

On the difficulty of distinguishing “good” and “poor” perceivers: A social relations analysis of empathic accuracy data

WILLIAM ICKES,^a ANN BUYSSE,^b HAO PHAM,^a
KERRI RIVERS,^a JAMES R. ERICKSON,^a
MELANIE HANCOCK,^a JOLI KELLEHER,^a AND PAUL R. GESN^c
^aUniversity of Texas at Arlington, ^bUniversity of Ghent, and ^cTexas Education Agency

Abstract

Two studies were conducted to explore the reasons why replicable individual-difference correlates of empathic accuracy have proved so difficult to find. In Study 1, we examined sources of variance in empathic accuracy data using the Social Relations Model (Kenny, 1988, 1994; Kenny & Albright, 1987; Malloy & Kenny, 1986). The results revealed substantial perceiver variance only in the type of research design in which a relatively large set of individual perceivers inferred the thoughts and feelings of the same set of target persons. In Study 2, we found evidence that even in this apparently optimal type of research design, the significant individual-difference correlates of empathic accuracy were fewer and subject to more unexpected qualifications than the results of Davis and Kraus's (1997) meta-analysis would suggest. So far, the “best candidate” predictor of empathic accuracy appears to be verbal intelligence, but it remains to be seen whether it and two other recently proposed predictors of interpersonal accuracy will survive the test of replicability.

One of the most intractable problems in the study of accuracy in interpersonal judgment has been the difficulty of identifying reliable individual-difference correlates of “good” (vs. “poor”) perceivers. As Funder (1995) has noted, “Although historically the good judge is the first potential moderator [of accuracy effects] to have been addressed by research, it remains the one for which, to date, the accuracy literature has the sparsest data and the fewest firm findings to report” (p. 660, brackets added).

The force of Funder's (1995) conclusion is underscored by the fact that the history to which he refers is a relatively long one. Forty years before Funder's article appeared in print, Taft (1955) published a qualitative review of the then-available re-

search on the ability to judge other people. Taft's review suggested that the best potential correlates of this ability for adult participants were intelligence, good psychological adjustment, and aesthetic interest. Forty-two years later, when Davis and Kraus (1997) published their quantitative meta-analysis of the data from 36 “post-Cronbach investigations” (251 effects involving 32 individual-difference variables and 30 interpersonal accuracy measures), their list of the best potential correlates was only slightly longer than Taft's. Expanding on Taft's finding that “good judges” tend to be intelligent, Davis and Kraus found that “good judges” also tend to be cognitively complex, field-independent, and nondogmatic. And qualifying Taft's finding that “good judges” tend to be psychologically well-adjusted, Davis and Kraus found that “good judges” tend to be more mature and well-socialized, but not less neurotic and anxious.

Correspondence concerning this article should be addressed to William Ickes, Department of Psychology, University of Texas, Arlington, TX 76019-0528. E-mail: ickes@uta.edu.

Interestingly, however, Davis and Kraus (1997) found no evidence that “good judges” score reliably higher than “poor judges” on self-report measures of empathically relevant skills and dispositions. Their meta-analysis revealed mean effect sizes of nearly zero for various measures of self-reported dispositional empathy (.01), femininity (.00), social intelligence (.08), and social sensitivity/thoughtfulness (.04). Paradoxically, although the respondents seemed unable to provide predictively valid self-reports about their own empathic ability, aggregate ratings of their social sensitivity by peer informants were associated with respondents’ scores on the performance measures of interpersonal accuracy, with an effect size of .26. Viewed collectively, this pattern of data suggests either (a) that individuals are so biased or so lacking in self-insight that they cannot report accurately/objectively about their own level of empathic ability, or (b) that individual differences on this dimension are remarkably subtle and difficult to discern (e.g., so that only by aggregating across the judgments of a number of peer raters can predictively valid assessments be made).

In the present study, we sought to determine if there was preliminary evidence for the second of these interpretations in the available studies of empathic accuracy (i.e., accuracy in inferring the specific content of other people’s thoughts and feelings; Ickes, Stinson, Bissonnette, & Garcia, 1990). If individual differences in empathic accuracy really are subtle and difficult to discern, then Social Relations Model (SRM) analyses of empathic accuracy scores should reveal relatively little “perceiver variance” (Kenny, 1994, Ch. 7). In other words, if individual differences in perceivers’ levels of empathic ability contribute very little to the total variance in empathic accuracy scores, then the lack of such “perceiver variance” might account for such findings as (a) the evidence that most perceivers fail to display accurate “metaknowledge” regarding their own empathic ability (Ickes, 1993; Marangoni, Garcia, Ickes, & Teng, 1995; Mortimer, 1996); (b) the corresponding failure of self-

reported dispositional empathy measures to predict interpersonal accuracy scores (Davis & Kraus, 1997); and (c) the need to use aggregated (peer-report) measures to achieve such prediction.

Conversely, if the results of the SRM analyses reveal substantial perceiver variance in any of the data sets, they would be useful in helping us specify the types of research designs in which significant individual-difference correlates of empathic accuracy might be identified. By applying what are presumably the most optimal predictor variables (e.g., those identified by Davis & Kraus, 1997) in what is presumably the most optimal type of research design, it should be possible to maximize the chances of identifying replicable and reliable individual-difference correlates of empathic accuracy.

We investigated these possibilities in the two studies reported below. In Study 1, we investigated the first possibility—that there is not enough perceiver variance in empathic accuracy scores for individual-difference measures to amount to a statistically significant extent. In Study 2, we investigated the second possibility—that reliable predictors of empathic accuracy might still be found if we can identify a type of research design in which the amount of perceiver variance is substantial. We begin by considering the theoretical and empirical background for Study 1, which is developed in the following section.

Relevant Findings from the Research on Empathic Accuracy

As in the broader literature on interpersonal accuracy, there is accumulating evidence in the empathic accuracy literature that (a) most perceivers fail to display accurate “metaknowledge” of their own empathic ability (Ickes, 1993; Marangoni et al., 1995; Mortimer, 1996), and (b) self-reported dispositional empathy measures generally fail to predict perceivers’ empathic accuracy scores (Ickes et al., 1990; Ickes, Hancock, Graham, Gesn, & Mortimer, 1994; Marangoni et al., 1995).

