of interpersonal perception. *Journal of Counseling Psychology*, 18, 51-54.
- Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology*, 10, 85.
- Davis, M. H. (1994). *Empathy: A social psychological approach*. Madison, WI: Brown & Benchmark.
- Deutsch, F., & Madle, R. A. (1975). Empathy: Historic and current conceptualizations, measurement, and a cognitive theoretical perspective. *Human Development*, 18, 267-287.
- Egan, G. (1976). *Interpersonal living: A skills/contract approach to human-relations training in groups*. Monterey, CA: Brooks/Cole.
- Eisenberg-Berg, N., & Mussen, P. (1978). Empathy and moral development in adolescence. *Developmental Psychology*, 14, 185-186.
- Ekman, P., & Friesen, W. V. (1975). *Unmasking the face: A guide to recognizing emotions from facial cues*. Englewood Cliffs, NJ: Prentice-Hall.
- Elliot, R., Filipovich, H., Harrigan, L., Gaynor, J., Reimschuessel, C., & Zapadka, J. K. (1982). Measuring response empathy: The development of a multicomponent rating scale. *Journal of Counseling Psychology*, 29, 379-387.
- Estes, S. G. (1938). Judging personality from expressive behavior. *Journal of Abnormal and Social Psychology*, 33, 217-236.
- Feldstein, J. C., & Gladstein, G. A. (1980). A comparison of the construct validities of four measures of empathy. *Measurement and Evaluation in Guidance*, 13, 49-57.
- Fletcher, G. (1993, October). *Empathic accuracy in married couples*. Paper presented at the annual meeting of the Society for Experimental Social Psychology, Santa Barbara, CA.
- Funder, D. C. (1994). *Accuracy theory*. Manuscript submitted for publication.
- Funder, D. C., & Colvin, C. R. (1988). Friends and strangers: Acquaintanceship, agreement, and the accuracy of personality judgment. *Journal of Personality and Social Psychology*, 55, 149-158.
- 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.
- Goldstein, A. P., & Michaels, G. Y. (1985). *Empathy: Development, training and consequences*. Hillsdale, NJ: Erlbaum.
- Graham, T. (1994). *Empathic accuracy in dyads composed of friends or strangers of the same sex*. Unpublished master's thesis, University of Texas at Arlington.
- Hall, J. A. (1978). Gender effects in decoding nonverbal cues. *Psychological Bulletin*, 85, 845-857.
- Hall, J. A. (1987). On explaining gender differences: The case of nonverbal communication. In P. Shaver & C. Hendrick (Eds.), *Review of personality and social psychology: Vol. 7. Sex and gender* (pp. 177-200). Newbury Park, CA: Sage.
- Hancock, M., & Ickes, W. (1995). *Empathic accuracy: When does the perceiver-target relationship make a difference?* Manuscript submitted for publication.
- Hathaway, S. R., & McKinley, J. C. (1967). *Minnesota Multiphasic Personality Inventory: Manual for administration and scoring*. New York: Psychological Corporation.
- Hoffman, M. L. (1977). Empathy, its development and prosocial implications. *Nebraska Symposium on Motivation* (Vol. 25, pp. 169-218). Lincoln: University of Nebraska Press.
- Hogan, R. (1969). Development of an empathy scale. *Journal of Consulting and Clinical Psychology*, 33, 307-316.
- Ickes, W. (1988). *An 18-item measure of empathic accuracy*. Unpublished scale and associated psychometric data, University of Texas at Arlington.
- Ickes, W. (1993). Empathic accuracy. *Journal of Personality*, 61, 587-609.
- Ickes, W., Bissonnette, V., Garcia, S., & Stinson, L. (1990). Implementing and using the dyadic interaction paradigm. In C. Hendrick & M. S. 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., Stinson, L., Bissonnette, V., & Garcia, S. (1990). Naturalistic social cognition: Empathic accuracy in mixed-sex dyads. *Journal of Personality and Social Psychology*, 59, 730-742.
- Ickes, W., & Tooke, W. (1988). The observational method: Studying the interactions of minds and bodies. In S. Duck, D. F. Hay, S. E. Hobfoll, W. Ickes, & B. Montgomery (Eds.), *Handbook of personal relationships: Theory, research, and interventions* (pp. 79-97). Chichester, England: Wiley.
- Kagan, N. (1972). *Influencing human interaction*. East Lansing, MI: Michigan State University Press.
- Kagan, N. (1977a). *Affective sensitivity test: Validity and reliability*. Pa-

