

Sources of Accuracy in the Empathic Accuracy Paradigm

Judith A. Hall
Northeastern University

Marianne Schmid Mast
University of Neuchâtel

In the empathic accuracy paradigm, perceivers make inferences about the naturalistically occurring thoughts and feelings of stimulus persons, and these inferences are scored for accuracy against the stimulus persons' self-reported thoughts and feelings. The present study investigated sources of accuracy in this paradigm by presenting the stimulus tape in several cue modalities (full video, audio, transcript, or silent video) and with differing instructions (infer thoughts and feelings, infer thoughts, or infer feelings). Verbal information contributed the most to accuracy, followed by vocal nonverbal cues. Visual nonverbal cues contributed the least, though still at levels above zero. When asked to infer feelings, perceivers appeared to shift attention toward visual nonverbal cues and away from verbal cues, and the reverse occurred when they were asked to infer thoughts. The study contributes to understanding of factors contributing to accuracy in the empathic accuracy paradigm.

Keywords: empathic accuracy, interpersonal sensitivity, nonverbal, channel, modality

Researchers who study person perception are often interested in identifying the cues that contribute to impressions about a person. Sometimes the search is focused on a specific cue—for example, whether perceivers take note of the muscle at the corner of the eye when deciding whether a smile indicates spontaneous enjoyment (Ekman, Davidson, & Friesen, 1990; Frank, Ekman, & Friesen, 1993). Sometimes the search is more exploratory, for example to identify what cues might matter most in forming impressions of a person's veracity (Zuckerman, Koestner, & Driver, 1981), personality (Borkenau & Liebler, 1995), intelligence (Murphy, Hall, & Colvin, 2003), or dominance (Hall, Coats, & Smith LeBeau, 2005). In other research, the goal might not be to identify specific cues but rather to describe the relative contributions of different channels or modalities—such as face, body, voice tone, or words—in forming such impressions (e.g., O'Sullivan, Ekman, Friesen, & Scherer, 1985; Scherer, Scherer, Hall, & Rosenthal, 1977).

The study of accuracy in interpersonal perception is also an area in which the contribution of different cues or channels has been studied. Sometimes a lens model approach is taken to identifying what cues are validly utilized by perceivers when making judgments (e.g., Bernieri & Gillis, 2001). Interest in different cues or channels has led some investigators to isolate

them at the outset, as in nonverbal sensitivity tests that present only the face (Nowicki & Duke, 1994), postures (Pitterman & Nowicki, 2004), or content-masked speech (Scherer, Banse, & Wallbott, 2001). Some test developers have gone even further in the molecular direction by isolating elements within such channels (e.g., eyes only, Baron-Cohen, Wheelwright, & Jolliffe, 1997; intonation contours only, Thompson, Schellenberg, & Husain, 2004). One test (Profile of Nonverbal Sensitivity) consists of separate face, body, and voice tone channels that are presented to perceivers both singly and in combination (Rosenthal, Hall, DiMatteo, Rogers, & Archer, 1979).

On the other hand, some investigators of accuracy seek to maximize ecological validity by using stimuli that are presented holistically, not separated into channels (e.g., Costanzo & Archer, 1989). This mode of presentation bears the greatest resemblance to how people make inferences about others in daily life, though in daily life the channels are sometimes restricted, as in talking on the telephone. When the stimuli are presented holistically, an opportunity presents itself for understanding which cues or channels carry the most useful information and/or are attended to preferentially by perceivers when they make their holistic judgments. For example, Archer and Akert (1977)'s Situational Interpretations Test (SIT) included both the full visual stimulus and the expressors' spontaneous speech. By analyzing the accuracy of perceivers who were given only a typed transcript of the words on which to base their judgments, Archer and Akert concluded that verbal content did not contribute to accuracy on the SIT task.

The present research is concerned with understanding sources of accuracy in the empathic accuracy paradigm, a method for assessing interpersonal accuracy that is based on perceivers' holistic judgments of others' expressive behavior (Ickes, Stinson, Bissonnette, & Garcia, 1990; Ickes, 2001). This paradigm addresses the difficult problem of how to measure interpersonal sensitivity in an interacting dyad. After engaging in an interaction, dyad members view their videotape individually and make two sets of judgments. First, they identify the thoughts and feelings

Judith A. Hall, Department of Psychology, Northeastern University; Marianne Schmid Mast, Department of Work and Organizational Psychology, University of Neuchâtel.

We thank William Ickes for his comments on this article; Yota Gikas and Hilary Reichlin for running participants and preparing tapes; and Heather Berry, Jared Majerle, Rie Nagasawa, Elizabeth Roderick, and Caitlin Tarwater for data scoring and entry.

Correspondence concerning this article should be addressed to Judith A. Hall, Department of Psychology, 125 NI, Northeastern University, Boston, MA 02115. E-mail: j.hall@neu.edu; or Marianne Schmid Mast, Department of Work and Organizational Psychology, University of Neuchâtel, Rue de la Maladiere 23, CH-2000, Neuchâtel, Switzerland. E-mail: marianne.schmid@unine.ch

they were having during the interaction, and second, they are asked to infer the thoughts and feelings of their partner at each point where the partner had indicated having a thought or feeling during his or her own tape review.

Accuracy is calculated by comparing each person's inferences about what the partner was thinking and feeling to what the partner reported thinking and feeling. For example, if the person inferred that the partner was "feeling irritated" at a given moment, and the partner had indeed said he or she was feeling irritated (or something equivalent in meaning), then the person would get credit for accurately perceiving the partner. Ickes called this accuracy "everyday mind reading" (2003) because it captures the kind of inference process that people engage in virtually continuously when interacting with another person.

