The Development of Meaning Contexts for Empathic Accuracy: Channel and Sequence Effects

Paul R. Gesn and William Ickes
University of Texas at Arlington

This study examined the effects of both immediate and cumulative contextual knowledge on empathic accuracy. Participants viewed excerpts from videotapes of 3 simulated psychotherapy sessions and attempted to infer each client's actual thoughts and feelings. Immediate contextual knowledge was varied by controlling the informational channels available. Cumulative contextual knowledge was varied by presenting the excerpts from each tape in either their original order or in a random order. The results revealed that empathic accuracy in this clinically relevant context was primarily dependent upon verbal, rather than nonverbal cues. Knowledge of the cumulative meaning context was also important, however, in that it led perceivers to respond in a schema-driven fashion that sometimes facilitated, but at other times impaired, their empathic accuracy.

Empathic accuracy is a measure of a perceiver's ability to accurately infer the specific content of another person's thoughts and feelings (Ickes, 1993). To date, none of the studies that used Ickes, Stinson, Bissonnette, and Garcia's (1990) procedure for assessing empathic accuracy has focused on the types of contextual cues that might influence empathic inference. Moreover, only two studies have examined the verbal and nonverbal cues that perceivers might use in making such inferences (Ickes et al., 1990; Stinson & Ickes, 1992). Accordingly, in the present study we examined the roles of both the immediate contextual cues (the verbal and nonverbal behaviors that precede each reported thought or feeling) and the cumulative contextual cues (the emergent meaning context that develops as a function of the temporal nature of an interaction) in the on-line inference of the specific content of a target person's successive thoughts and feelings.

The essential questions we are posing in the present study are (a) What is the nature of the context that promotes the greatest accuracy in a perceiver's inferences about a target person's specific thoughts and feelings? and (b) Will the emergent cumulative meaning context facilitate such accuracy beyond the effect of the immediate meaning context provided by the words and actions that immediately precede each of the target person's reported thoughts and feelings?

Cumulative Contextual Knowledge

In a social interaction context, one acquires cumulative contextual knowledge about a target by tracking the person's behavior as it unfolds over the course of the interaction, such that accurate perception at time \( x \) is dependent upon and emerges from the perception of information available at time \( x - 1 \) or other previous times (Ginsburg & Smith, 1993). We propose that cumulative contextual knowledge is important because it can make a unique contribution to the perceiver's accuracy in "reading" the target person's thoughts and feelings, that is, one that goes beyond the contribution made by the target's immediate verbal and nonverbal cues. In some instances, of course, cumulative contextual knowledge will merely complement and reinforce the knowledge derived from the target's immediate behavior. In other instances, however, cumulative contextual knowledge will add new and independent information that can dramatically qualify the meaning of the more immediate contextual knowledge—and, in some cases, even contradict it entirely. For example, only Jane's friend Barbara, who has known her for years, knows that Jane's 100-watt smile during the surprise birthday party that her new coworkers have thrown for her masks Jane's long-term antipathy to being made the center of attention in social situations (Colvin, Vogt, & Ickes, 1997, p. 182).

Our approach to studying cumulative contextual knowledge in the present study was inspired in part by Newtson's (1973) research on behavior perception. Newtson asked participants to view videotaped sequences depicting an actor performing a concrete behavior and to indicate the naturally occurring, meaningful acts of the sequence by pressing a button on a continuous-event recorder as they viewed the tapes. The boundaries of action units indicated by the participants were referred to as breakpoints. Nombreakpoints were points in the sequence within any particular action unit; that is, they represented points of continuity. Interestingly, participants who were presented with slides of the series of still breakpoints identified by the previous participants (i.e., a series resembling the still images of a story board) were able to extract as much meaningful information as did the participants who had viewed the complete dynamic sequence of behavior on videotape.
Two robust findings that Newtson and his colleagues have reported are of particular relevance to the present study. First, the order of presentation of the breakpoint slides reliably affected the participants' ability to accurately interpret the behavioral sequences (Newton, Engquist, & Bois, 1977). The accurate interpretation of the complete behavioral sequences from the slides of the breakpoints alone occurred only when the slides were presented in the correct order. In contrast, comprehension of the action was significantly impaired when the slides were presented out of order.

Second, the degree of order information contained in triads of breakpoints was reliably greater than the information contained in the component pairs. Newton et al. (1977) presented half of the participants in their study with triads of slides from the behavioral sequences. They presented the remaining participants with the same three slides but in pairs; that is, for each triad, the three component pairs of slides were shown to the participants. For this condition, Newton et al. determined the joint probability of accuracy of judgment for each triad. They found breakpoint triads to be more accurately interpreted and judged as to correctness of presentation order than could be predicted from the triads' component pairs (Newton & Engquist, 1976). According to Newton et al. (1977), this finding implies that "breakpoints are selected not only on the basis of immediate changes in the ongoing behavior stream, but also with respect to a coherent, ongoing interpretation of the action" (p. 857).

Newton et al. (1977) concluded, however, that isolated behavioral segments are not sufficient to provide an accurate interpretation of ongoing behavior. They argued that a necessary condition for perceptual accuracy is that it occur within the "frame" of a coherent ongoing interpretation of the temporally extended action sequence. This view assumes that there is an orderly and coherent flow to behavioral sequences that can be apprehended in terms of the emerging gestalten or patterns described by Asch (1952). If this view is correct, it implies that immediate contextual knowledge (i.e., that derived from the words and actions that immediately precede a target's reported thought or feeling) may not be sufficient for the perceiver to achieve a high degree of empathic accuracy. Such accuracy should also depend in part on the cumulative contextual knowledge that is constructed by the perceiver from the ongoing flow of events.

If empathic accuracy depends in part on the cumulative contextual knowledge that the perceiver constructs from the entire flow of an interaction sequence, then participants presented with excerpts of an interaction in the order in which they actually occurred should have no advantage compared with participants who are presented the same excerpts in a randomly shuffled order. This outcome, which would uphold the sufficiency of immediate contextual knowledge while calling into question the unique contribution of cumulative contextual knowledge, is not one that can be readily dismissed as implausible. Although the importance of cumulative contextual knowledge may seem obvious in a task such as tracking the unfolding plot of a story, it is possible that inferring a target's thoughts and feelings during a social interaction is not necessarily subject to the same cross-temporal constraints. The experience of speaking with an interaction partner who jumps back and forth from topic to topic helps to make this point clear: Will the perceiver's empathic accuracy necessarily suffer in this case?

Immediate Contextual Knowledge

In addition to testing for a unique influence of cumulative contextual knowledge on perceivers' empathic accuracy in the present study, we also tested the effect of constraining the different (verbal and nonverbal) information channels available in the brief time windows immediately preceding each thought or feeling (i.e., constraining the channels through which the immediate contextual knowledge can be obtained).

Much of the research on "channel effects" on human communication has reflected Darwin's (1872) cross-species emphasis on the accurate encoding and decoding of nonverbal cues (Archer & Akert, 1984; Funder, 1995; Funder, Sneed, & Carl, 1993; Zebrowski, 1990, 1997). An important goal of this research has been to compare the effects of information provided in different channels of communication on perceivers' perceptual or inferential accuracy—ideally, by presenting the information in each channel in a relatively "pure" form that isolates it from the information provided by other channels. For example, the Profile of Nonverbal Sensitivity, developed by Rosenthal, Hall, DiMatteo, Rogers, and Archer (1979), was designed with this goal in mind.

In the present study, we sought to determine which information channels contribute most to perceivers' empathic accuracy when they infer the thoughts and feelings that different female clients reported in their interactions with a male client-centered therapist. Our general prediction was that, in this clinically relevant setting, the verbal channel would carry most of the information on which accurate inferences are based. Because clients in therapy settings are typically motivated to express many (though not all) of their thoughts and feelings in words, and because their body movements are typically constrained by the chair in which they are sitting, the therapy context is one in which verbal cues should exert the most influence on perceivers' empathic accuracy scores.