The most direct evidence that most perceivers fail to display accurate metaknowledge about their own empathic accuracy comes from studies by Marangoni et al. (1995) and Mortimer (1996). In the first study, Marangoni et al. (1995) found that although perceivers displayed strong cross-target consistency in their empathic accuracy scores (intertarget correlations averaging about .60), they were unable to postdict their performance levels immediately after the task was complete. In the second study, Mortimer (1996) assessed perceivers' metaknowledge on an inference-by-inference basis, asking them to rate how accurate they thought they were following each of the inferences they made. When he correlated the perceivers' actual empathic accuracy scores with their self-estimated accuracy ratings across the entire set of 40 thought/feeling inferences, Mortimer found that the average "metaknowledge correlation" was negative and nonsignificant ($r = -.20$.) Moreover, this within-subject correlation was positive and significant for only a small percentage of the participants, leading Mortimer to estimate that 85% to 90% of them lacked even minimally accurate metaknowledge about the inference-by-inference variation in their own empathic accuracy scores.

Given this evidence that most perceivers have little or no metaknowledge regarding their own empathic ability, it is perhaps not surprising that—as in the much larger set of interpersonal accuracy studies reviewed by Davis and Kraus (1997)—self-reported dispositional empathy measures have consistently failed to predict perceivers' performance on empathic accuracy tasks. The two most relevant studies are replete with nonsignificant correlations between perceivers' empathic accuracy scores and a range of self-report measures that include (a) our own composite scales for assessing "everyday mind reading" ability, perspective-taking, and emotional contagion ($r_s = -.06, .05, \text{ and } .11$, respectively; Ickes et al., 1994); and (b) Davis's (1983) measures of perspective-taking, empathic concern, fantasy identification, and personal distress ($r_s = -.14,$

$.04, -.11, \text{ and } .06$, respectively; Ickes et al., 1990). Even more disturbing from the standpoint of Taft's (1955) and Davis and Kraus's (1997) reviews, the more intellectual measures of need for cognition and grade point average also failed to predict aggregated empathic accuracy scores ($r_s = .00 \text{ and } -.05$, respectively) in a study involving 128 participant-perceivers (Ickes et al., 1994).

Taken together, the findings from these and other empathic accuracy studies lead us to conclude (a) that perceivers generally lack accurate metaknowledge about their own empathic ability, and (b) that researchers have, to date, been unable to identify a single individual-difference correlate of empathic accuracy that is both statistically reliable and replicable from one study to the next. If we assume that perceiver differences in empathic accuracy are substantial and easy to discern, these data patterns are puzzling indeed. Conversely, if we assume that perceiver differences in empathic accuracy are subtle and difficult to discern (i.e., that the amount of perceiver variance is generally quite low), these data patterns are both sensible and—in retrospect, at least—to be expected.

Study 1

In Study 1, we tested the hypothesis that Social Relations Model (SRM) analyses would consistently reveal relatively little perceiver variance in empathic accuracy scores. If the level of perceiver variance is found to be consistently low across studies, this finding might be sufficient to explain (a) why most perceivers seem to lack metaknowledge of their own empathic ability, and (b) why reliable individual-difference correlates of this ability have proved so difficult to find. Conversely, if the level of perceiver variance is found to be substantial in any of the studies, additional work would be needed to (a) determine whether or not reliable individual-difference correlates of empathic accuracy are found in studies in which the level of perceiver variance is substantial, and (b) attempt to spec-

ify the features of studies in which relatively high versus low levels of perceiver variance are found.

Method and Results

To our knowledge, none of the existing studies of empathic accuracy were specifically designed for a Social Relations Model analysis. A careful search eventually revealed, however, that the data sets from four previous empathic accuracy studies (Buysse & Ickes, 1999; Hancock & Ickes, 1996; Kelleher, 1998; Marangoni et al., 1995) were amenable to some form of SRM analysis, as discussed by Kenny (1988, 1994). In response to the editorial feedback we received on an earlier version of this article, we subsequently added the data from a fifth and just-completed study designed to help answer some of the major questions we have raised. Descriptions of the five data sets used in this study and their corresponding analyses are provided below.

The Hancock and Ickes study

In the Hancock and Ickes (1996) study, two pairs of same-sex friends (i.e., 4 men or 4 women) were run in each session. One of the friends from the first pair engaged in an initial, unstructured interaction with one of the friends from the second pair. The 6-minute interaction between these two strangers was unobtrusively videotaped. The two interactants subsequently viewed the videotape in separate cubicles and paused the tape at the appropriate self-designated points to record each of the actual thoughts and feelings they remembered having had during the interaction. The two interactants then viewed the tape a second time for the purpose of inferring their interaction partner's thought or feeling at each of the partner-designated "tape stops." The friends of these two interactants then replaced the interactants in the two cubicles where they also viewed the tape of the interaction twice and attempted to infer the thoughts and feelings actually reported by

each of the two interactants. (For additional details, see Hancock & Ickes, 1996.)

The empathic accuracy data provided by the friends of the two interactants were amenable to a social relations analysis using Kenny's (1994, pp. 232–234) half-block design. Guided by this design, we analyzed the empathic accuracy data for 30 same-sex, 4-person groups. In each of these groups, two perceivers (each a friend of one of the interactants) inferred the thoughts and feelings of two target individuals whose initial, unstructured interaction had been recorded on videotape. (As noted above, the two interacting target individuals also subsequently inferred each other's thoughts and feelings from the videotape, but their data were not included in the present analysis because they did not meet the fundamental condition that each perceiver must make judgments about more than one target; see Hancock & Ickes, 1996.) Consistent with the requirements of the half-block design, the two noninteracting perceivers within each group did not serve as targets, and the two interacting targets within each group did not serve as perceivers.