The empathic accuracy paradigm has also been adapted to a standardized test format, so that the accuracy of many perceivers can be tested using one set of audiovisual excerpts for which accuracy is measured as the match between the inferences made by perceivers and the self-reported thoughts and feelings of the individuals shown on the videotape. As an example, Marangoni, Garcia, Ickes, and Teng (1995) showed perceivers three videotapes of clients talking in a psychotherapy session, each tape containing 30 thought-feeling entries that had been identified by the original clients. Both the dyadic and the standardized test versions of the empathic accuracy paradigm enable a researcher to gather naturalistic (i.e., unposed, unscripted) episodes of behavior and then score them for accuracy using a criterion that has high ecological validity. As such, it represents an important addition to the existing methods for measuring interpersonal sensitivity that often involve posed or otherwise artificially generated stimuli. The empathic accuracy paradigm has been used to address questions relating to gender (Ickes, Gesn, & Graham, 2000), acquaintanceship (Thomas & Fletcher, 2003), motivation (Klein & Hodges, 2000), and relationship threat (Simpson, Ickes, & Blackstone, 1995).

The present study addresses two issues that remain unresolved in research using the empathic accuracy paradigm. The first concerns the contributions of different cue modalities to accuracy, and the second concerns the impact of how the task instructions are worded.

Modality Influences on Accuracy

When viewing and listening to a partner, or to a person having a conversation on videotape, what cues are important for achieving accuracy about that person's thoughts and feelings? One study has been conducted to examine the contribution of different cue modalities, or channels, in this paradigm. Gesn and Ickes (1999) developed a standardized empathic accuracy test and presented it to perceivers in three modalities: original full video, audio only, and video plus electronically filtered audio. In the latter, the voice tones could be heard, but the words were made unintelligible by electronic filtering. These three modalities vary in the kinds and amounts of information available. Full video has visual nonverbal cues, vocal nonverbal cues, and words; audio omits the visual nonverbal cues while leaving vocal nonverbal cues and words; and video plus filtered speech omits the words while leaving visual and vocal nonverbal cues.

Gesn and Ickes found that accuracy was dramatically higher, and approximately the same, in the two conditions that contained

the words compared to the condition that omitted the words, leading them to conclude that the verbal content is the primary source of information for achieving accuracy in this paradigm. However, there was some accuracy even in the video plus filtered speech (i.e., no words) modality, indicating that at least some useful information about thoughts and feelings was available even when only nonverbal cues were present.

Although Gesn and Ickes (1999) provided important information about the contribution of different cue modalities to accuracy, the study did not allow a full understanding of how different cue channels contribute to accuracy. Gesn and Ickes' cue modalities each contained two or three channels (combinations of visual nonverbal, vocal nonverbal, and/or verbal). There were no modalities containing only one channel. In the present study, we constructed a standardized version of an empathic accuracy test and presented it to perceivers in four modalities: original full video, audio only, transcript only, and silent video only. Thus, compared to the full video modality that contained three channels (visual nonverbal cues, vocal nonverbal cues, and verbal cues), the audio modality contained two channels (vocal nonverbal cues and verbal cues), and the transcript and silent video modalities contained one channel each.¹

With these four cue modalities, we could replicate and extend Gesn and Ickes' (1999) finding that accuracy depends mostly on the other person's words. In their study, this conclusion was reached by comparing modalities that did and did not include words, but they did not examine a modality that contained only the words. Moreover, although Gesn and Ickes argued that using electronically filtered speech to separate words from vocal nonverbal cues was an effective way to test whether vocal nonverbal cues matter, filtered speech is only one way to approach this problem and one that has its own difficulties. Possibly, listening to such an unusual (and frustrating) stimulus channel may have distracted perceivers and hurt their accuracy. Furthermore, filtering distorts the voice and may eliminate cues that are important to accurate judgment. Wallbott and Scherer (1986) found that perceivers had lower accuracy for judging emotions in electronically filtered standard-content voice clips than for unfiltered standard-content voice clips. Therefore, the contribution of vocal nonverbal cues to accuracy was examined only partially in the Gesn and Ickes study.

In addition to appraising the role of verbal cues in accuracy, we also examined an additional hypothesis, also addressed by Gesn and Ickes, that accuracy depends on how many sources of information are present in the stimulus. Therefore, this study complements that of Gesn and Ickes and offers the possibility of a more comprehensive understanding of channel effects.

Wording of Judgment Task Instructions

The second issue addressed in the present study concerns the nature of the judgment that perceivers are asked to make. In the

¹ Of course, the concept of "channel" is arbitrary. We recognize that an indeterminate number of more narrowly defined channels may comprise what we are calling a channel (e.g., for the face channel, each region of the face could be considered a separate channel).

standard paradigm, thoughts and feelings are considered together. The interactants who are videotaped are asked to identify their “thoughts and feelings” when they perform the tape review, and the perceivers who are subsequently doing the inferring (whether they are the stimulus person’s interaction partner or strangers who view the tape later) are similarly asked to write down “thoughts and feelings.”

Could this distinction matter? It could matter if perceivers assume that feelings are conveyed mainly through nonverbal cues and thoughts are conveyed mainly through words, a belief we think is likely to be widespread. Indeed, we confirmed this belief in an informal study of 19 adults who were not psychologists. They were asked in a brief written questionnaire to imagine people’s nonverbal cues and words in a conversation, and to indicate whether nonverbal cues would be likely to express thoughts more than feelings, feelings more than thoughts, or equal amounts of thoughts and feelings. They were then to make the same judgment about the verbal cues (words). Although some respondents said that nonverbal and verbal cues would express equal amounts of thoughts and feelings, there was a clear association of nonverbal cues with feelings and words with thoughts (see Table 1).

On the basis of these lay beliefs, we thought it likely that perceivers in the empathic accuracy paradigm might direct their attention to different kinds of cues if they were asked separately about feelings versus thoughts, and accuracy might be influenced accordingly. Specifically, they might rely more on nonverbal cues when judging feelings and more on words when judging thoughts (or when judging thoughts and feelings together). Possibly, the standard instruction which mentions both thoughts and feelings directs perceivers’ attention to the “thoughts” component and therefore disproportionately to verbal cues. Such a pattern would be consistent with Gesn and Ickes’ (1999) conclusion that words were the most important source of information.