The importance of also taking into account the effect of context in this situational sense of the term has been emphasized by DePaulo and Friedman (1998), who noted that
Attempts are sometimes made to ascertain the relative importance to social meaning of verbal versus nonverbal cues (e.g., Mehrabian, 1972). But these are ill-fated because the importance is context dependent (Friedman, 1978). In a letter or other written correspondence, the meaning is by definition all verbal, whereas in a silent, intimate encounter, most if not all of the meaning may be nonverbally created. (p. 27)

This injunction—that the importance of different information channels can depend entirely on the situational context (see also Munn, 1940)—provides an essential corrective to an earlier theme in the work on nonverbal decoding that argued for the "primacy" of visual cues over those in other modalities. Although the results of a number of studies seemed to support the greater importance or primacy of visual cues (e.g., DePaulo, Rosenthal, Eisenstat, Rogers, & Finkelstein, 1978; Gitter, Kozel, & Mostofsky, 1972; Mehrabian & Ferris, 1967; Jonides & Yantis, 1988), researchers have generally abandoned that position in favor of the context-dependent position espoused by Friedman (1978), Munn (1940), and DePaulo and Friedman (1998).

In the present study, we contrasted three different channel conditions: video and audio (VA), video and filtered audio (VFA), and audio only (AO). We considered, but ultimately decided against, including video only, filtered audio only, or verbal transcript only conditions as well. Our pilot research convinced us that few, if any, participants would be willing to complete 2-hr sessions in which they would be asked to infer the thoughts and feelings of three clients from silent videotapes or from audiotapes in which the client–therapist conversations were completely unintelligible because of electronic filtering. Even if a few participants were willing to complete the experimental sessions in these two conditions, they would likely have been so highly self-selected (on dimensions such as compliance, tolerance for frustration, etc.) that such individual-difference characteristics would inevitably have confounded the interpretation of any significant effects obtained. The inclusion of a verbal transcript would have resulted in a different, but no less serious, confound. Because it would have introduced a communication channel (reading comprehension) that was not present in the original client–therapist interactions, the interpretation of any significant effects involving this condition would be compromised by the fact that the decoding skill required in this condition was not one required in the original client–therapist interactions.

Our predicted main effect for the channel manipulation—that verbal information will have the greatest impact on empathic accuracy—can be further specified in terms of the three theoretical positions proposed by Archer and Akert (1980) regarding the encoding of meaning in social interactions. In their significant clue theory, the one most consistent with our prediction, Archer and Akert explain accuracy as being primarily a function of specific kinds of clues that are extremely meaningful to the task at hand, whereas other clues are uninformative or even misleading. Given the assumed greater importance of verbal content in empathic accuracy as compared with other types of social inference tasks, the participants in the conditions in which the verbal content is available should have higher empathic accuracy scores than should participants in conditions in which the verbal content is not available. The prediction according to this theory, then, is that participants in the VA condition and the AO condition should achieve empathic accuracy scores that are similar to each other but that are significantly greater than those of participants in the VFA condition.

In their additive theory, Archer and Akert (1980) explain accuracy as a linear function of the amount of information available. If the empathic accuracy task is additive in nature, vis-à-vis the immediate cues attended to, then empathic accuracy scores should be highest in the VA condition and roughly equal in the AO and VFA conditions. The VA condition contains visual cues, paralinguistic (i.e., vocalic) cues, and verbal content cues. In both the AO and VFA conditions, one cue is subtracted from these three. In the AO condition the visual cues are absent, whereas the paralinguistic and verbal content cues remain. In the VFA condition the verbal content cues are absent, whereas the visual and paralinguistic cues remain. Thus, if the additive theory is correct, accuracy should be about equal for the last two conditions.

In a third alternative, called the diffusion theory, Archer and Akert (1980) propose that all cues are informative, even in isolation. In other words, each piece of information provides a unique clue to solving the puzzle. This theory suggests that if different perceivers are presented with different channels of information, each perceiver should be accurate about some aspect of the message to be decoded. The implication for the present study is the possibility that empathic accuracy scores could be roughly equivalent in all three channel conditions.

Validity of the Empathic Accuracy Measure

A number of studies conducted during the past 10 years provide converging evidence of the validity of our empathic accuracy measure. If the measure is valid, we would expect to find that friends would display more empathic accuracy than would strangers when inferring the content of each other’s thoughts and feelings. This prediction was tested in studies reported by Stinson and Ickes (1992) and Graham (1994). The results of both studies revealed that the empathic accuracy of friends was indeed reliably greater than that of strangers.

In a more applied study, Marangoni, Garcia, Ickes, and Teng (1995) measured empathic accuracy in a clinically relevant setting. In this study, perceivers viewed videotapes of female clients interacting with a male client-centered therapist and attempted to infer the content of the specific thoughts and feelings that the clients had previously reported. Marangoni et al. hypothesized that (a) perceivers’ empathic accuracy would be greater at the end of the videotapes than at the beginning, and (b) perceivers who received feedback about the actual content of the clients’ thoughts and feelings during a (middle) training phase would achieve higher empathic accuracy scores during a (final) test phase than perceivers who did not receive feedback. Both of these predictions were supported by the data.

In another study, Kelleher (1998) predicted that perceivers who had been given an accurate frame for interpreting the nature of an interaction between a confederate and a target person would be more accurate in inferring the target’s frame-relevant thoughts and feelings than would perceivers who had been given either an inaccurate frame or no frame. This prediction was also supported by the data, thus contributing to a growing body of findings that offer converging evidence for the validity of our empathic accuracy measure.
Method

Participants and Design

The participants were 72 undergraduate students at the University of Texas at Arlington who participated in the study in exchange for credit in their introductory psychology course. The experimental design was a 3 X 2 X (3) mixed factorial. Channel and sequence were independently manipulated as the between-subjects factors, and stimulus tape was manipulated as a within-subjects (repeated measures) factor. The channel factor contrasted the three modes of stimulus presentation (VA, VFA, and AO). The sequence factor contrasted sets of excerpts from the three tapes presented in either their original sequence or a random sequence. Finally, because all three stimulus tapes were presented within the same Channel X Sequence modality, stimulus tape was treated as a repeated measures factor, with the order of tape presentation counterbalanced. Twelve participants, 6 men and 6 women, were randomly assigned to each of the six experimental conditions.

Stimulus Materials

The stimulus materials consisted of a preexperimental instructional videotape and the three stimulus videotapes. The three stimulus tapes were highly edited versions of the tapes used in the previous study by Marangoni et al. (1995). Each tape depicted a female client discussing a genuine personal problem with a male client-centered therapist. In two of the tapes—Divorce 1 and Divorce 2—the female clients talked about their actual (Divorce 1) or anticipated (Divorce 2) divorce. In the third tape—Role Conflict—the female client discussed the stress she experienced in attempting to balance her responsibilities as a wife and mother with those of her career as an attorney.

Preparation of the stimulus tapes. The stimulus tapes, developed originally for the study by Marangoni et al. (1995), were prepared as follows. Three women (all White, college-educated, 24–32 years old, from middle- to upper-middle-class backgrounds) volunteered to participate individually in a simulated psychotherapy session. They 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. Each of the female clients knew beforehand that her therapy session would be videotaped for use in future (unspecified) research. Each client also signed a consent form granting her permission for the tape to be used for this purpose.

The simulated psychotherapy sessions were each videotaped without any rehearsal, and the genuineness and spontaneity of the sessions were evident in the clients’ range of emotional expression (e.g., one woman wept openly as she discussed her recent divorce). Immediately after their respective sessions with the therapist were completed, each client was debriefed and asked to sign a second consent form indicating her willingness to participate in an assessment of the specific thoughts and feelings she had had during the videotaped session. After signing the second consent form, each client was then escorted to the University of Texas at Arlington Social Interaction Lab and seated in a cubicle. The client’s task was to view the videotape of the just-completed session and, pausing the tape at the appropriate times, to make a list of all the thoughts and feelings she distinctly remembered having had during the session along with the times at which they occurred. The experimenter emphasized to the clients both the importance of their being as honest and as accurate as possible in reporting their actual thoughts and feelings and the importance of not reporting any thoughts or feelings that occurred to them for the first time while they were viewing the videotape.

From the resulting thought-feeling protocols, 30 thoughts and feelings were available for each client in edited versions of the videotaped sessions that ranged from 26 to 36 min in length. These 30 reported thoughts and feelings for each client served as the objective criteria against which the accuracy of the empathic inferences of the participants in both the Marangoni et al. (1995) study and the present study were assessed. Sample thoughts and feelings reported by the three clients were as follows: “I was thinking about when I first started to resent him” (Divorce 1); “I was thinking about how hard it would be to face him and admit it is over” (Divorce 2); “I was feeling that although my husband tries outwardly not to impose on me, he does in fact impose by actions or insinuations” (Role Conflict). (For additional details, see Marangoni et al., 1995, and Ickes, Marangoni, & Garcia, 1997).