The variance partitioning of the perceiver and target effects in a half-block design is a random-effects analysis of variance (Kenny, 1994, pp. 233–234). Using the empathic accuracy data from all 30 groups of two perceivers and two targets, the SRM analyses were first calculated for each group separately. The resulting perceiver and target variances were then averaged across the 30 groups and used to estimate the perceiver and target effects in the larger population from which our sample was drawn. Specifically, the following steps were taken to derive these variance estimates for the data from the Hancock and Ickes (1996) study.

For each of the 30 groups of two perceivers and two targets, we (1) estimated the perceiver effect for each of the two perceivers, the target effect for each of the two targets, and the relationship-plus-residual effect for each of the four perceiver-target pairs; (2) calculated the mean square for the two perceivers, the mean square for the two

targets, and the mean square for the four perceiver \times target interactions; (3) calculated the perceiver, target, and relationship-plus-residual variance estimates; and (4) tested the statistical significance of the perceiver and target variances using *F*-tests. As the next steps in the analysis, we then (5) computed the average perceiver, target, and relationship-plus-residual variances across the 30 groups, and (6) also averaged the corresponding *F*-tests across the 30 groups.

Finally, to estimate the percentages of variance explained by the perceiver, target and relationship-plus-residual effects, we divided the mean perceiver variance, the mean target variance, and the mean relationship-plus-residual variance by the sum of these three variances. The resulting percentage variance estimates for the Hancock and Ickes (1996) data are reported in the first row of Table 1.¹

It is important to note that in the standard use of SRM analysis (i.e., the trait attribution domain) what is termed “perceiver variance” does not measure individual differences across targets (as it does in the present case). Instead, it measures something like the tendency for each judge to rate all targets in a similar way; and the accuracy of perceiver effects, in turn, equates to something like person-level stereotype accuracy. In the present case, however, when per-

ceiver variance is calculated for empathic accuracy scores using the half-block SRM analysis, it is equivalent to the assessment of individual differences between judges across targets. The reason for this difference is that the criterion measure and the inferential judgment are essentially “folded” into one measure of accuracy in deriving the empathic accuracy scores (whereas in trait-judgment SRM analyses the judgments and the criterion scores are kept separate). For this reason, our use of the term “perceiver variance” properly refers to individual differences in accuracy across targets.

The Marangoni et al. study

Data from the Marangoni et al. (1995) study were amenable to a social relations analysis using Kenny's (1994, pp. 232–234) half-block design. Guided by this design, we analyzed the empathic accuracy data for the 40 perceivers (20 male, 20 female) in the no-feedback (i.e., control) condition of the Marangoni et al. (1995) study. In this condition, the 40 perceivers—who were tested individually—inferred the thoughts and feelings of the same three female “clients” in simulated psychotherapy sessions. These women had each reported their thoughts and feelings immediately following a 50-minute videotaped psychotherapy session with the same male client-centered therapist. In each of these sessions, the client discussed with the therapist a real-life relationship problem that she was currently trying to resolve. The resulting tapes were later edited so that they were each approximately 30 minutes in length and represented the points in the conversation at which 30 of the client's thoughts and feelings had been reported.

Consistent with the requirements of the half-block design, the 40 perceivers within the control (no-feedback) condition did not serve as targets, and the three targets did not serve as perceivers. For the empathic accuracy data from this condition, we (1) estimated the perceiver effect for each of the 40 perceivers, the target effect for each of the three targets, and the relationship-

1. In 19 of the 30 groups the estimated target variance was negative, and in 14 groups the estimated perceiver variance was negative. In such cases, the most common interpretation is that the negative variance estimate provides evidence that the true value of the component of variance is zero. How negative estimates should be used in further calculations is, however, problematic. In his authoritative text, Searle (1971) lists six courses of action and labels them all as unsatisfactory for one reason or another. For the numbers shown in Table 1, the negative estimates were used when the means were calculated, and then the negative mean perceiver variance was set equal to zero. If the negative group estimates are set equal to zero before averaging, the mean perceiver and target variances change, and the variance proportions become .07 for perceivers, .46 for targets, and .46 for relationship and residual. Neither procedure is obviously better, but either leads to the conclusion that target variance is substantial and perceiver variance is not.

Table 1. Variance partitioning of empathic accuracy scores in five studies

Study	Sources of Variance		
	Perceiver	Target	Relationship and Residual ^a
<i>Unique Perceiver-Target Subgroups</i>			
Hancock & Ickes (1996) (30 subgroups) (2 perceivers, 2 unique targets per subgroup)	.00	.41*	.59
Buyse & Ickes (1999) (4 perceivers, 4 unique targets per subgroup)	.00	.26*	.74
<i>M</i>	.00	.34	.66
<i>Many Perceivers-Constant Targets</i>			
Marangoni et al. (1995) (control condition) (40 perceivers, 3 constant targets)	.57**	.14**	.29
Pham and Rivers (1998) (68 perceivers, 3 constant targets)	.51**	.11*	.38
Kelleher (1998) (48 perceivers, 8 constant targets)	.09**	.23**	.67
<i>M</i>	.39	.16	.45

^aNot tested.* $p < .05$. ** $p < .005$.

plus-residual effect for each of the 120 perceiver-target pairs; (2) calculated the mean square for the 40 perceivers, the mean square for the three targets, and the mean square for the 120 perceiver \times target interactions; (3) calculated the average perceiver, target, and relationship-plus-residual variance; (4) tested the statistical significance of the perceiver and target variances using *F*-tests; and (5) estimated the percentage of variance explained by the perceiver, target, and relationship-plus-residual effects (see Table 1, fourth row).

The Buyse and Ickes study

The Buyse and Ickes (1999) study used a round-robin design in which the four members of 28 mixed-sex quartets (2 males, 2 females) inferred each other's thoughts and feelings during brief (6 minute) laboratory discussions about safer sex. Each quartet consisted of two heterosexual dating couples. The two members of each couple took turns discussing safer sex—on one occasion with their own dating partner and on a second occasion (order counterbalanced) with the opposite-sex dating partner from the

other couple. (For additional details, see Buyse & Ickes, 1999.)