To test this hypothesis, we varied the task instructions so that perceivers were asked to infer thoughts, infer feelings, or infer both thoughts and feelings (i.e., the standard instruction). We predicted that the pattern of means in the different conditions as well as the pattern of correlations between modalities would show evidence that the amount of attention paid to words versus nonverbal cues would shift depending on instruction condition.

In summary, we sought to understand further the sources of accuracy in the empathic accuracy paradigm by varying the modalities of cue presentation (full video, audio only, transcript only, or silent video only) and the nature of the instructions (infer thoughts, infer feelings, or infer thoughts and feelings).

Table 1
Survey of Lay Beliefs About Expression of Thoughts and Feelings

Response	Nonverbal cues	Words
Express thoughts more than feelings	0	11
Express feelings more than thoughts	10	1
Express equal amounts of feelings and thoughts	9	7

Method

Participants

Perceivers were 197 (91 male, 104 female; 2 gender unreported) students who were recruited from introductory psychology classes in Northeastern University’s Psychology Department and given partial course credit for participation. Participants had a mean age of 19.0 years (range 17–24), and their self-reported ethnicity was 77% Caucasian, 9% Asian, 5% Hispanic, 3% African American, and 7% other. In addition, 6 students (2 male, 4 female) served as expressors; they were selected from an initial sample of 24 students (mostly female), recruited in the same manner as the perceivers. Finally, 6 students (gender unmeasured), also recruited in the same way, served as preraters of the stimulus materials.

Test Tape

Twenty-four expressors (targets) were videotaped during a competitive interaction with 24 peers. Before the interaction, each dyad member generated a list of the five most influential movies in the past 10 years and then, during the subsequent 8-min videotaped interaction, tried to convince the other about his or her choices during an unstructured discussion. The person who succeeded in putting more of his or her initial movies on their final list won \$2.00. After this interaction, the expressors watched the videotape again and were asked to stop it at exactly the location where they had a thought or a feeling during the interaction and to write down the thought or feeling, along with the corresponding time on the video timer. They were explicitly instructed to report only thoughts and feelings they distinctly remembered experiencing while interacting and not to report any thoughts and feelings they experienced for the first time while watching the tape.

From the pool of 24 expressors, nine were selected on the basis that they expressed both thoughts and feelings (according to the information they provided during tape review). Typed transcripts of these self-reported thoughts and feelings were then presented to six preraters who rated them on a scale that went from 1 (*mostly a thought*) through 5 (*thought/feeling mix*) to 9 (*mostly a feeling*). Reliability among the six preraters was good (Cronbach’s alpha = .76). The eight responses receiving the highest mean rating (purest feelings) and the eight responses receiving the lowest mean ratings (purest thoughts) were selected to be edited onto the stimulus tape (representing six expressors). Examples of thoughts were “Why am I doing this?” and “Stop going so fast”; examples of feelings were “bored” and “beaten.”

On the test tape, excerpts of interaction averaged 27.56 s in duration (range = 9–41 s). The camera was focused on the expressor with a profile of the partner sometimes in view, and the voices of both were audible. Each excerpt was preceded by a number on the screen and terminated immediately after the point at which the expressor had described having the thought or feeling, consistent with Ickes’ (2001) standard methodology. Thoughts and feelings were intermingled and were not identified as such to perceivers.

Procedure

Perceivers were run in small groups, with each group being assigned quasi-randomly to one presentation modality (full video,

audio, silent video, or transcript). Within each group, perceivers were randomly assigned to judgment condition: infer thoughts, infer feelings, or infer thoughts and feelings. Wording of the instructions appears in the Appendix. Table 2 shows the cell sizes for the 4×3 experimental design.

In the full video (sound turned on) and silent video (sound turned off) conditions, as well as in the audio (no video) condition, the tape was stopped after each excerpt (at exactly the same moments in each condition) for perceivers to write down their inferences. In these conditions, it was obvious to perceivers who the target was. In the audio condition, the procedure was the same except that the experimenter informed perceivers of which of the two interactants was the target. For the transcript condition, a verbatim transcript of each excerpt was prepared that showed the words uttered by both individuals in the interaction, with those of the designated target clearly identified. Perceivers read the transcribed utterances at their own pace and made their inferences as in the other conditions.

Scoring of Accuracy

Scoring was done following Ickes' (2001) general description by two trained coders who both scored all perceivers. For each of the 16 excerpts, the coders compared the perceiver's inference to the original target's self-reported thought or feeling, using a 0 (*no similarity*) to 2 (*fully acceptable degree of similarity*) scale (with .5 increments), and scores in between meant degrees of similarity. A score of 2 could be given for a verbatim match in wording or an inference that the coders considered synonymous with the target's self-report. To illustrate, the correct answer for item 16 was "BORED!" (i.e., this is what the expressor said his or her feeling was at this point in the interaction). For this item, the perceiver who responded "she/he's probably feeling kind of interested because she/he couldn't explain and now the other person is explaining because he or she doesn't interrupt like they have been" earned no points; the perceiver who responded "calm, just listening; no emotion" earned 1 point; and the perceiver who responded "the person is thinking they are uninterested and feeling bored and

uncomfortable" earned 2 points. Across all 197 perceivers, agreement between the two coders was excellent, with a median correlation (Pearson r) across the individual 16 test items of .80 (range of r s = .65 to .93). Intercoder reliability for total score was $r(195) = .86, p < .001$. The two coders' scores were averaged, and summary scores were calculated for actual thoughts, actual feelings, and total.