Editing of the stimulus tapes. To present the current study’s participants with versions of the client-therapist interactions that corresponded to the particular Channel X Sequence condition to which they had been randomly assigned, it was necessary to create several edited versions of the Marangoni et al. (1995) tapes using an audiovisual mixer. Making edited versions of the original sequence tapes was a fairly straightforward task that is described below. However, making edited versions of the random sequence tapes involved an additional consideration. Specifically, to help ensure the generality of any sequence effects that might be obtained, we edited the tapes for the random sequence conditions to create two different random orderings of the taped excerpts. Therefore, we edited each stimulus tape (Divorce 1, Divorce 2, and Role Conflict) to create three channel combinations (VA, VFA, and AO) crosscut by each of three sequential orderings (original sequence, random sequence 1, and random sequence 2). Within the random sequence condition, we presented one half of the participants with the Random Sequence 1 tapes and the other half with the Random Sequence 2 tapes.

We edited the tapes such that each tape comprised a set of thirty 15-s excerpts taken from each of the Marangoni et al. (1995) tapes. In other words, instead of seeing and/or hearing the original 26–36 min tapes of each of the three therapy sessions, the participants in the present study saw and/or heard a set of thirty 15-s excerpts from each session (representing each of the 30 thought-feeling stop points that had been identified by each of the three clients). Trial and error led us to conclude that 15-s excerpts represented the optimal trade-off of experimental realism, external validity, and experimental control. On the one hand, to preserve the real-life quality of the stimulus segments, the usual 1–3-s excerpts used in most studies of emotion perception were clearly not sufficient. On the other hand, because some of the tape stops reported by the clients occurred relatively close together, any excerpt length greater than 15 s would have caused overlap between segments and potentially would have blurred the distinction between the immediate and the cumulative meaning context.

The necessary trade-off, then, required selecting an excerpt length that both preserved some degree of the cumulative context in the original sequence condition and ensured that enough information was available in the immediate context (i.e., the excerpts preceding each tape stop) in both sequence conditions so that a fair test of the relative contributions of the

1 Practical constraints cited by Marangoni et al. (1995, i.e., the lengthiness of the procedure) mandated that the number of clients in the videotapes be limited to no more than three (see also Ickes et al., 1997). Female clients were chosen on the assumption that female targets would more willingly disclose personally meaningful, intimate concerns, and that they would do so in a more expressive fashion than would male targets. Finally, the decision to use “normal” clients in simulated psychotherapy sessions rather than clinically disturbed clients in real psychotherapy sessions was based on several considerations. These included (a) the likelihood that clinically disturbed individuals would exhibit less insight and more confusion about their thoughts and feelings; (b) the likelihood that disturbed individuals would be less willing to reveal thoughts and feelings that might make them feel ashamed, deviant, unworthy, vulnerable, and so forth; and (c) the legal and ethical problems associated with requiring disturbed individuals to participate in a self-disclosure experience whose potential impact on their lives and therapeutic outcomes was, at best, unclear.
immediate versus cumulative context could be achieved. On the one hand, presenting participants with extremely short (e.g., 1-3 s) excerpts of the client-therapist interaction would have restricted the information available in the immediate context to such an extent that the magnitude of its contribution to empathic accuracy might not have received a fair test. On the other hand, using an excerpt length of more than 15 s might have unduly confounded the cumulative context with the immediate one(s) and would also have caused overlap between successive excerpts. Thus, our decision to use 15-s excerpts was to a large extent dictated by the practical and theoretical trade-offs that were involved.

For the original sequence conditions, we presented the thirty 15-s excerpts from each of the three tapes to the participants in their original order. For the random sequence conditions, we presented the thirty 15-s clips to the participants in one of the two random orders. In the VA condition, the participants had access to both the verbal and nonverbal channels for each of the three tapes. In the AO condition, the participants heard the audio portion of each tape, but we placed a black posterboard in front of the video screen to cover it completely. In the VFA condition, we ran the audio portion during playback through a filter built by the psychology department’s electronics and computer technician, which rendered the verbal content unintelligible while preserving the paralinguistic cues.

For the participants in the VFA condition, the experimenter manually adjusted the electronic filter at the start of each stimulus tape to achieve the most effective preservation of the emotional and paralinguistic characteristics of the clients’ voices while making their actual words unintelligible. Consistent with our intent, virtually all of the participants in this condition reported that they could understand almost none of the verbal content of the conversations in the stimulus tapes. In fact, they typically had to be reassured that this was indeed what they were supposed to be hearing.

### Procedure

The experimenter met the participants at their scheduled times and escorted them to individual cubicles. Depending on the sign-ups for a given day, it was possible to run either 1 or 2 participants during each experimental session. After taking their seats in the cubicles and reading and signing informed consent forms, the experimenter informed the participants that they would be asked to respond to a set of videotapes containing excerpts from three psychotherapy sessions in which different female clients discussed their personal problems with the same male therapist. The experimenter then returned to the control room, where he started the instructional tapes for the participants. After the participants had viewed the instructional tapes and were ready to continue, the experimenter started playing the first stimulus tape.

Following the presentation of each 15-s excerpt of the client-therapist interaction, the experimenter paused the tape for the participant. The participants’ task during each of these stops was to (a) record the time of the stop, as indicated by a display that appeared in the upper left-hand corner of the video image; (b) check a box to indicate whether they thought the target was having a thought or a feeling at that point in the interaction; and (c) write down in sentence form their inference regarding the specific content of that thought or feeling (cf. Marangoni et al., 1995).

When the participants had finished recording their inferences for a given thought–feeling entry, they restarted the tape by means of a remote control device located in each cubicle. The tape then continued through each 15-s excerpt of the client-therapist interaction until the participants had responded to all 30 excerpts. After the participants completed this task, the experimenter returned to the cubicule, collected the thought–feeling inference forms used for that tape, and provided the new forms to be used for the next tape. The empathic inference procedure then continued as before. After the participants had completed their inferences for all three tapes, the experimenter conducted a complete debriefing and encouraged the participants to ask any questions they had about the procedure.

### Computation of Empathic Accuracy Scores

The participants’ empathic accuracy scores were subsequently computed for each of the psychotherapy tapes in the following manner. Using a custom software program, “Collect your thoughts,” developed by Stephen Trued and William Ickes, eight independent raters compared the content of each of the actual thoughts and feelings reported by the clients with the corresponding inferences made by the experimental participants. The software program allows each actual thought or feeling and the corresponding paired inference to be presented together on the computer screen (for a full description of this program, see Ickes, Bissinnette, Garcia, & Stinson, 1990). Using the coding software, the raters successively viewed each actual–inferred pair on the computer screen and, at the appropriate prompt, entered their rating of the degree of perceived similarity between the content of the actual thought–feeling entry and the content of the corresponding thought–feeling inference. The degree of similarity was rated on a 3-point scale (0 = essentially different content, 1 = somewhat similar, but not the same, content, 2 = essentially the same content). The computer program kept track of all the rating data and created files of both the raw and the aggregated (i.e., averaged) rating data for subsequent analysis.

To determine whether the data from all eight raters should be used, we calculated the interrater reliabilities for each tape, using Cronbach’s alpha. The reliabilities for the three tapes were uniformly high: Divorce 1, $\alpha = .92$; Divorce 2, $\alpha = .89$; Role Conflict, $\alpha = .89$. Accordingly, the item-by-item empathic accuracy scores were averaged across the eight raters, summed across all 30 entries, and divided by 60 (the maximum number of accuracy points that could be obtained for each stimulus tape). The resulting empathic accuracy scores for each tape are analogous to percentage scores, having a potential range of 0 (zero accuracy) to 100 (perfect accuracy).

### Results

The experimental design was a $3 \times 2 \times (3)$ mixed factorial in which channel (VA, VFA, and AO) and sequence (original and random) were the between-subjects factors and stimulus tape (Divorce 1, Divorce 2, and Role Conflict) was the within-subjects factor.

### Cross-Target Consistency in Empathic Accuracy

Intercorrelations of the empathic accuracy scores for the three tapes are shown in Table 1. The average cross-tape correlation was .84, and the intraclass correlation of empathic accuracy scores across all three tapes was .91, comparable to the .86 coefficient reported by Marangoni et al. (1995). These findings suggest that the overall empathic accuracy scores reflect an underlying skill dimension on which perceivers displayed an impressive level of cross-target consistency in performance.