As in the case of the Hancock and Ickes (1996) data, the SRM analyses were calculated for each quartet separately. The estimated perceiver and target variances were then averaged across the 28 quartets. The variance partitioning of the perceiver and target effects in a round-robin design requires essentially the same procedure used for a half-block design, although the corresponding formulas are somewhat more complicated. Consistent with Kenny's (1994, pp. 236–238) recommendations, the following steps were taken to derive the appropriate variance estimates for the empathic accuracy data from the Buyse and Ickes (1999) study.

For each of the 28 quartets, we (1) estimated the perceiver effect for each of the four perceivers, the target effect for each of the four targets, and the relationship-plus-residual effect for each of the 12 perceiver-target pairs; (2) calculated the mean square for the four perceivers, the mean square for the four targets, and the mean square for the 12 perceiver \times target interactions; (3) calculated the perceiver, target and relationship-

plus-residual variance estimates; and (4) computed the average perceiver, target, and relationship-plus-residual variances across the 28 quartets. Finally, we (5) tested the statistical significance of the average perceiver and target variances (in the manner described below); and (6) estimated the percentages of variance explained by the perceiver, target and relationship-plus-residual effects (see Table 1, second row).

Significance testing for the round-robin design required that, after the appropriate variance estimates had been computed for each group (steps 1–3), one-sample *t*-tests were then conducted to determine whether the mean of the variance estimates for a given effect was significantly different from zero. Because the mean perceiver variance was negative for the empathic accuracy data in the Buysse and Ickes (1999) study, its value was set to zero (consistent with Kenny's [1994] recommendation), but its negative *t*-value was retained.

The Kelleher study

In the Kelleher (1998) study, 48 perceivers (counterbalanced by gender) inferred the thoughts and feelings of the same eight target persons (4 randomly-matched mixed-sex pairs) whose initial (6-minute) interactions were recorded on videotape. In each of the four videotapes, one target person had been given a "hidden agenda"—to try to keep the other person laughing throughout the interaction. In two of the tapes, the male target person had the hidden agenda; in the other two tapes, the female target person had the hidden agenda. (For additional details, see Kelleher, 1998.)

The resulting empathic accuracy data were amenable to analysis using Kenny's (1994, pp. 232–234) half-block design with one group of 48 perceivers and eight targets. As in the case of the Hancock and Ickes (1996) data, the perceivers did not serve as targets and the targets did not serve as perceivers. Following Kenny's (1994, pp. 233–234) recommendations, we (1) estimated the perceiver effect for each of the 48 perceivers, the target effect for each of the eight

targets, and the relationship-plus-residual effect for each of the 384 perceiver-target pairs; (2) calculated the mean square for the 48 perceivers, the mean square for the eight targets, and the mean square for the 384 perceiver \times target interactions; (3) calculated the average perceiver, target, and relationship-plus-residual variance; (4) tested the statistical significance for the perceiver and target variances using *F*-tests; and (5) estimated the percentages of variance explained by the perceiver, target and relationship-plus-residual effects (see Table 1, sixth row).

The Pham and Rivers study

In the Pham and Rivers (1998) study, the three videotapes originally developed in the Marangoni et al. (1995) study were used again, but in an even more highly edited form. The videotapes of the three female clients participating in individual, simulated psychotherapy sessions were edited as follows. For each "stop point" at which the client reported having had a specific thought or feeling, the 15 seconds immediately preceding that point in the interaction with the therapist were recorded onto the new stimulus videotape. This editing procedure resulted in a new stimulus tape that contained thirty 15-second excerpts from each of the three psychotherapy sessions for each of the three clients. The thirty 15-second excerpts for each client were shown to the participants in the Pham and Rivers (1998) study in their original, chronological order.

A total of 68 perceivers (27 men and 41 women) inferred the thoughts and feelings of the same three female targets. The resulting empathic accuracy data were amenable to analysis using Kenny's (1994, pp. 232–234) half-block design, with one group of 68 perceivers and three targets. Again, consistent with the requirements of the half-block design, the perceivers did not serve as targets, and the targets did not serve as perceivers.

Using Kenny's model, we (1) estimated the perceiver effect for each of the 68 per-

ceivers, the target effect for each of the three targets, and the relationship-plus-residual effect for each of the 204 perceiver-target pairs; (2) calculated the mean square for the 68 perceivers, the mean square for the three targets, and the mean square for the 204 perceiver \times target interactions; (3) calculated the average perceiver, target, and relationship-plus-residual variance; (4) tested the statistical significance for the perceiver and target variances using *F*-tests; and (5) estimated the percentages of variance explained by the perceiver, target, and relationship-plus-residual effects (see Table 1, fifth row).

Summary of the variance partitioning in the five studies

Results of the social relations variance partitioning for the five empathic accuracy studies are summarized in Table 1. These data appear to support the following tentative conclusions:

1. There are substantial differences across studies in the amount of perceiver variance in empathic accuracy scores.
2. In the Hancock and Ickes (1996) and Buysse and Ickes (1999) studies, in which (a) perceivers inferred the thoughts and feelings of only two or three target individuals and (b) the sets of targets who were judged were different for each set of perceivers, there was apparently no reliable perceiver variance at all.
3. In contrast, in the remaining three studies, those in which a relatively large set of perceivers inferred the thoughts and feelings of the same set of multiple target individuals, there was a statistically reliable amount of perceiver variance in each case. (For additional evidence that reliable perceiver differences in empathic accuracy are found in studies of this type, see Gesn, 1997; and Graham, 1996.)
4. In contrast to the striking differences in the amount of perceiver variance across the five studies, the amount of target variance in empathic accuracy scores appears to be somewhat more consistent from one study to the next. On average, the target variance accounts for about 25% of the total variance.
5. In general, the largest component of variance in empathic accuracy scores is that which combines relationship (perceiver \times target) variance with any residual sources of variance (the largest source of which is presumed to be associated with measurement error). A better understanding of this component, which accounts on average for about 55% of the total variance, requires designs in which relationship variance can be separated from the residual variance.