Overall accuracy was $M = 6.87$ ($SD = 3.26$), where the maximum could have been 32 (16 items \times 2 points each). Comparison of overall accuracy against zero (no accuracy) was highly significant, $t(196) = 29.54, p < .001$.² Male and female perceivers did not differ significantly in their accuracy overall, for actual thoughts and feelings separately, or according to modalities or instruction conditions ($p > .26$), with one exception: when judging actual feelings in transcript (across instruction conditions), women were more accurate than men, $t(60) = 2.10, p < .05$, effect size $r = .26$. Because gender effects were not otherwise evident, gender was not entered into subsequent analyses.

Results

Modality Effects in the Standard Instruction Condition

To parallel the Gesn and Ickes (1999) study, we first examined accuracy for the participants in the standard instruction condition (infer thoughts and feelings). A 4×2 analysis of variance (ANOVA) was run in which presentation modality (full video, audio, transcript, or silent video) was a between-participants factor and content (actual thoughts or actual feelings) was a repeated measure factor. The main effect of content was not significant, $F(1, 68) = 1.02, p < .32$, meaning that actual thoughts and actual feelings were equally easy to judge in this condition. There was a significant condition effect, $F(3, 68) = 11.66, p < .001$. Table 2, top panel, shows the means for this analysis.

Two planned comparisons were conducted on the total scores. The first compared the modalities containing verbal information (first three listed in table) to the modality having no words (silent video), using weights of 1, 1, 1, and -3 , respectively. This contrast was highly significant, $F(1, 68) = 23.79, p < .001$, effect size $r = .51$, supporting the conclusion of Gesn and Ickes (1999) that the words carry more useful information than the other modalities (M for modalities with words = 8.01, M for silent video = 4.15).

The second contrast tested an ordering of modalities as a function of how many channels were represented in each (weights = 1.25 for full video, .25 for audio, $-.75$ for transcript, and $-.75$ for silent video). This contrast was also highly significant, $F(1, 68) = 26.10, p < .001$, effect size $r = .53$. Examination of the means shows that although the contrast was highly significant, it was not perfect because the two single-channel modalities were not identical in accuracy (transcript being somewhat better than silent

Table 2
Accuracy as a Function of Modalities, Content, and Instructions

Modality	Actual thoughts	Actual feelings	Total (SD)	Cell n
Infer thoughts and feelings (standard instruction)				
Full video	4.06	5.00	9.06 (2.65)	21
Audio	4.90	3.96	8.86 (2.79)	13
Transcript	2.74	3.38	6.12 (3.02)	24
Silent video	1.86	2.29	4.15 (2.41)	14
Infer thoughts				
Full video	3.02	4.27	7.29 (3.78)	14
Audio	3.59	3.52	7.11 (2.56)	11
Transcript	3.39	3.02	6.41 (2.83)	21
Silent video	2.59	2.11	4.70 (2.77)	11
Infer feelings				
Full video	3.80	5.20	9.00 (3.57)	16
Audio	3.38	4.56	7.94 (3.05)	13
Transcript	3.00	3.38	6.38 (3.38)	19
Silent video	2.44	2.94	5.38 (2.36)	20

² Because scores could only range upwards from zero, any mean greater than zero would likely produce a significant effect. An alternative "no-accuracy" baseline value was developed by rearranging the responses of a sample of 17 participants, in a different order for each participant, and then scoring them in the usual manner (see also Gesn & Ickes, 1999, for a similar procedure). The new (pseudo) mean accuracy for these 17 individuals was 1.89. A t -test of the overall mean accuracy against this alternative no-accuracy baseline was also significant, $t(196) = 21.41, p < .001$.

video). The two contrasts were not orthogonal (correlation between the weights = .52).

Further insight came from post hoc (Tukey) tests between the modalities (see Table 2). Full video was better than silent video ($p < .001$) and transcript ($p < .01$), and audio was better than silent video ($p < .001$) and transcript ($p < .03$). Silent video and transcript were not significantly different from each other, nor were full video and audio significantly different from each other. Accuracy was greater than zero in all four modalities by one-sample t tests ($p < .001$). Overall, this pattern suggests that although words mattered, the nonverbal cues contained in the voice also mattered because both full video and audio were significantly better than transcript. On the other hand, the visual nonverbal cues were the least helpful: full video was not significantly better than audio only, and silent video was the least accurate channel (though not significantly worse than transcript). Silent video did, however, produce accuracy greater than zero.

Modality Effects in the Infer Thoughts and Infer Feelings Conditions

Infer thoughts condition. In this condition, the same 4×2 ANOVA as above was conducted (between perceivers: modality; within perceivers: content, i.e., actual thoughts vs. actual feelings). This ANOVA yielded no significant effects, $p > .11$, though the totals for the four modalities (Table 2, middle panel) showed the same rank order as in the standard instruction condition described in the preceding section.

Infer feelings condition. The same ANOVA conducted in this condition yielded two significant effects (Table 2, bottom panel). There was a main effect of content, $F(1, 64) = 6.80$, $p = .01$, showing that actual feelings were more accurately judged than actual thoughts, unlike in the other two instruction conditions. Table 2 shows that the instruction to infer feelings increased accuracy for actual feelings (compared to the other two instruction conditions) while not influencing accuracy for actual thoughts. This suggests that when asked to infer feelings, perceivers paid special attention to the nonverbal cues. There was also a main effect of modality, $F(3, 64) = 4.71$, $p = .005$, showing a rank order of means that was the same as in the other instruction conditions.