The high level of cross-target stability observed in this study might in part be attributable to the fact that the three stimulus tapes

---

2 The following functional description of this system was provided by Golden Strader, the electronics technician who built it. First, the audio signal from the videotapes was sent through a high-pass filter to remove background sounds below the vocal range. Second, the signal was passed through a series of 12 low-pass filters. This was necessary because the range of frequencies was so large that 12 filters were required to effectively remove the enunciation harmonics of both male and female voices (the male therapist and the female client). Third, a clipping filter was used to flatten the top and bottom of the waveform, essentially removing most of the remaining linguistic information while preserving virtually all of the paralinguistic information.
were relatively homogeneous in terms of both the clients (e.g., all were middle-class, college-educated, White women) and the problems they discussed (women’s relationship issues). To see if it might also be attributable to between-channel differences in the diagnosticity of the information available, we calculated intraclass correlations within each level of the channel variable: VA, α = .73; VFA, α = .62; AO, α = .85. These data, which are reasonably consistent across channels, are compelling in their implication that individual differences in empathic accuracy can be measured very reliably using our videotape-assessment procedure.

Statistically Controlling for Tape (Target) Differences

As we expected, there was a significant main effect for tape, $F(2, 70) = 23.88, p < .001$, that replicated the one reported previously by Marangoni et al. (1995). Consistent with Marangoni et al.’s study, we found that the average level of empathic accuracy was significantly lower for the Divorce 2 tape ($M = 22.8$) than for the Role Conflict or Divorce 1 tapes, for which the mean levels of empathic accuracy were not significantly different ($M_s = 28.6$ and 29.2, respectively). As Marangoni et al. noted when discussing this difference, the client in the Divorce 2 tape is particularly difficult to read because of the ambivalence she expressed throughout the tape about whether to get a divorce. In contrast, the clients in the other two tapes were easier to read because both of them expressed more consistent reactions to the problems they discussed (see also Ickes et al., 1997; and for a general discussion of target judgability, see Colvin, 1993).

Because the main effect of tape was not of theoretical interest in the present study, we standardized the empathic accuracy scores (converted to z scores) within each tape, and we used the standardized scores in all subsequent analyses. A preliminary analysis of the standardized scores revealed no significant effects involving either the tape variable or the order-of-presentation variable. Accordingly, unless otherwise noted, we collapsed the data across these variables in the analyses reported below. Moreover, because it is more informative to report empathic accuracy means on the original 0–100 scale (the values are easier to interpret, and they facilitate cross-study comparisons), we elected to follow the convention of our previous studies and report our findings in terms of the percentage analogue (0–100) scale.

Channel Effects

Archer and Akert (1980) proposed three theories of how meaning might be decoded in social interactions. If, consistent with the significant clue theory, accuracy is primarily a function of uniquely diagnostic information that must be present in at least one channel (presumably, the verbal channel in the case of empathic accuracy), then the mean empathic accuracy scores should be approximately the same in the VA and the AO conditions, but in both cases significantly higher than in the VFA condition. If, consistent with the additive theory, accuracy is a linear function of the sheer amount of information available across different channels, then the mean empathic accuracy scores should be highest in the VA condition and about equal in the AO and VFA conditions. Finally, if, consistent with diffusion theory, all clues are meaningful in their own right and provide information about some aspect of the situation, then the mean empathic accuracy scores should differ little, if at all, across the three channel conditions.

As we expected, the results we obtained were most consistent with the significant clue theory (see Table 2 and Figure 1). The analysis of variance revealed a large channel effect for each of the three tapes and for overall empathic accuracy: Divorce 1, $F(2, 67) = 60.36, p < .001$; Divorce 2, $F(2, 67) = 30.88, p < .001$; Role Conflict, $F(2, 67) = 28.56, p < .001$; overall empathic accuracy, $F(2, 67) = 50.24, p < .001$. Tukey’s multiple comparison test indicated that there was no difference between the VA condition ($M = 34.14$) and the AO condition ($M = 32.70$), but that the means for these two conditions both differed significantly from that of the VFA condition ($M = 10.03$; $p < .01$ for both comparisons).

These results indicate that the most diagnostic channel (i.e., the most significant clue) for empathic accuracy in the present study was the verbal, rather than the visual, channel. The linguistic content of the client–therapist conversation was particularly im-

### Table 1

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>All tapes</td>
<td>.90</td>
<td>.91</td>
<td>.93</td>
<td>.91</td>
</tr>
<tr>
<td>Role Conflict</td>
<td>.89</td>
<td>.75</td>
<td>.75</td>
<td>.75</td>
</tr>
<tr>
<td>Divorce 1</td>
<td>.92</td>
<td>.79</td>
<td>.79</td>
<td>.79</td>
</tr>
<tr>
<td>Divorce 2</td>
<td>.89</td>
<td>.89</td>
<td>.89</td>
<td>.89</td>
</tr>
</tbody>
</table>

**Note.** Values on the diagonal are interrater reliabilities. All correlations, $p < .0001$.

### Table 2

<table>
<thead>
<tr>
<th></th>
<th>Video and audio</th>
<th>Audio only</th>
<th>Video and filtered audio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stimulus tape</td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
</tr>
<tr>
<td>All tapes</td>
<td>34.14</td>
<td>6.93</td>
<td>32.70</td>
</tr>
<tr>
<td>Role Conflict</td>
<td>37.60</td>
<td>8.75</td>
<td>33.14</td>
</tr>
<tr>
<td>Divorce 1</td>
<td>37.24</td>
<td>10.53</td>
<td>36.78</td>
</tr>
<tr>
<td>Divorce 2</td>
<td>27.57</td>
<td>6.67</td>
<td>28.16</td>
</tr>
</tbody>
</table>

* $p < .0001$.
portant in this regard, as evidenced by the findings that (a) empathic accuracy was highest when the client–therapist conversation was intelligible, (b) removing all of the visual cues did not impair empathic accuracy as long as the speech was intelligible, and (c) making the conversation unintelligible through electronic filtering did impair empathic accuracy, despite the availability of all of the visual cues.

Although the nonverbal (visual and paralinguistic) cues were not the most diagnostic ones in this study, they still appeared to contribute something to the perceivers’ empathic accuracy. When we tested the mean of the VFA condition \((M = 13.9)\) against the null hypothesis that the true mean was zero, we found that the difference was significant and substantial, \(t(23) = 15.47, p < .001\), Cohen’s \(d = 6.82\). Moreover, when we tested the same VFA mean \((M = 13.9)\) against an average baseline-accuracy score \((M = 4.5)\) calculated from three previous dyad studies of empathic accuracy (Gesn, 1995; Ickes et al., 1990; Stinson & Ickes, 1992), the difference was again significant, \(t(23) = 2.13, p < .05\), Cohen’s \(d = 0.90\). These comparisons suggest that although the nonverbal cues were collectively less diagnostic than the verbal (linguistic) cues in the present study, they were nonetheless sufficient to enable the VFA perceivers to achieve levels of empathic accuracy that were reliably greater than a typical chance level of responding.

In contrast to the highly diagnostic verbal cues, the visual cues themselves appeared to add little or nothing to the perceivers’ empathic accuracy, as indicated by the nonsignificant difference between the means of the VA and AO conditions. The generalizability of this null finding may be limited, however, by the restricted range of variation in the information available in the visual channel. Although there was substantial variation in the range of emotions the clients expressed, they all remained seated throughout the entire interaction, and their expressive behaviors were limited primarily to facial expressions, hand gestures, and head nods. In addition, their facial cues may have been ones in which expressions of negative or neutral affect predominated.

**Sequence Effects**

We predicted that empathic accuracy would be significantly greater in the original sequence condition than in the random sequence condition. This prediction was based on the expectation that, for the complex inferential process involved in the empathic accuracy task, preservation of a cumulative meaning context would add significantly to the perceivers’ diagnostic and interpretive accuracy, beyond the effect of the immediate cues available to the perceivers.