Separation of the relationship (perceiver \times target interaction) variance and the residual (error) variance requires at least two "scores" for each inference made by each perceiver about each of the target persons.² In other words, each perceiver would have to rate each target more than once on each "item" being judged. This requirement is problematic because, normally, the repeated ratings would not be independent. However, with the present data, an estimate of perceiver-target reliability could be obtained because each perceiver inferred several of each target's thoughts or feelings. Because empathic accuracy is known to increase with practice (Marangoni, et al., 1995) the empathic accuracy scores for odd-numbered and even-number thought/feeling inferences were scored separately in two of the studies: the Marangoni et al. (1995) study in which 40 perceivers rated the same three targets and the Hancock

2. If targets and perceivers are assumed to be random and p = the number of perceivers, t = the number of targets, and n = the number of ratings of each event, the estimate of error variance is MS residual, the estimate of relationship or target by perceiver interaction variance is (MS interaction - MS residual)/ n , the estimate of perceiver variance is (MS perceiver - MS interaction)/ nt , and the estimate of target variance is (MS target - MS interaction)/ np .

and Ickes (1996) study in which two perceivers rated two targets in each of 30 different groups.

After separating the empathic accuracy scores for the odd-numbered and the even-numbered empathic inferences in the Marangoni et al. (1995) study, we obtained these estimates: MS (Error) = 78.63, MS (perceiver \times target) = 77.44, MS (perceivers) = 543.97 and MS (targets) = 1568.45. The estimate of interaction variance is negative, so it was set equal to zero.³ The estimates of perceiver and target variance are 77.7 and 18.6, respectively, resulting in proportions of variance of 44.4 for perceivers, 10.6 for targets, and 44.9 for error. The proportions of perceiver variance and target variance are similar to, but somewhat lower than, the corresponding values shown in Table 1, perhaps reflecting the lower level of aggregation in the present analysis. (Error variance in the present analysis includes odd/even unreliability, and has a different interpretation than "residual" variance in the analysis summarized in Table 1, so no comparison between these two variance estimates is appropriate in this case.)

After separating the odd-numbered and even-numbered empathic inferences for each group in the Hancock and Ickes (1996) study, and estimating the components of variance for each group, the mean variance estimates over groups were error variance = 216.50, perceiver \times target interaction = -7.58 (thus the estimate was set equal to 0), perceiver variance = -10.42 (thus the estimate was set equal to 0) and target variance = 55.86.⁴ From these values,

3. The expected value of MS(interaction) is σ^2 error + $n\sigma^2$ interaction, and the expected value of MS(error) is σ^2 error. So the estimate of σ^2 interaction is $(MS_{\text{interaction}} - MS_{\text{error}})/n$, which is negative, and so is set to zero.

4. When variance estimates are set equal to 0 if they were negative before means were calculated, the estimates are error variance = 216.54, perceiver variance = 14.61, target variance = 90.74, and perceiver \times target variance = 52.22. The resulting proportions are error variance = .578, perceiver variance = .039, target variance = .242 and perceiver \times target variance = .140.

the estimated proportions of variance were .205 for the target variance and .795 for the error variance.

The mean variance estimates in the Hancock and Ickes (1996) study do not tell the whole story, however. The correlation between the perceiver \times target variance and either the perceiver variance or the target variance were -.64 and -.46, respectively. Basically, when the perceiver \times target variance was greater than 0 for a given group, the perceiver variance or the target variance (or both) tended to be 0, and when the interaction variance was equal to 0, perceiver or target variance tended to be positive. About half the time, one of the two targets in a group was easier to rate; and about one-third of the time, one of the two perceivers was more accurate. About one-third of the time there was a substantial target-by-perceiver interaction such that friends were rated more accurately than were strangers, in which case perceiver or target differences tended to be relatively small. The reasons why perceiver/target interactions appeared in some groups but not in others are not clear, and this question awaits future research for clarification.

Discussion

To our knowledge, Study 1 was the first to examine the sources of variation in empathic accuracy scores using the logic of the Social Relations Model (Kenny, 1988; 1994; Kenny & Albright, 1987; Malloy & Kenny, 1986). With regard to the amount of perceiver variance, the results we obtained were quite variable across the five data sets we analyzed. There was no reliable perceiver variance in the two studies in which (a) perceivers inferred the thoughts and feelings of only two or three target individuals and (b) the sets of targets who were judged were different for each set of perceivers (Hancock & Ickes, 1996; Buysse & Ickes, 1999). In contrast, the amount of perceiver variance was both substantial and reliable in the three studies in which many perceivers inferred the thoughts and feel-

ings of the same set of multiple target individuals.

These findings may have important implications for our understanding of perceiver effects in empathic accuracy and the conditions in which they are most likely to be observed. Apparently, perceiver variance is dramatically curtailed in designs in which different perceivers try to infer the thoughts and feelings of different targets, and the perceivers and the targets are nested within small groups. Our more detailed analysis of the Hancock and Ickes (1996) data suggested that variations in the degree of the perceivers' acquaintance with the targets might play a role in this outcome, such that when the perceiver \times target variance is greater than 0 for a given group, the perceiver variance or the target variance (or both) tends to be 0. In contrast, because the cross-group data indicate substantial target variance, at the level of the cross-group data it may be primarily the perceiver variance that suffers in the trade-off with the perceiver \times target interaction variance in studies of this type.

These data suggest that the amount of perceiver variance might be affected by two potential moderating factors—the “confounding” of unique perceivers and unique targets who are nested within small groups, and the trade-off in the proportions of interaction variance and perceiver variance in the cross-group data for acquainted and unacquainted perceiver-target pairs. Unfortunately, however, the present data permit us to do no more than speculate regarding the potential importance of one or both of these factors. We should also emphasize that these two factors by no means exhaust the set of possible factors that might account for the cross-study differences in the amount of perceiver variance in empathic accuracy scores. It is not yet clear what specific features of the designs and data-analytic procedures are most responsible for these differences, either individually or in combination. Providing anything like a definitive answer to this question is likely to require a statistical/conceptual analysis of how perceiver variance estimates are influ-

enced by averaging across small nested subgroups of unique perceivers and targets versus deriving a single estimate from a large group of perceivers who all rate the same targets. It may also require Monte Carlo simulations of data representing these different types of designs, and meta-analyses of actual empathic accuracy data from a much larger number of studies than we were able to examine here.