The contrast comparing the three modalities that contained words to silent video was significant, $F(1, 64) = 7.77$, $p < .01$, effect size $r = .33$, showing again that accuracy was higher when the words were available than when words were not available (M for modalities with words = 7.77, M for silent video = 5.38); however, this effect was smaller than the corresponding contrast for the standard instruction condition, suggesting that words did not matter as much when perceivers were asked to infer feelings. Note that accuracy on silent video was greater for actual feelings than in the other instruction conditions, though not significantly so (see Table 2). Post hoc tests for the infer feelings condition showed that full video was marginally better than transcript ($p = .07$) and significantly better than silent video ($p < .01$), and audio was marginally better than silent video ($p = .10$). The planned contrast that compared the modalities on how many channels were represented in each was also significant, $F(1, 64) = 12.86$, $p < .001$, effect size $r = .41$, indicating that accuracy increased as more channels were included in the stimulus.

Correlations Between Full Video and Other Modalities

Thus far, we have addressed the question of what cues perceivers attended to by comparing mean accuracy in different cells. Another way to examine this question is to correlate accuracy between full video and the other modalities (cf. Ekman, Friesen, O'Sullivan, & Scherer, 1980; Hall, Braunwald, & Mroz, 1982; Scherer et al., 1977). To illustrate this logic, imagine that accuracy in full video was correlated highly with transcript accuracy and not at all with silent video accuracy. This would strongly suggest that when judging full video, perceivers relied on words to the exclusion of visual nonverbal cues. By comparing these correlations for the different instruction conditions, we can gain insight into how attention might have shifted depending on instructions. We performed this analysis by averaging accuracy across perceivers for each item, within modalities, and calculating intermodality correlations with N being the 16 items on the test.

Table 3 shows the correlations between accuracy in full video and accuracy in the other modalities. When perceivers were instructed to infer both thoughts and feelings, accuracy in full video was very highly correlated with accuracy in the transcript modality, consistent with previous analyses suggesting that the words contributed a great deal to overall accuracy in the standard instruction condition. Both of the modalities that contained nonverbal cues (silent video and audio) were less strongly correlated with full video, though both appeared to contribute (not significantly in the case of silent video).

The correlations in the other two instruction conditions suggest a somewhat different allocation of attention. In the infer thoughts condition, the silent video modality appeared not to contribute at all to full-video accuracy, suggesting that perceivers shifted attention away from visual nonverbal cues when asked to infer thoughts. On the other hand, audio and transcript both were positively correlated with full-video accuracy, suggesting that perceivers put their attention preferentially onto the words. In the infer feelings condition, in contrast, full-video accuracy also seemed to be based on audio and transcript, but now the correlation with silent video was of considerably stronger magnitude (p values for this analysis were limited by the fact that df was based on the number of items in the test, which was only 16). Thus, when inferring feelings, perceivers in the full-video modality appeared to distribute their attention more evenly across all of the channels, giving more attention to visual nonverbal cues than in the infer thoughts condition.

Table 3
Correlations Between Full Video and Other Modalities

Instruction condition	Correlation of full video with		
	Audio	Transcript	Silent video
Infer thoughts and feelings (standard instruction)	.64**	.90***	.33
Infer thoughts	.70**	.56*	-.06
Infer feelings	.73***	.70**	.47+

Note. $N = 16$ items (averaged across perceivers).
+ $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

Discussion

The present study contributes to an understanding of the empathic accuracy paradigm for measuring interpersonal sensitivity (Ickes, 2001, 2003). In this paradigm, a perceiver watches and listens to a person in naturalistic interaction and is scored for how well he or she can discern the thoughts and feelings of that person, using that person's postinteraction self-reports as the scoring criteria. Because this is a complex, holistic judgment task, there is merit in investigating the judgment process to gain an understanding of what, exactly, it is measuring and how accuracy is attained.

Gesn and Ickes (1999), in their study using full video, video with electronically filtered audio, and unfiltered audio conditions, found that verbal cues contributed far more than nonverbal cues to the attainment of accuracy. This is an important finding for researchers working in the field of interpersonal sensitivity measurement because generally there is an assumption that interpersonal sensitivity depends mainly on the processing of nonverbal cues. Gesn and Ickes did find that exposure to nonverbal cues alone (video accompanied by filtered audio) allowed for some degree of accuracy, but it was far less than occurred in the two conditions in which perceivers could hear the targets' words.

The present study pursued this question further by presenting perceivers with a somewhat different array of cue channels, specifically full video, unfiltered audio, words (transcript), and silent video. We found that when perceivers could judge only silent video, accuracy was quite low but still above zero. When they could hear both words and voice (audio condition), and when they had both visual and vocal nonverbal cues along with the words (full video), accuracy was significantly better than when they could only read the words. Thus, it would be wrong to conclude that all perceivers really needed was the words, for accuracy based on transcript fell substantially behind the modalities that combined words with nonverbal cues. We should therefore conclude that nonverbal cues played a role in accuracy along with verbal cues though, consistent with Gesn and Ickes (1999), verbal cues dominated as a source of accuracy.

The fact that verbal cues are very important in this paradigm may help to explain an apparent paradox with regard to gender differences. Gender differences are not reliably present in the empathic accuracy paradigm as meta-analytically reviewed by Ickes et al. (2000) and as found in the present study. This stands in contrast to the well-established female superiority on interpersonal sensitivity tasks that exclude verbal cues or minimize their role (see meta-analytic reviews by Hall, 1978, 1984, and by McClure, 2000). It may be that the socialization of women produces superior interpersonal judgment skills mainly in the nonverbal realm and that tasks that rely mainly on verbal cues for the attainment of accuracy will not show a marked gender difference. It is interesting to note, however, that in the present study the only significant evidence of female superiority was in the transcript condition, with no gender difference in the conditions that contained nonverbal cues. Thus, the evidence regarding gender differences in the empathic accuracy paradigm is very mixed at present.

Do Verbal Cues Really Matter More Than Nonverbal Cues?