In apparent contradiction to this hypothesis, however, there was no overall main effect for sequence: overall empathic accuracy, \(F(1, 60) = 1.25, p > .25\); Divorce 1, \(F(1, 67) < 1\), ns; Divorce 2, \(F(1, 67) = 2.40, p > .12\); Role Conflict, \(F(1, 67) = 1.07, p > .25\). There were also no interactions between sequence and any of the other independent variables. Although the participants in the original sequence condition \((M = 27.8)\) did not perform significantly better than the participants in the random sequence condition \((M = 26.0)\), this null result should not be interpreted to mean that the sequence variable was unimportant. As we will see, one of the most powerful effects that emerged in this study was an unexpected Sequence \(\times\) Inferential Difficulty interaction.
Breakdown of the Empathic Accuracy Scores by Difficulty Level

The lack of a significant overall main effect for the sequence manipulation was puzzling, given that the overall coherence of the client-therapist interactions was obviously disrupted in the random sequence condition. We suspected that the key to this puzzle might be found in the potential moderating influence of the inferential difficulty of the individual thought–feeling entries. The individual thought–feeling entries reported by each of the clients in the psychotherapy tapes differed substantially in their level of inferential difficulty, as assessed by a group of raters in the Marangoni et al. (1995) study. In that study, seven independent raters viewed each of the three tapes and assessed the difficulty of inferring each thought or feeling on the basis of the information available in the 10 s immediately before each point at which the client had reported it. The level of inferential difficulty was rated on a 3-point scale (1 = very difficult to infer, 2 = somewhat difficult to infer, 3 = easy to infer). The interrater reliability (Cronbach’s alpha) for this judgmental task, as reported by Marangoni et al., was .80.

For the easy-to-infer entries, substantial information is available to the perceiver about what the client is thinking or feeling. For example, in many of these instances the client’s thoughts and feelings are expressed directly in what she is saying to the therapist. In such instances, the meaning of the client’s current words and actions both contributes to and is partially determined by the cumulative meaning context of the client’s unfolding account of her problem(s). For the difficult-to-infer entries, however, the available information is considerably less diagnostic. In many instances, the client’s thought or feeling is so personal, off-topic, or unrelated to the context of her current discussion with the therapist that information about the cumulative meaning context might actually be misleading.

On the one hand, the perceivers’ construction of cumulative meaning contexts for the tapes in the original sequence condition might have facilitated their empathic accuracy for those easy-to-infer thoughts and feelings that most directly contributed to (or were most consistent with) those same cumulative meaning contexts. On the other hand, to the extent that the perceivers relied on these cumulative meaning contexts, they may also have been predisposed to make plausible but inaccurate inferences about the more difficult-to-infer thoughts and feelings—the ones that were typically the most personal, off-topic, and out of context, and thus generally not consistent with the cumulative meaning contexts.

What would happen, however, if the perceivers found it difficult to construct such cumulative meaning contexts? This was the situation faced by the participants in the random sequence condition of the present study. Because these perceivers should have found it difficult to construct cumulative meaning contexts in which to frame their empathic inferences, it was unlikely that cumulative meaning contexts could have facilitated their empathic accuracy for the easy-to-infer thoughts and feelings or impaired their accuracy for the difficult-to-infer thoughts and feelings. Lacking such larger interpretive frames, these perceivers should have been forced to rely on the immediate cues available in making each individual thought–feeling inference, thereby attenuating the influence of the cumulative meaning contexts on inter-item differences in inferential difficulty.

The pattern of the results just described would take the form of a crossover interaction of Sequence × Inferential Difficulty. An interesting feature of this type of interaction is that, consistent with the results of our original analysis, no overall main effect for the sequence variable would be evident. To examine this possibility, two analyses were used to test the potential moderating influence of item difficulty on the relationship between sequence and empathic accuracy.

Analysis 1: Breakdown of the scores by rated difficulty. In the first analysis, we used the inferential-difficulty ratings that were collected in the Marangoni et al. (1995) study to categorize the 90 thought–feeling entries (3 tapes × 30 entries per tape) into three levels of difficulty: very difficult to infer (28 entries), moderately difficult to infer (36 entries), and easy to infer (26 entries). Our decision to aggregate the thought–feeling entries in each category across the three tapes was justified on the grounds that (a) the item-by-item empathic accuracy scores had already been standardized (i.e., converted to z scores) within each tape and (b) such aggregation permitted a more powerful test of the moderating variable hypothesis. However, because the variances in the empathic accuracy scores were necessarily heterogeneous across the three levels of inferential difficulty (i.e., were proportional to the means within each condition), we used the logarithms (base 10) of the standardized empathic accuracy scores to further normalize the data in this analysis.

The perceivers’ aggregated log-transformed empathic accuracy scores for the subsets of thoughts and feelings at each of the three levels of inferential difficulty were entered into a repeated measures analysis of variance, with channel and sequence as the independent variables. As expected, the results of the new analysis revealed a significant main effect of inferential difficulty, \( F(2, 65) = 10.85, p < .001 \). The perceivers’ empathic accuracy scores were highest for what the raters had judged to be the easy-to-infer thoughts and feelings and lowest for what the raters had judged to be the difficult-to-infer thoughts and feelings (easy, \( M = 31.0 \); moderate, \( M = 27.0 \); difficult, \( M = 24.4 \)). These data support the validity of the difficulty ratings obtained in the Marangoni et al. (1995) study.

The Channel × Difficulty interaction was also significant in this analysis, \( F(4, 128) = 4.32, p < .005 \). As can be seen in Figure 2, for the VA and AO conditions, the mean empathic accuracy scores decreased linearly as the inferential difficulty of the thought–feeling entries increased. In the VFA condition, however, the mean empathic accuracy scores did not differ as a function of inferential difficulty. Apparently, the raters’ judgments of inferential difficulty in the Marangoni et al. (1995) study depended so much on the information available in the specific content of the targets’ speech that the impact of the difficulty variable was not evident in the VFA condition, in which all speech content had been rendered unintelligible through electronic filtering.

Although no overall sequence effect was evident in our original analysis, when we broke down the empathic accuracy data by the inferential-difficulty level of the thought–feeling entries, a strong Sequence × Difficulty interaction was evident, \( F(4, 128) = 40.71, p < .001 \).
p < .001. To determine the nature of this interaction, we conducted univariate analyses at each level of inferential difficulty. The results of these analyses are shown in Figure 3.

For the easy-to-infer thought–feeling entries, empathic accuracy in the original sequence condition ($M = 35.2$) was significantly better than in the random sequence condition ($M = 26.9$), $F(1, 65) = 17.18$, $p < .001$, Cohen's $d = 1.04$. However, for the difficult-to-infer thought–feeling entries, empathic accuracy in the original sequence condition ($M = 21.0$) was significantly worse than in the random sequence condition ($M = 27.9$), $F(1, 65) = 9.51$, $p < .005$, Cohen's $d = 0.77$. For the thought–feeling entries of moderate difficulty, empathic accuracy scores in the original ($M = 29.1$) and random ($M = 24.9$) sequence conditions did not differ significantly, $F(1, 64) = 1.17$, $p > .25$, Cohen's $d = 0.28$.

This pattern of results suggests that, relative to the perceivers in the random sequence condition, the perceivers in the original sequence condition were better able to construct a cumulative meaning context for each tape that made it easier for them to infer the content of the generally context-consistent easy-to-infer thoughts and feelings. However, their ability to construct such cumulative meaning contexts became a liability with respect to the more context-inconsistent difficult-to-infer thoughts and feelings, presumably because using the cumulative meaning context as a basis for making inferences about these out-of-context thoughts and feelings was inappropriate and therefore misleading.

We think the most straightforward interpretation of this Sequence X Inferential Difficulty interaction is that the participants in the original sequence condition were more likely to rely on schematic context-consistent representations of the clients and their problems than were participants in the random sequence condition. Schematic representations for each client should have been more readily formed by the participants in the original sequence condition because it was in that condition that the cumulative meaning context was preserved. If we assume that the clients’ easy-to-infer thoughts and feelings were the most consistent with these schematic representations, it makes sense that the participants in the original sequence condition should have outperformed those in the random sequence condition on this subset of thought–feeling entries.

The perceivers' tendency to rely on such schematic representations in the original sequence condition should, by the same reasoning, have impaired their empathic accuracy for the more schema-inconsistent difficult-to-infer thoughts and feelings. By relying more heavily on context-based schemas to guide their inferences, the participants in the original sequence condition may have actually been at a disadvantage compared with the partici-
pants in the random sequence condition, who would have been forced to pay more attention to whatever immediate cues were most diagnostic of these out-of-context, off-topic, and difficult-to-infer thoughts and feelings.