In terms of the goals of the present study, however, the important point is not why the amount of perceiver variance differs so greatly in the two types of studies. Rather, the important point is that such obvious differences in the amount of perceiver variance might be useful in helping us understand why it has been so difficult to identify reliable and replicable individual-difference correlates of empathic accuracy. If perceiver variance is negligible in studies in which small subgroups of unique perceivers rate unique targets, but is substantial in studies in which a large group of perceivers rate a set of common targets, the important question then becomes: Can reliable and replicable individual-difference correlates of empathic accuracy be found in the second type of study? It was to help answer this important question that Study 2—the Pham and Rivers (1998) study—was conducted.

Study 2

The Pham and Rivers (1998) study was conducted to determine whether reliable individual-difference correlates of empathic accuracy would be found if the conditions for detecting them were optimal, in terms of the best information that was currently available. Specifically, Pham and Rivers sought to predict empathic accuracy scores using the “best candidate” individual difference measures (those identified in Davis and Kraus's 1997 meta-analysis) in the type of design in which perceiver variance appeared to be the greatest (that originated by Marangoni et al., 1995).

Because Davis and Kraus's (1997) meta-analysis had identified a number of individ-

ual-difference constructs that were reliably associated with interpersonal sensitivity measures in general, Pham and Rivers used Davis and Kraus's findings to select some best-candidate measures for predicting empathic accuracy scores in their research. Specifically, Pham and Rivers included in their study measures of intelligence, dogmatism, locus of control, and interpersonal trust—constructs that had all been related to interpersonal sensitivity in Davis and Kraus's (1997) meta-analytic review. After pretesting each of the perceivers in their study with paper-and-pencil tests that assessed these four constructs, Pham and Rivers asked each perceiver to view highly edited versions of the three client–therapist videotapes that were originally used in the Marangoni et al. (1995) study and attempt to infer the content of each client's specific thoughts and feelings. Correlations were later computed to determine whether any of the four individual-difference variables predicted participants' global empathic accuracy scores in this study, in which a substantial and reliable amount of perceiver variance was clearly evident (see Table 1, fifth row).

Method

Participants and stimulus materials

The participants were 74 (29 male and 45 female) undergraduates who participated in the study for credit in their introductory psychology course. (The data from two men and four women were later dropped from the SRM analyses previously reported in Study 1 because they did not view all three stimulus tapes.)

The three videotapes originally used in the Marangoni et al. (1995) study were further edited to reduce their presentation time. Specifically, for each of the 30 "stop points" at which each client had reported having had a specific thought or feeling, only the 15 seconds of the videotaped client–therapist interaction immediately preceding that point were recorded onto the stimulus videotapes. Hence, instead of

viewing an edited 30-minute psychotherapy session, the participants in Study 2 viewed thirty 15-second excerpts from the session. For each tape, the 30 excerpts were shown in their original, chronological order.

Setting and procedure

An experimenter met each participant in a waiting room and escorted him or her to a cubicle containing a 25-inch color television monitor, a two-way intercom speaker, a signal button, and a remote control that was connected to a videocassette recorder (VCR) located in the adjacent control room (see Ickes et al., 1990, for a diagram of the laboratory arrangement).

Each participant was informed that the study would be conducted in two parts. The first part involved completing a set of paper-and-pencil questionnaires, whereas the second part involved viewing and making judgments about three videotaped interactions. The participant was then given an informed-consent form to read and voluntarily sign while the experimenter went to get the questionnaire measures. (All participants agreed to sign the consent form.) The experimenter then returned to administer the set of questionnaires, which included Shipley's (1940) Institute of Living Scale (to assess verbal IQ), Rokeach's (1956) Dogmatism Scale, Rotter's (1966) Internal-External Locus of Control Scale, and Rotter's (1967, 1971) Interpersonal Trust Scale.

The Shipley Institute of Living Scale (SILS; Shipley, 1940) is a measure of verbal intelligence. The SILS was highly correlated (.83) with the Kaufman Brief Intelligence Test in a study by Bowers and Pantle (1998), and has also been found to be a good predictor of the revised Wechsler Adult Intelligence Scale (Pringle & Hanstad, 1971; Wood, Conn, & Harrison, 1977). The Rokeach Dogmatism Scale, a measure of open versus closed belief systems, has been shown to differentiate members of politically right and left groups from members of the political center (Rokeach, 1956; Di Renzo, 1967). Rotter's locus of control

measure has been extensively validated (for reviews, see Lefcourt, 1984, and Phares, 1975), and several studies support the construct validity of his interpersonal trust measure (Boroto, 1970; Geller, 1968; Hamsher, Geller, & Rotter, 1968; Roberts, 1967). In the present sample, coefficient alpha for these measures ranged from .76 to .84.

The participant was instructed to complete these measures in the order given, and was then left alone to complete them in private. A signal button mounted on the wall enabled the participant to notify the experimenter when he or she was finished with the questionnaire portion of the study. At that time, the experimenter returned to the cubicle and collected the questionnaire materials.

The experimenter then gave the participant a supply of standardized thought/feeling inference forms (see Ickes et al., 1990, p. 29) and explained that the participant would now watch an instructional videotape before completing the second part of the study. The instructional videotape reviewed the procedure to be followed during the empathic inference task and gave the participant an opportunity to signal the experimenter in the event that it was necessary to clarify these instructions. When the participant was ready to proceed with the task, the experimenter started the first stimulus tape, pausing it at the "tape stop" that came at the end of each 15-second excerpt.

Each time the videotape was paused by

the experimenter, the participant recorded on the empathic inference form (a) whether the client appeared to be having a thought or a feeling, and (b) the specific inferred content of that thought or feeling, written in sentence form in the space provided. When the participant had finished recording the required information for a given thought/feeling entry, he or she used the remote control to restart the videotape. The videotape then continued until the experimenter paused it at the next stop point, when the participant inferred the client's thought or feeling at that point in the interaction. This procedure was repeated until the participant had inferred all 30 of the thoughts and feelings reported by the clients in each of the three tapes, and had therefore made a total of 90 thought/feeling inferences.