To find that verbal cues matter a great deal and nonverbal cues by themselves matter comparatively little may seem surprising in

light of previous experiments on cue combinations. Such experiments have often found that nonverbal cues matter more than verbal cues in impression formation (Argyle, Alkema, & Gilmour, 1972; Bugental, Kaswan, & Love, 1970; Mehrabian & Ferris, 1967; Mehrabian & Wiener, 1967). There are exceptions to this finding (Bugental, Kaswan, Love, & Fox, 1970), and sometimes the relative contribution of verbal versus nonverbal cues depends on what kind of judgment the perceiver is making (Mehrabian, 1970) or what kind of behavior is being judged (O'Sullivan et al., 1985). Nevertheless there is much evidence that nonverbal cues can contribute heavily to perceivers' impressions and, presumably, to accuracy. What, then, accounts for the apparent inconsistency between the present results and this research tradition? Does it mean that when people engage in everyday mind reading (Ickes, 2003), as opposed to judging experimentally contrived stimuli, nonverbal cues actually do not matter much? In the following paragraphs we suggest possible reasons why nonverbal cues are less important than verbal cues in the empathic accuracy paradigm.

Words may predominate because perceivers look first to the words as a source of information and turn only secondarily to nonverbal cues. Perceivers may turn to nonverbal cues mainly when the words are uninformative for the judgment in question or there is a glaring verbal-nonverbal contradiction. Therefore, if the words people utter in a naturalistic situation typically contain useful information, it is not unreasonable to expect that the words will predominate as a source of accuracy. By this logic, if an interpersonal sensitivity test were designed to include only uninformative verbal cues, it should show a proportionate increase in the importance of nonverbal cues. This is, in fact, the case with the Interpersonal Perception Task (Costanzo & Archer, 1989) and the Situational Interpretations Test (Archer & Akert, 1977); on both tests, the nonverbal cues predominate over the verbal cues in determining accuracy.³

Moreover, if perceivers are attempting to identify the targets' thoughts, this may further accentuate reliance on verbal cues. In the present study, some perceivers were asked to identify thoughts and feelings (the standard instruction within this paradigm), but some were only asked to identify thoughts and others were only asked to identify feelings. Analyses of these different conditions revealed that the words mattered most in the two conditions in which perceivers were asked to identify thoughts (either thoughts only or both thoughts and feelings), while nonverbal cues gained in importance when perceivers were only asked to identify feelings. Thus, the instructions influenced what cues perceivers used, consistent with our informal survey respondents' belief that words reveal more about thoughts and nonverbal cues reveal more about feelings.

³ It is interesting that vocal nonverbal cues outweigh visual nonverbal cues in the empathic accuracy paradigm, considering that research generally finds lower accuracy for judging vocal nonverbal cues than visual (especially facial) nonverbal cues (Rosenthal et al., 1979; Wallbott & Scherer, 1986). The fact that the vocal channel weighs more than the visual channel in the empathic accuracy paradigm may reflect the fact that perceivers are attending to the good information contained in the words and therefore are strongly oriented toward the voice. Therefore, the relatively high value of vocal nonverbal cues (compared to visual nonverbal cues) may be a by-product of perceivers' close monitoring of speech.

One implication for judgments of real-life naturalistic communication is that the relative impact of verbal cues should vary with whether perceivers are seeking insight into cognitive versus emotional processes. However, perceivers in daily life are probably interested in both thoughts and feelings and would have trouble making the distinction in any case. Therefore, the standard instruction condition used in the empathic accuracy paradigm, which asks perceivers to attend to both, is probably most similar to what perceivers do in daily life, and if this is the case, then verbal cues probably predominate in everyday communication as well.

The hypothesis that perceivers generally turn first to verbal information implies that nonverbal cues are less informative than verbal cues. Nonverbal cues could be less informative for two possible reasons. The first possibility is that thoughts and feelings are not predictably revealed through nonverbal cues. Thoughts, in particular, may not have many nonverbal correlates—as our informal survey respondents believed. But even feelings may not always have nonverbal correlates. Although research has established some nonverbal correlates of basic emotions and of a few other states such as anxiety, embarrassment, and pride (see review in Knapp & Hall, 2005), there are many other states that a layperson would call “feelings” that have not been investigated with regard to associated nonverbal cues. It is unknown, therefore, whether the “feelings” identified by targets in studies using the empathic accuracy paradigm are typically accompanied by informative nonverbal cues.

The second possibility is that even if thoughts and feelings are accompanied by nonverbal cues, such cues may be ambiguous in meaning. Nonverbal cues are, indeed, notoriously ambiguous. Studies that succeed in finding high rates of accuracy in judging nonverbal cues typically use posed expressions, expressions selected to be easily judged, prototypical expressions, and/or basic emotions (e.g., Ekman et al., 1987; Matsumoto et al., 2000; Nowicki & Duke, 1994; Rosenthal et al., 1979; Wallbott & Scherer, 1986). But to find that these kinds of nonverbal cues are judged with high accuracy does not mean that the nonverbal cues displayed in everyday interactions will be similarly easy to judge. Indeed, the thoughts and feelings experienced by people in a naturalistic situation may have complex, unfamiliar, and possibly idiosyncratic displays associated with them. Thus, even if nonverbal cues are useful in principle, they would not be useful in practice if perceivers are unsure what the cues mean. Perceivers may therefore rely on the words because they involve less guesswork. In short, because there is more “mind reading” involved in interpreting nonverbal cues, perceivers may take the path of least resistance by relying on the words if this is at all possible.

Thus far, our discussion would apply to any situation in which a person is judging naturalistic communication, whether it is in daily life or in the empathic accuracy paradigm per se. A final possible explanation for the predominance of verbal cues stems from the specific methodology employed in the empathic accuracy paradigm. Here, the concern is that the scoring criteria may be biased toward verbal cues. As described earlier, the “right answers” used for scoring empathic accuracy are retrospective self-reports made by expressors while reviewing their own videotape. These self-reports may combine accurate recollections along with reconstructions or new inferences based on watching and listening to themselves (Bem, 1967; Nisbett & Wilson, 1977). To the extent that expressors may be influenced by their own words that they

hear upon replay, the scoring criteria would be weighted toward verbal information. Consequently, a perceiver would be more likely to be scored correct if he or she also attended to the verbal information. Though this bias may not be very likely considering the use of clear instructions to expressors that they should report only thoughts and feelings that they are sure they experienced during the interaction, as well as the accumulated evidence for the validity of the paradigm (Ickes, 2001, 2003; Ickes, Robertson, Tooke, & Teng, 1986); nevertheless, it cannot be entirely ruled out.