**Analysis 2: Breakdown of the scores by actual difficulty.** To ensure that the moderating effect of item difficulty did not depend on some unidentified subjective bias(es) of the raters in the Marangoni et al. (1995) study, we repeated the analysis using an empirically based (as opposed to a ratings-based) assignment of the 90 thought–feeling entries to the three difficulty levels. We began by simply rank ordering the average empathic accuracy scores for each of the individual thought–feeling entries from each tape. The 10 entries from each tape with the highest empathic accuracy scores were designated as easy to infer, the 10 with the lowest empathic accuracy scores were designated as difficult to infer, and the 10 intermediate entries were designated as moderately difficult to infer. Aggregated (i.e., average) empathic accuracy scores were then computed across the three tapes for the 30 easy-to-infer, 30 intermediate, and 30 difficult-to-infer thoughts and feelings.

When the data were reanalyzed using this empirically based method of assessing inferential difficulty, the pattern of results was essentially the same as in the analysis using the Marangoni difficulty ratings (see Figure 4). Once again, there was the expected main effect of inferential difficulty, \( F(2, 65) = 175.00, p < .001 \). Also, as before, there was the same Channel \( \times \) Difficulty interaction, \( F(4, 128) = 21.30, p < .001 \). Of greatest theoretical interest, however, was the strong crossover interaction of Sequence \( \times \) Inferential Difficulty, \( F(2, 128) = 33.80, p < .001 \).

We again conducted univariate analyses at each level of inferential difficulty. For the easy-to-infer thought–feeling entries, empathic accuracy in the original sequence condition (\( M = 38.5 \)) was significantly greater than in the random sequence condition (\( M = 30.1 \)), \( F(1, 65) = 13.46, p < .001 \). However, for the difficult-to-infer thought–feeling entries, empathic accuracy in the original sequence condition (\( M = 17.2 \)) was significantly worse than in the random sequence condition (\( M = 21.8 \)), \( F(1, 65) = 8.51, p < .005 \). For the thought–feeling entries of moderate difficulty, empathic accuracy scores in the original (\( M = 11.7 \)) and random (\( M = 11.6 \)) sequence conditions did not differ significantly, \( F(1, 65) < 1 \), \( ns \).

In summary, whether inferential difficulty was defined empirically or in terms of raters’ more subjective judgments, the data provided converging evidence of a strong crossover interaction of Sequence \( \times \) Inferential Difficulty. This convergence is especially striking considering that the agreement between the two methods for determining inferential difficulty was far from perfect. When Cohen’s kappa statistic was used to test the degree of overlap in classification of items into difficulty categories, the value obtained could be characterized as only moderately strong, \( \kappa = .48, p < .05 \). As one might expect, however, the instances of disagreement that did occur always involved assignment to the easy versus moderate categories or to the moderate versus difficult categories.

**Testing the Schema-Consistency Interpretation**

Our schema-based interpretation of the Sequence \( \times \) Inferential Difficulty interaction was somewhat speculative in the absence of more direct empirical support. Accordingly, we sought to obtain such support by asking eight raters to independently judge the degree of schema consistency of each of the individual thoughts and feelings that were originally reported by the clients in the Marangoni et al. (1995) tapes.

Two problems had to be solved before we could proceed with this plan. First, we had to decide what the specific schemas were for each of the Marangoni et al. (1995) tapes. Second, whatever these schemas were, it was essential to ensure that we did not construct them in a way that would bias the results in favor of our schema-consistency interpretation of the Sequence \( \times \) Inferential Difficulty interaction.

Fortunately, the solution to both problems was already available. In a study that was completed before the present study was designed, Graham (1996) developed specific paragraph-length schemas for two of the Marangoni et al. (1995) tapes: the Divorce 1 and Role Conflict tapes (see the Appendix). After drafting two schema paragraphs that provided a capsule summary of each client’s specific problem and how she addressed it during the therapy session, Graham gave the paragraphs to three research assistants and asked them to compare their content with the events that occurred in the two videotaped therapy sessions and suggest any changes in the schema paragraphs that seemed appropriate. She then used the feedback provided by the three research assistants to revise the schema paragraphs until a complete consensus

---

**Figure 4.** Sequence \( \times \) Empirical Difficulty interaction.
was reached that each paragraph provided a good capsule summary of the client's problem and how she addressed it in the Divorce 1 and Role Conflict stimulus tapes.

Using the revised schema paragraphs that Graham (1996) developed for these two tapes, we provided eight independent raters with printed copies of each schema along with a list of the 30 thoughts and feelings reported by the client who appeared in the tape. We then asked them to view the tapes of these two client-therapist interactions. At each of the tape stops at which the client had indicated having had a specific thought or feeling, the raters paused the tape and, on the basis of the information contained in the interactions and in the transcripts of the clients' thoughts and feelings, rated the degree to which each thought or feeling was consistent with the content of the appropriate schema paragraph. These ratings were made on a 3-point scale (0 = schema inconsistent, 1 = ambiguous, 2 = schema consistent).

The reliability of the schema consistency ratings for the Role Conflict tape, as indexed by Cronbach's alpha, was .73, and the reliability of the schema consistency ratings for the Divorce 1 tape was .88. The results of the alpha analysis indicated that deleting the data from one or more raters would not improve overall reliability for the 30 entries in the Divorce 1 tape. Accordingly, on the basis of the mean judgments of all eight raters, we categorized the thoughts and feelings for each tape into three levels of schema consistency: schema inconsistent, schema ambiguous, and schema consistent. We then computed aggregated empathic accuracy scores for the thoughts and feelings that defined each level of schema consistency.

Analysis of Sequence × Schema Consistency interaction. The aggregated empathic accuracy scores for the subsets of thoughts and feelings at each of the three levels of schema consistency were entered into a repeated measures analysis of variance, with sequence as the independent variable. The results of this analysis revealed a significant main effect of schema consistency, F(2, 140) = 112.50, p < .0001. The perceivers' empathic accuracy scores were highest for what the raters had judged to be the schema-consistent thoughts and feelings and lowest for what the raters had judged to be the schema-inconsistent thoughts and feelings (schema consistent, M = 32.3; schema ambiguous, M = 29.1; schema inconsistent, M = 25.4).

More interestingly, and consistent with our earlier speculation, the Sequence × Schema Consistency interaction was also highly significant, F(2, 140) = 35.07, p < .0001. To determine the nature of this interaction, we conducted univariate analyses at each level of schema consistency. For the schema-consistent thought-feeling entries, empathic accuracy in the original sequence condition (M = 38.8) was significantly higher than in the random sequence condition (M = 25.8), F(1, 70) = 11.56, p < .001, Cohen's d = 0.82. However, for the schema-inconsistent thought-feeling entries, empathic accuracy in the original sequence condition (M = 22.6) was significantly worse than it was in the random sequence condition (M = 28.3), F(1, 70) = 4.77, p < .03, Cohen's d = 0.52. For the schema-ambiguous thought-feeling entries, empathic accuracy scores in the original (M = 30.7) and random (M = 27.5) sequence conditions did not differ significantly, F(1, 70) < 1, n.s., Cohen's d = 0.22.

This pattern of results clearly parallels the results for the original Sequence × Inferential Difficulty analyses. It therefore lends direct support to our claim that the participants in the original sequence condition were better able to form specific capsule-summary schemas than were the participants in the random sequence condition for the interactions between the clients and the therapist—schemas that facilitated their performance for the easy thought-feeling entries but hindered their performance for the difficult thought-feeling entries.

Analysis of Sequence × Residual Difficulty interaction. To pursue this reasoning further, we retested the Sequence × Inferential Difficulty interaction after first statistically removing the effects of schema consistency. The purpose of this analysis was to determine whether the contrasting univariate effects of sequence for the subsets of easy and difficult thought-feeling entries would still be significant once the schema consistency-inconsistency component had been statistically removed. If both of the univariate effects were no longer significant, the pattern of evidence would suggest that the schema consistency or inconsistency of the thought-feeling entries, rather than other sources of their inferential difficulty, was responsible for the significant Sequence × Item Difficulty interaction we originally observed.