- per presented at the 85th Annual Convention of the American Psychological Association, San Francisco, CA.
- Kagan, N. (1977b). *Interpersonal process recall*. East Lansing, MI: Michigan State University Press.
- Kagan, N., Krathwohl, D. R., & Farquhar, W. W. (1965). *Interpersonal process recall: Stimulated recall by videotape in exploratory studies of counseling and teaching-learning—Final report (National Defense Education Act Title VII, Grant No. OE 7-32-0410-216)*. East Lansing, MI: Michigan State University Press.
- Kagan, N., Krathwohl, D. R., Goldberg, D., Campbell, R., Schauble, P., Greenberg, P., Danish, B., Resnikoff, A., Bowes, J., & Bondy, S. (1967). *Studies in human interaction: Interpersonal process recall stimulated by videotape*. East Lansing, MI: Michigan State University Press.
- Keefe, T. (1976). Empathy: The critical skill. *Social Work, 21*, 10–14.
- Keefe, T. (1979). The development of empathic skill: A study. *Journal of Education for Social Work, 15*, 30–37.
- 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.
- Knudson, R. A., Sommers, A. A., & Golding, S. L. (1980). Interpersonal perception and mode of resolution in marital conflict. *Journal of Personality and Social Psychology, 38*, 251–263.
- Kurtz, R. R., & Grummon, D. L. (1972). Different approaches to the measurement of therapist empathy and their relationship to therapy outcomes. *Journal of Consulting and Clinical Psychology, 39*, 106–115.
- Laing, R. D., Phillipson, H., & Lee, A. R. (1966). *Interpersonal perception: A theory and a method of research*. New York: Springer.
- Lamber, M. J., DeJulio, S. S., & Stein, D. M. (1978). Therapist interpersonal skills: Process, outcome, methodological considerations, and recommendations for future research. *Psychological Bulletin, 85*, 467–489.
- Levenson, R. W., & Ruef, A. M. (1992). Empathy: A physiological substrate. *Journal of Personality and Social Psychology, 63*, 234–246.
- Marks, S. E., & Tolsma, R. J. (1986). Empathy research: Some methodological considerations. *Psychotherapy, 23*, 4–20.
- McCrae, R. R. (1982). Consensual validation of personality traits: Evidence from self-reports and ratings. *Journal of Personality and Social Psychology, 43*, 293–303.
- Mehrabian, A., & Epstein, N. (1972). A measure of emotional empathy. *Journal of Personality, 40*, 525–543.
- Mill, J. (1984). High and low self-monitoring individuals: Their decoding skills and empathic expression. *Journal of Personality, 52*, 372–388.
- Neimeyer, R. A., Neimeyer, G. L., & Landfield, A. W. (1983). Conceptual differentiation, integration, and empathic prediction. *Journal of Personality, 51*, 185.
- Newmark, C. S., Woody, G., & Ziff, D. (1977). Understanding and similarity in relation to marital satisfaction. *Journal of Clinical Psychology, 33*, 83–86.
- Noller, P. (1980). Misunderstandings in marital communication: A study of couples' nonverbal communication. *Journal of Personality and Social Psychology, 39*, 1135–1148.
- Noller, P. (1981). Gender and marital adjustment level differences in decoding messages from spouses and strangers. *Journal of Personality and Social Psychology, 41*, 272–278.
- Noller, P., & Venardos, C. (1986). Communication awareness in married couples. *Journal of Social and Personal Relationships, 3*, 31–42.
- Norman, W. T., & Goldberg, L. R. (1966). Raters, ratees, and randomness in personality. *Journal of Personality, 4*, 681–691.
- 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.
- Patterson, C. H. (1984). Empathy, warmth, and genuineness in psychotherapy: A review of reviews. *Psychotherapy, 21*, 431–438.
- Rogers, C. R. (1954). The case of Mrs. Oak: A research analysis. In C. R. Rogers & R. F. Dymond (Eds.), *Psychotherapy and personality change* (pp. 259–348). Chicago: University of Chicago Press.
- Rogers, C. R. (1957). The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting Psychology, 21*, 95–103.
- Rogers, C. R. (1975). Empathic: An unappreciated way of being. *The Counseling Psychologist, 5*, 2–10.
- Rogers, C. R., & Dymond, R. F. (Eds.). (1954). *Psychotherapy and personality change*. Chicago: University of Chicago Press.
- Rogers, C. R., & Meador, B. (1977). In R. Corsini (Ed.), *Current psychotherapies*. Itasca, IL: Peacock.
- 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.
- Selman, R. L. (1980). *The growth of interpersonal understanding*. New York: Academic Press.
- Sillars, A. L. (1989). Communication, uncertainty and understanding in marriage. In B. Dervin, L. Grossberg, B. O'Keefe, & E. Wartella (Eds.), *Rethinking communication: Vol. 2. Paradigm exemplars* (pp. 307–328). Newbury Park, CA: Sage.
- Sillars, A. L., & Scott, M. D. (1983). Interpersonal perception between intimates: An integrative review. *Human Communication Research, 10*, 153–176.
- Simpson, J., Ickes, W., & Blackstone, T. (1994). *When the head protects the heart: Empathic accuracy in dating relationships*. Manuscript submitted for publication.
- Stinson, L., & Ickes, W. (1992). Empathic accuracy in the interactions of male friends versus male strangers. *Journal of Personality and Social Psychology, 62*, 787–797.
- Truax, C. B., & Carkhuff, R. R. (1967). *Toward effective counseling and psychotherapy: Training and practice*. Chicago: Aldine.
- Wilson, F. R., & Griswold, M. L. (1985). The effects of method and comprehensiveness of training on the reliability and validity of ratings of counselor empathy. *Measurement and Evaluation in Counseling and Development, 18*, 3–11.

Received June 7, 1994

Accepted October 11, 1994 ■