Thus, there are plausible reasons why verbal cues matter so much in the empathic accuracy paradigm as well as in the naturalistic communication situations we confront all the time in daily life. It is important to emphasize, however, that this does not mean that nonverbal cues did not matter. Both Gesn and Ickes (1999) and the present study found that the greatest accuracy was achieved with a combination of channels (the more, the better). Having more channels can increase redundancy and also increase the likelihood that relevant information will occur in at least one of the channels. In addition, we showed that the relative contribution of different channels is influenced by the kind of judgment perceivers were making, in this case thoughts versus feelings. A similar demonstration of a shift of focus was made by Zuckerman, Spiegel, DePaulo, and Rosenthal (1982), who showed that perceivers suspecting deceit (wisely) discounted the face as a source of information about the truthfulness of the message. As that study and others cited earlier suggest, the relative importance of verbal and nonverbal cues may depend very much on the particulars of the cues to be judged, the expressors’ and perceivers’ goals, and other situational factors.

Limitations

The present study captured only a small number of excerpts of naturalistic communication; therefore, the results have unknown generalizability. However, it is encouraging that key results replicated those of Gesn and Ickes (1999), who used a different collection of videotaped excerpts. Our task proved to be somewhat more difficult than theirs, though it is not clear whether this could have influenced the comparisons we made. Considering that the likelihood of perceivers getting correct answers by guessing was extremely low (see footnote 2), and that even our relatively low accuracy was significantly above this level, we think it unlikely that the difficulty of our task posed a threat to the validity of the results.

A limitation in terms of ecological validity stems from the fact that, like Gesn and Ickes (1999), our study was based on a standardized stimulus tape rather than the reactions of perceivers in live dyadic interaction. Whether the present results would pertain in live interaction cannot be assessed at present. The use of a standardized tape is, however, required not only to permit the presentation of cues in different modalities but also to avoid any confounding influence of an interaction partner. In fact, research on the live-dyad version of the empathic accuracy paradigm can be ambiguous for this very reason.

References

- Archer, D., & Akert, R. (1977). Words and everything else: Verbal and nonverbal cues in social interaction. *Journal of Personality and Social Psychology, 35*, 443–449.