To conduct this follow-up analysis, we first regressed the inferential difficulty ratings onto the ratings of schema consistency. Then, using the residual difficulty ratings as a basis for recategorization, we ranked ordered the 30 thoughts and feelings for each of the Divorce 1 and Role Conflict tapes and again created three subsets of thought-feeling entries, which in this case reflected three levels of residual difficulty. After aggregating the mean empathic accuracy ratings within each of these three levels, we then tested the sequence effect at each level of residual difficulty. Consistent with our schema-consistency interpretation, none of the univariate analyses revealed a statistically significant effect of sequence (all ps > .15). These findings suggest that schema consistency, rather than other aspects of item difficulty, was indeed the primary influence underlying the original Sequence × Item Difficulty interaction.

Discussion

Our hypotheses concerned the types of immediate and cumulative contextual cues that might affect empathic accuracy in a clinically relevant setting. First, we hypothesized that the verbal channel would carry most of the information on which accurate inferences were based because of the nature of the psychotherapy context, which not only encourages clients to express their thoughts and feelings in words but also imposes certain constraints (e.g., restricted movement) on their nonverbal behavior. Second,
we hypothesized that empathic accuracy should be affected not only by the immediate meaning context established by the clients’ and therapist’s most proximal words and actions but also by the cumulative meaning context established over the entire chronology of the client–therapist interaction.

Consistent with our expectation, we found that the perceivers’ empathic accuracy depended more on the targets’ verbal behavior than on their nonverbal behavior. The effect of channel or information modality was the strongest between-groups effect to emerge. It revealed that the average empathic accuracy scores in the VA and AO conditions were comparable and significantly larger than the average empathic accuracy score obtained in the VFA condition. This pattern of results is consistent with Archer and Akert’s (1980) significant clue theory in suggesting that, of the immediate contextual cues available to perceivers, those conveyed in the verbal channel of the client–therapist discussion were the most diagnostic. In contrast, the availability of visual cues in the VA condition did not significantly augment the accuracy based solely on the auditory cues available in the AO condition.

Also consistent with our expectation, empathic accuracy in the present study was determined not only by the targets’ words and actions that were available in the immediate meaning contexts (i.e., the 15-s intervals before each tape stop) but also by whether the cumulative meaning context of the client–therapist interaction was preserved or disrupted. Unexpectedly, however, the effect of preserving versus disrupting the cumulative meaning context was significant only in its interaction with the inferential difficulty of the individual thought–feeling items. For items that were generally consistent with the overall interpretational schema implied by the cumulative meaning context, preservation of the cumulative meaning context significantly enhanced the perceivers’ empathic accuracy. However, for items that were generally inconsistent with the overall schema, preservation of the cumulative meaning context significantly impaired their empathic accuracy.

Finally, replicating the effect reported by Marangoni et al. (1995), substantial cross-tape consistency was evident in the perceivers’ empathic accuracy scores. This effect is discussed in the section below, followed by discussions of the more theoretically novel effects for the channel and sequence variables.

**Reliability of the Empathic Accuracy Measure**

Although the stimulus tapes were relatively homogeneous in nature, the large intraclass correlation of the perceivers’ empathic accuracy scores across the three targets (both overall and within each level of channel) was nevertheless quite striking. This high cross-target correlation reveals that participants were remarkably consistent in the levels of empathic ability they displayed. Participants who displayed high empathic accuracy on one tape displayed comparably high empathic accuracy on the other two tapes, whereas participants who displayed low empathic accuracy on one tape displayed comparably low empathic accuracy on the other two. This finding suggests that the empathic accuracy measure reflects a stable and reliably measured social skill.

Interestingly, the intertwap correlations of the empathic accuracy scores in the present study were actually somewhat higher than those reported in the Marangoni et al. (1995) study, in which relatively unedited versions of the same stimulus tapes were used. These data suggest that, in future research, our highly edited tapes (in particular, those for the original sequence VA condition) could be used in place of the original Marangoni et al. tapes, permitting a major reduction in the amount of time required to conduct each experimental session, with no loss of reliability in the measurement of empathic accuracy.

**Channel Effects**

In terms of magnitude, the main effect of information channel was the strongest between-groups effect in the present study. With respect to empathic accuracy, the pattern of results was most consistent with Archer and Akert’s (1980) significant clue theory. According to this theory, different social inference tasks each rely on one or more primary types of information, without which performance suffers. This significant clue information is generally sufficient to enable perceivers to attain a satisfactory degree of accuracy, and the addition of information in other channels provides, at best, a diminishing return.

Consistent with the notion that the verbal content of the client–therapist discussion is the major significant clue for empathic accuracy, participants in the VA and the AO conditions of the present study obtained comparably high levels of empathic accuracy. Moreover, subtracting the significant clue of verbal content substantially impaired empathic accuracy (comparisons of the VFA condition with the VA and AO conditions), whereas adding the available visual cues did not further enhance it (comparison of the VA and AO conditions).

The results of the present study therefore suggest that the verbal content of the client–therapist discussion is the primary cue, along with the paralinguistic information contained therein (the secondary cue), may be the significant and sufficient clues for making relatively accurate thought–feeling inferences in this clinically relevant situation. This conclusion is consistent with the results of other studies in which the importance of linguistic information in other social information processing tasks has been reliably demonstrated (e.g., Berry, Pennebaker, Mueller, & Hiller, 1997; Borkenau & Liebler, 1992). For example, in the Berry et al. study, various linguistic variables (e.g., number of negative versus positive emotion words and number of self-referents) significantly predicted perceivers’ tendencies to form accurate impressions of targets even after the effects of traditional visual variables (e.g., attractiveness and expressiveness) had been statistically accounted for.

Our conclusion about the relative importance of verbal versus visual cues should be tempered, however, by the recognition that there was a restricted range of variation of behavioral information available in the visual channel—a feature of the present study that could readily account for why visual cues did not significantly facilitate empathic accuracy beyond the level attributable to auditory cues alone. On the one hand, the present findings stand in contrast to those of emotion decoding studies in which perceivers seemingly attend to visual cues to a greater extent than audio cues and give visual cues the most weight in judgments of a target person’s affect (e.g., DePaulo et al., 1978; Gitter et al., 1972; Mehrabian & Ferris, 1967). On the other hand, we agree with DePaulo and Friedman (1998) that such channel effects should not be overinterpreted and overgeneralized, as they will always be dependent on the situational context in which they are studied.
Should we have attempted to include additional channel conditions in the present study? Probably not. Our pilot testing convinced us that few, if any, participants would be willing to complete 2-hr sessions in which they tried to infer the thoughts and feelings of three clients from silent videotapes (a video only condition) or from audiotapes in which the client–therapist conversations were completely unintelligible because of electronic filtering (a filtered audio only condition). The few participants who would be willing to complete the experimental sessions in these two conditions would likely have been so highly self-selected (on dimensions such as compliance, tolerance for frustration, etc.) that such individual-difference characteristics would inevitably have confounded the interpretation of any significant effects obtained.

Including a verbal-transcript-only condition would have been even more difficult to justify. In this condition, participants would neither see nor hear the client on the tape; instead, they would be presented with only a verbal transcript of the client’s and therapist’s conversation. It is clear that this condition would have introduced a number of confounding elements. First, because perceivers would be required to rely on a communication channel (reading comprehension) that was not present in the original client–therapist interactions, the interpretation of any significant effects involving this condition would be compromised by the fact that the decoding skill required in this condition was not one required in the original client–therapist interactions. Second, the perceivers in this condition would have to work with written information that has been abstracted from the client’s and therapist’s behavior rather than with information that is inherent in their concrete behavior, thus introducing a confound on the dimension of informational abstraction or concreteness. Third, there are other potentially confounding elements (e.g., differences between the participants’ reading time and the speaking time of actual client–therapist conversations) that would also cloud the interpretation of any results obtained.

**Sequence Effects**

The strong moderating effect of the inferential difficulty of specific thoughts and feelings on the relation between the sequence manipulation and empathic accuracy was the most unexpected of the findings but also the most intriguing from a theoretical standpoint. The participants watched the thirty 15-s segments of all three stimulus tapes either in the order that they actually occurred or in one of two random orders that should have disrupted the cumulative meaning context of the client–therapist interaction. The question of interest was, Would empathic accuracy depend in part on the cumulative meaning context, beyond the effect of the immediate context defined by the verbal and nonverbal cues present within each of the 15-s segments?