Results

The results of Study 2 are presented in Table 2. They reveal that the dogmatism and locus-of-control measures did not predict the participants' empathic accuracy scores, whereas the verbal intelligence and interpersonal trust measures did predict empathic accuracy—but in ways not entirely consistent with the results of Davis and Kraus's (1997) meta-analysis.

First, a hierarchical moderated multiple regression analysis revealed that the relationship between verbal intelligence and empathic accuracy was significantly moderated by gender, $F(1,63) = 5.57, p < .025$. The

Table 2. Correlates of empathic accuracy in the Pham and Rivers study (Study 2)

Individual difference measure	Global empathic accuracy measure		
	Men	Women	All participants
Verbal intelligence (Shipley, 1940)	.35	-.23	.02
Dogmatism (Rokeach, 1956)	-.12	.01	-.05
Locus of control (Rotter, 1966)	.13	-.09	.00
Interpersonal trust (Rotter, 1967)	-.30	-.21	-.25*

Note: For the men, $n = 29$; for the women, $n = 45$; for all participants, $N = 74$. The actual degrees of freedom varied slightly across the four predictors because of occasional missing data.
 $p < .075$. * $p < .05$.

relationship was positive and approached significance for the men in the sample ($r = .35, p < .075$), but was negative and nonsignificant for the women ($r = -.23, ns$). Second, a significant correlation was found between the participants' interpersonal trust scores and their global empathic accuracy scores ($r = -.25, p < .05$). However, this correlation is opposite in sign to the positive correlation between interpersonal trust and interpersonal sensitivity that Davis and Kraus (1997) identified in their meta-analytic study.

Discussion

The results of Study 2 further emphasize the difficulty of identifying reliable and replicable individual-difference correlates of empathic accuracy scores. Seeking to maximize their chances of identifying such correlates, Pham and Rivers (1998) used what Davis and Kraus's (1997) meta-analysis suggested were the optimal individual-difference constructs as their predictor variables. They then applied these variables in a research design in which the perceiver variance in empathic accuracy scores was substantial and significant—accounting for over half of the total variance of the global empathic accuracy measure (see Table 1, fifth row).

Despite these efforts, however, results revealed that only two of the four predictor variables—verbal intelligence and interpersonal trust—were significantly associated with the empathic accuracy measure, and in ways not entirely consistent with Davis and Kraus's (1997) meta-analytic review. The predicted positive relationship between verbal intelligence and empathic accuracy was evident for the men but not for the women in the study, and the negative relationship between interpersonal trust and empathic accuracy was opposite in sign from the positive relationship expected on the basis of Davis and Kraus's (1997) review.

Until these findings are replicated, we should probably resist any temptation to try to interpret them or attempt to reconcile them with the findings reported by Davis

and Kraus (1997). The safest conclusion to draw at this point is that a reliable and replicable individual-difference correlate of empathic accuracy has still not been identified. The most promising candidate—in terms of both the Study 2 findings and the Davis and Kraus (1997) findings—is verbal intelligence, but there is no obvious reason why this variable should differentially predict the empathic accuracy of male, versus female, perceivers. Similarly, although the unanticipated negative correlation between interpersonal trust and empathic accuracy is intriguing and deserving of further study, this relationship would also have to survive the test of replicability before any strong claims could be made, particularly given the opposite finding reported by Davis and Kraus (1997).

On a more positive note, the results of Study 2 provide at least some evidence that individual difference correlates of empathic accuracy can be identified in a study in which the amount of perceiver variance is substantial. If future researchers want to maximize their chances of either replicating these relationships or discovering additional ones, the present findings may be useful in suggesting the type of research design they should use and the types of individual-difference variables they should assess.

General Discussion

The present studies contribute important pieces of information to the puzzle of why it has been so difficult to find reliable individual-difference correlates of empathic accuracy. The results of Study 1 suggest that, to identify such correlates, researchers must use a research design in which the amount of perceiver variance in empathic accuracy scores is substantial enough for individual-difference correlates to matter. The optimal research design for this purpose appears to be one in which a relatively large set of perceivers individually infer the thoughts and feelings of the same, standardized set of target persons.

Given the results of Study 1, Pham and Rivers conducted Study 2 to see whether

reliable individual-difference correlates of empathic accuracy could be found in a study using what were presumably an optimal research design and an optimal set of predictor variables. Perceiver variance was substantial in this study, accounting for over half of the variance in empathic accuracy scores, and two of the four predictor variables—verbal intelligence and interpersonal trust—were found to be related to empathic accuracy. However, verbal intelligence was positively correlated with the men's, but not the women's, empathic accuracy scores. And contrary to the findings of Davis and Kraus's (1997) meta-analysis, interpersonal trust was negatively, rather than positively, correlated with empathic accuracy.

The present findings may have implications for the larger study of interpersonal accuracy. From a methodological standpoint, we can now question the wisdom of using face-valid, self-report measures of empathic skills and dispositions to predict interpersonal accuracy in research designs in which small subsets of perceivers are nested with similarly small subsets of unique targets. Viewed in conjunction with the results of Davis and Kraus's (1997) meta-analytic review, the findings of the present Study 1 suggest that studies of this type are particularly likely to yield null results. It is ironic, therefore, that relationship researchers who are interested in interpersonal accuracy have traditionally favored such studies, using self-report empathy/sensitivity measures to try to predict interpersonal accuracy/sensitivity in samples composed of unique or nested dyads such as friends, roommates, dating partners, or married couples. We suspect that researchers who want to account for individual differences in interpersonal sensitivity will have to find a better research paradigm to use, with the best available candidate being a paradigm in which a relatively large group of individual perceivers judge the same standardized set of target persons.