- Argyle, M., Alkema, F., & Gilmour, R. (1972). The communication of friendly and hostile attitudes by verbal and non-verbal signals. *European Journal of Social Psychology, 1*, 385–402.
- Baron-Cohen, S., Wheelwright, S., & Jolliffe, T. (1997). Is there a “language of the eyes”? Evidence from normal adults, and adults with autism or Asperger syndrome. *Visual Cognition, 4*, 311–331.
- Bem, D. J. (1967). Self-perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological Review, 74*, 183–200.
- Bernieri, F. J., & Gillis, J. S. (2001). Judging rapport: Employing Brunswik’s lens model to study interpersonal sensitivity. In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal sensitivity: Theory and measurement* (pp. 67–88). Mahwah, NJ: Erlbaum.
- Borkenau, P., & Liebler, A. (1995). Observable attributes as manifestations and cues of personality and intelligence. *Journal of Personality, 63*, 1–25.
- Bugental, D. E., Kaswan, J. W., & Love, L. R. (1970). Perception of contradictory meanings conveyed by verbal and nonverbal channels. *Journal of Personality and Social Psychology, 16*, 647–655.
- Bugental, D. E., Kaswan, J. W., Love, L. R., & Fox, M. N. (1970). Child versus adult perception of evaluative messages in verbal, vocal, and visual channels. *Developmental Psychology, 2*, 367–375.
- Costanzo, M., & Archer, D. (1989). Interpreting the expressive behavior of others: The Interpersonal Perception Task. *Journal of Nonverbal Behavior, 13*, 225–245.
- Ekman, P., Davidson, R. J., & Friesen, W. V. (1990). The Duchenne smile: Emotional expression and brain physiology: II. *Journal of Personality and Social Psychology, 58*, 342–353.
- Ekman, P., Friesen, W. V., O’Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., et al. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology, 53*, 712–717.
- Ekman, P., Friesen, W. V., O’Sullivan, M., & Scherer, K. (1980). Relative importance of face, body, and speech in judgments of personality and affect. *Journal of Personality and Social Psychology, 38*, 270–277.
- Frank, M. G., Ekman, P., & Friesen, W. V. (1993). Behavioral markers and recognizability of the smile of enjoyment. *Journal of Personality and Social Psychology, 64*, 83–93.
- Gesn, P. R., & Ickes, W. (1999). The development of meaning contexts for empathic accuracy: Channel and sequence effects. *Journal of Personality and Social Psychology, 77*, 746–761.
- Hall, J. A. (1978). Gender effects in decoding nonverbal cues. *Psychological Bulletin, 85*, 845–857.
- Hall, J. A. (1984). *Nonverbal sex differences: Communication accuracy and expressive style*. Baltimore: Johns Hopkins University Press.
- Hall, J. A., Braunwald, K. G., & Mroz, B. J. (1982). Gender, affect, and influence in a teaching situation. *Journal of Personality and Social Psychology, 43*, 270–280.
- Hall, J. A., Coats, E. J., & Smith LeBeau, L. (2005). Nonverbal behavior and the vertical dimension of social relations: A meta-analysis. *Psychological Bulletin, 131*, 898–924.
- Ickes, W. (2001). Measuring empathic accuracy. In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal sensitivity: Theory and measurement* (pp. 219–241). Mahwah, NJ: Erlbaum.
- Ickes, W. (2003). *Everyday mind reading: Understanding what other people think and feel*. Amherst, NY: Prometheus Books.
- Ickes, W., Gesn, P. R., & Graham, T. (2000). Gender differences in empathic accuracy: Differential ability or differential motivation? *Personal Relationships, 7*, 95–109.
- Ickes, W., Robertson, E., Tooke, W., & Teng, G. (1986). Naturalistic social cognition: Methodology, assessment, and validation. *Journal of Personality and Social Psychology, 51*, 66–82.
- Ickes, W., 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.
- Klein, K. J. K., & Hodges, S. D. (2000). Gender differences, motivation, and empathic accuracy: When it pays to understand. *Personality and Social Psychology Bulletin, 27*, 720–730.
- Knapp, M. L., & Hall, J. A. (2005). *Nonverbal communication in human interaction*, 6th Ed. Belmont, CA: Wadsworth.
- Marangoni, C., Garcia, S., Ickes, W., & Teng, G. (1995). Empathic accuracy in a clinically relevant setting. *Journal of Personality and Social Psychology, 68*, 854–869.
- Matsumoto, D., LeRoux, J., Wilson-Cohn, C., Raroque, J., Kookan, K., Ekman, P., et al. (2000). A new test to measure emotion recognition ability: Matsumoto and Ekman’s Japanese and Caucasian Brief Affect Recognition Test (JACBART). *Journal of Nonverbal Behavior, 24*, 179–209.
- McClure, E. B. (2000). A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. *Psychological Bulletin, 126*, 424–453.
- Mehrabian, A. (1970). When are feelings communicated inconsistently? *Journal of Experimental Research in Personality, 4*, 198–212.
- Mehrabian, A., & Ferris, S. R. (1967). Inference of attitudes from nonverbal communication in two channels. *Journal of Consulting Psychology, 31*, 248–252.
- Mehrabian, A., & Wiener, M. (1967). Decoding of inconsistent communications. *Journal of Personality and Social Psychology, 6*, 109–114.
- Murphy, N. A., Hall, J. A., & Colvin, C. R. (2003). Accurate intelligence assessments in social interactions: Mediators and gender effects. *Journal of Personality, 71*, 465–493.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231–259.
- Nowicki, S., Jr., & Duke, M. P. (1994). Individual differences in the nonverbal communication of affect: The Diagnostic Analysis of Nonverbal Accuracy Scale. *Journal of Nonverbal Behavior, 18*, 9–35.
- O’Sullivan, M., Ekman, P., Friesen, W. V., & Scherer, K. R. (1985). What you say and how you say it: The contribution of speech content and voice quality to judgments of others. *Journal of Personality and Social Psychology, 48*, 54–62.
- Pitterman, H., & Nowicki, S., Jr. (2004). A test of the ability to identify emotion in human standing and sitting postures: The Diagnostic Analysis of Nonverbal Accuracy-2 Posture Test (DANVA2-POS). *Genetic, Social, and General Psychology Monographs, 130*, 146–162.
- 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.
- Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology, 32*, 76–92.
- Scherer, K. R., Scherer, U., Hall, J. A., & Rosenthal, R. (1977). Differential attribution of personality based on multi-channel presentation of verbal and nonverbal cues. *Psychological Research, 39*, 221–247.
- Simpson, J. A., Ickes, W., & Blackstone, T. (1995). When the head protects the heart: Empathic accuracy in dating relationships. *Journal of Personality and Social Psychology, 69*, 629–641.
- Thomas, G., & Fletcher, G. J. O. (2003). Mind-reading accuracy in intimate relationships: Assessing the roles of the relationship, the target, and the judge. *Journal of Personality and Social Psychology, 85*, 1079–1094.
- Thompson, W. F., Schellenberg, E. G., & Husain, G. (2004). Decoding speech prosody: Do music lessons help? *Emotion, 4*, 46–64.
- Wallbott, H. G., & Scherer, K. R. (1986). Cues and channels in emotion recognition. *Journal of Personality and Social Psychology, 51*, 690–699.
- Zuckerman, M., Koestner, R., & Driver, R. (1981). Beliefs about cues associated with deception. *Journal of Nonverbal Behavior, 6*, 105–114.
- Zuckerman, M., Spiegel, N. H., DePaulo, B. M., & Rosenthal, R. (1982).

Nonverbal strategies for decoding deception. *Journal of Nonverbal Behavior*, 6, 171–187.

Appendix

Wording of Instructions

Full video and silent video. You will see a series of short excerpts of two people discussing movies that they found influential. They were asked to agree on the best ones from both of their lists. The focus of the camera is on one person, called the subject. Your job is to pay close attention to the conversation and, as soon as the tape stops, write down what you think the subject was thinking (feeling; thinking or feeling) just before the tape stops. The tape will remain stopped while you write down your answers.

Audio. You will hear a series of short excerpts of two people discussing movies that they found influential. They were asked to agree on the best ones from both of their lists. The experimenter will be telling you to focus on one person, called the subject. Your

job is to pay close attention to the conversation and, as soon as the tape stops, write down what you think the subject was thinking (feeling; thinking or feeling) just before the tape stops. The tape will remain stopped while you write down your answers.

Transcript. You will read a series of short excerpts of two people discussing movies that they found influential. They were asked to agree on the best ones from both of their lists. The focus of the transcript is on one person, called the subject. Your job is to pay close attention to the conversation and, as soon as the dialogue ends, write down what you think the subject was thinking (feeling; thinking or feeling) just before the dialogue stops.

Note. In all modality conditions, the heading above the answer spaces said “The person was thinking,” “The person was feeling,” or “The person was thinking/feeling.”

Received January 30, 2006

Revision received November 7, 2006

Accepted January 15, 2007 ■