We predicted that empathic accuracy would be impaired in the random sequence condition as compared with the original sequence condition. Such an effect would be analogous to the impairment in understanding that might be expected if an individual were to try to follow the plot of a movie with the scenes presented in a random order. Contrary to this prediction, however, in our initial analyses we found neither a main effect of sequence nor any interactions of sequence with the other independent variables. Our attempt to understand this puzzling outcome led us to speculate that the sequence variable might have interacted with the inferential difficulty of the specific thoughts and feelings reported by the clients in the psychotherapy tapes. Specifically, we hypothesized that the expected sequence effect would be evident only for those thoughts and feelings that were related to the cumulative meaning context of the client–therapist discussion (i.e., for those entries that were categorized as easy to infer in the context they appeared). In contrast, for those thoughts and feelings that were unrelated to the cumulative meaning context (i.e., those categorized as difficult to infer in the context they appeared), preservation of the cumulative meaning context might actually bias the perceivers’ inferences in a way that would reduce their empathic accuracy.

When we tested this possibility using both subjective and empirical ratings of inferential difficulty, we found strong crossover interactions of Sequence × Inferential Difficulty. In both analyses, there was a facilitative effect of sequence for the easy-to-infer entries, with the participants in the original sequence condition displaying greater empathic accuracy than the participants in the random sequence condition. However, for the difficult-to-infer entries, the sequence effect was also significant but in the opposite direction. That is, participants in the random sequence condition actually achieved greater empathic accuracy for the difficult-to-infer entries than did the participants in the original sequence condition. (There was no sequence effect for the thought–feeling entries of intermediate difficulty.)

Follow-up analyses taking into account the consistency of each thought–feeling entry with a capsule-summary schema for the client and her problem indicated that the participants in the original sequence condition were indeed more likely to rely on such schematic representations of the clients and their problems than were the participants in the random sequence condition. Viewed collectively, the results of the follow-up analyses suggest that their schematic representations of the clients biased the inferences of the original sequence participants in a way that facilitated their empathic accuracy for the clients’ schema-consistent thoughts and feelings (cf. Andersen, Klatzky, & Murray, 1990; Andersen, Klatzky, & Roberta, 1987), but impaired their accuracy for the clients’ schema-inconsistent thoughts and feelings (cf. Sillars, 1997).

This latter finding may help to explain the “illusion of understanding,” which, according to Sillars and his colleagues (Sillars, 1995, 1997; Sillars, Pike, Jones, & Murphy, 1984; Sillars, Weissberg, Burggraf, & Zeitlow, 1990), is relatively common among intimates and other well-acquainted individuals (see also Ickes & Simpson, 1997; Thomas, Fletcher, & Lange, 1997). As Sillars (1997) has noted in this regard,

---

6 The use of a transcript-only condition would necessarily introduce other confounding elements as well. For example, would the participants’ reading time vary systematically from the speaking time of the clients’ and therapist’s actual conversations? Should the transcribed conversation be presented in closed caption on the otherwise blank TV monitors or in printed form on sheets of paper? If the former, should the time each captioned speaking turn appears on the screen reflect (a) the actual speaking time it occupied during the actual conversation, (b) whatever time each participant might need to read and comprehend it, or (c) a constant interval that takes no account of either type of variation? Because any or all of these elements could confound the interpretation of any significant effects involving a transcript-only condition, it would be difficult to justify including such a condition in a study of channel effects.
Although being another person’s “expert” allows one to make sophisticated and organized inferences, it also increases the tendency to view potential sources of information selectively and to make new inferences fit existing relationship theories (Thomas & Fletcher, 1997). Over time, frequently activated inferences may become increasingly entrenched and monitoring of the relationship for new information might decline as people listen more selectively and assume they “have heard it all before.” (p. 80)

In extreme cases, the illusion of understanding can even be fatal to the individuals’ relationship, as, for example, in the case of a partner whose day-to-day enactment of the “faithful spouse” role is so familiar and predictable that the other partner fails to attend to the subtle schema-inconsistent cues that would signal adulterous thoughts and feelings to a less schema-driven perceiver.

The difficult-to-infer thoughts and feelings in the present study derived at least part of their difficulty from their schema inconsistency. Our present findings suggest that when such schemas cannot readily be developed (i.e., when the cumulative meaning contexts are disrupted), schema-inconsistent thoughts and feelings are actually significantly easier to infer than in cases in which such schemas can be developed (see Figures 3 and 4). The current study may therefore provide a useful empirical and methodological precedent for studying the illusion of understanding in future research and establishing the processes that underlie it and the conditions in which it occurs.

Conclusions

The present study is one of the first to examine the contextual cues that affect empathic accuracy in a clinically relevant setting. Categorizing the various cues that influence empathic accuracy under the headings of immediate contextual knowledge and cumulative contextual knowledge provided a useful conceptual framework for examining the channel and sequence variables that were the primary focus of this study. Immediate contextual knowledge can be characterized by a target’s verbal and nonverbal behavior immediately prior to his or her reported thought or feeling. In contrast, cumulative contextual knowledge must be constructed by the perceiver over a more extensive time period by tracking the unfolding meaning of the target persons’ words and actions.

Immediate contextual knowledge was varied by presenting participants with the three stimulus tapes in one of three conditions that were distinguished by the different information channels that were available. Unlike studies of the judgment of discrete emotional states, in which the visual channel is often accorded primary importance, the data from the present study revealed that the verbal content of the target person’s conversation was the most significant clue for empathic accuracy. Having the visual cues available in addition to the verbal cues added very little to the perceivers’ accuracy. When the verbal content was rendered unintelligible, leaving only paralinguistic cues, empathic accuracy was impaired to a great degree, but the residual accuracy component was still significantly different from zero. We would argue against the overgeneralization of any such channel effects, however, because of their dependence on the situational context in which they occur (DePaulo & Friedman, 1998; Friedman, 1978). Rather than argue for the general primacy of a particular channel, it may be more productive to determine the conditions under which each channel may serve as the significant clue for the judgment task at hand.

Cumulative contextual knowledge was varied by presenting the participants with the three stimulus tapes in one of two sequences: original or random. The most theoretically intriguing finding of the present study was an unexpected but very robust Sequence X Schema Consistency interaction effect. It suggests that when perceivers in a clinically relevant setting are able to use the cumulative meaning context to develop schematic representations of individual clients and their problems, these representations will increase their empathic accuracy for schema-consistent thoughts and feelings but impair their accuracy for schema-inconsistent ones. Relative to uninformed perceivers who have not been able to develop such schematic representations, such informed perceivers may be lulled into a false sense of confidence about how well they really know what the target is thinking or feeling. When the disposition (thought, feeling, motive, or trait) to be inferred is a schema-consistent one, schema-driven judgments may be expected to facilitate performance. However, when the disposition to be inferred is schema inconsistent, or when there is non-diagnostic information present along with diagnostic information, perceivers may fall victim to a form of “interpretive overkill” that leads to an inferior judgment (Tetlock & Boettger, 1989; see also Bond, Omar, & Bonser, 1990). The theoretical and practical implications of this finding with respect to empathic accuracy should provide promising avenues for future research.

References


DePaulo, B. M., Rosenthal, R., Eisenstal, R. A., Rogers, P. L., & Finkel-
Appendix

Schemas for Stimulus Tapes

Schema for the Divorce 1 Tape

The therapy session you are about to view depicts a woman whose four-year marriage had recently ended in divorce. She recalls how she and her husband met and, soon after, married. She attributes the divorce to the fact that they knew each other for only a short time before marrying, and she describes how she quickly began to resent his domineering personality, even though she initially adopted a role that enabled him to act out his dominance. Since then, she has been adjusting to being single again and has also attempted to develop a better, more honest relationship with her parents, both of whom are remarried. She also discusses how she has grown comfortable with being herself, as opposed to adopting roles that she thinks others might prefer.

Schema for the Role Conflict Tape

The therapy session you are about to view depicts a woman who is feeling overwhelmed by her attempt to juggle her responsibilities as a wife with her responsibilities as a career professional. She explains that she spends her weekends doing time-consuming chores around the house, while her husband chooses to play golf rather than spend time with her or help around the house. She is becoming resentful toward her husband and loses her temper over minor hassles both at home and at work. She discusses possible solutions to her problems including exercise, relaxation techniques, changing her belief system with respect to housework, and learning not to worry about circumstances that are beyond her control.

Received January 29, 1999
Revision received April 26, 1999
Accepted April 26, 1999