Conversely, because other designs might be appropriate for other research purposes, it may be useful to think about different

kinds of perceiver/target interactions. One extreme would be a case where a large number of raters judge a small number of targets, as when people make judgments about candidates in an election. (The Marangoni et al., 1995, study approximates this situation.) In such a case, the candidates might be rather similar, but there could be substantial disagreement among raters about them, so perceiver variance would be large compared to target variance, with little perceiver \times target interaction variance.

At another extreme, a small number of judges rate a large number of targets (as when judges rate divers or figure skaters at an athletic competition). Here, perceiver variance should be smaller than target variance, and, again, there should be little interaction variance. Somewhere between, when a small number of raters judge a small number of targets, and there are individual differences in judge/target relationships, interaction variance among judge/target sets would be expected to become more important than inter-perceiver or inter-target differences. (Another middle ground, in which a large number of perceivers rate a large number of targets has been little explored, and predictions are difficult to make.)

Prior to this study, perhaps the most plausible explanation for the lack of reliable individual-difference correlates of empathic accuracy was the "lack of metaknowledge" interpretation that was first proposed by Ickes (1993) and later elaborated by Mortimer (1996) and Davis and Kraus (1997). The findings from the present Study 2 are not incompatible with the "lack of metaknowledge" hypothesis, which has received at least a modest degree of preliminary empirical support (Marangoni et al., 1995; Mortimer, 1996). It is possible, for example, that the perceivers' lack of metaknowledge affects the predictive validity of face-valid, self-report measures of empathy/sensitivity even in research designs in which there is a substantial amount of perceiver variance. If so, the specific lack of predictive validity for such measures could account for the data pattern identified by Davis and Kraus (1997), in which

self-report measures of empathically relevant traits and dispositions consistently fail to predict performance on interpersonal sensitivity tasks, whereas self-report measures of some seemingly less-relevant constructs do a better job in this regard.

In the final analysis, which constructs are most likely to successfully differentiate "good" versus "poor" perceivers? Although it is still too early to say for sure, verbal intelligence may be one of the strongest candidates to date (see also Hughes & Cutting, in press). Interpersonal trust is another con-

struct that warrants further investigation, despite the conflicting findings of Davis and Kraus (1997) and the present Study 2. As for additional candidates, Vogt and Colvin (1998) have reported evidence suggesting that a communal interpersonal orientation may be positively associated with interpersonal accuracy, whereas Thomas (1998) has made a similar claim for the construct of attributional complexity. It remains to be seen, however, whether any of these constructs will survive the critical test of cross-study replicability.

References

- Boroto, D.R. (1970). *The Mosher forced-choice inventory as a predictor of resistance to temptation*. Unpublished master's thesis, University of Connecticut, Storrs.
- Bowers, T.L., & Pantle, M.I. (1998). Shipley Institute for Living Scale and the Kaufman Brief Intelligence Test as screening instruments for intelligence. *Assessment*, 5, 187-195.
- Buysse, A., & Ickes, W. (1999). Topic-relevant cognition and empathic accuracy in laboratory discussions of safer sex. *Psychology and Health*, 14, 351-366.
- Davis, M.H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, 51, 167-184.
- Davis, M.H., & Kraus, L.A. (1997). Personality and empathic accuracy. In W. Ickes (Ed.), *Empathic accuracy* (pp. 144-168). New York: Guilford Press.
- Di Renzo, G.J. (1967). *Personality, power, and politics*. Notre Dame, IN: University of Notre Dame Press.
- Funder, D.C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, 102, 650-672.
- Geller, J.D. (1968). Some personal and situational determinants of interpersonal trust. *Dissertation Abstracts*, 28 (11-B), 4755.
- Gesn, P.R. (1997). *The development of meaning contexts for empathic accuracy: Channel and sequence effects*. Unpublished doctoral dissertation, University of Texas at Arlington.
- Graham, T.M. (1996). *Effects of feedback and frames on empathic accuracy*. Unpublished doctoral dissertation, University of Texas at Arlington.
- Hamsher, J.H., Geller, J.D., & Rotter, J.B. (1968). Interpersonal trust, internal-external control, and the Warren Commission Report. *Journal of Personality and Social Psychology*, 9, 210-215.
- Hancock, M., & Ickes, W. (1996). Empathic accuracy: When does the perceiver-target relationship make a difference? *Journal of Social and Personal Relationships*, 13, 179-199.
- Hughes, C., & Cutting, A.L. (in press). Nature, nurture and individual differences in early understanding of mind. *Psychological Science*.
- Ickes, W. (1993). Empathic accuracy. *Journal of Personality*, 61, 587-610.
- Ickes, W., Bissonnette, V., Garcia, S., & Stinson, L. (1990). Implementing and using the dyadic interaction paradigm. In C. Hendrick & M. Clark (Eds.), *Review of personality and social psychology: Volume 11, research methods in personality and social psychology* (pp. 16-44). Newbury Park, CA: Sage.
- Ickes, W., Hancock, M., Graham, T., Gesn, P.R., & Mortimer, D.C. (1994). *(Nonsignificant) individual-difference correlates of perceivers' empathic accuracy scores*. Unpublished data, University of Texas at Arlington.
- 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.
- Kelleher, J. (1998). *The effects of frames of reference on empathic accuracy*. Unpublished master's thesis, University of Texas at Arlington.
- Kenny, D.A. (1988). Interpersonal perception: A social relations analysis. *Journal of Social and Personal Relationships*, 5, 247-261.
- Kenny, D.A. (1994). *Interpersonal perception: A social relations analysis*. New York: Guilford Press.
- Kenny, D.A., & Albright, L. (1987). Accuracy in interpersonal perception: A social relations analysis. *Psychological Bulletin*, 102, 390-402.
- Lefcourt, H.M. (Ed.). (1984). *Research with the locus of control construct*. (Vol. 3). New York: Academic Press.
- Malloy, T.E., & Kenny, D.A. (1986). The Social Relations Model: An integrative model for personality research. *Journal of Personality*, 54, 199-225.
- 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.
- Mortimer, D.C. (1996). *"Reading" ourselves "reading" others: Actual versus self-estimated empathic accuracy*. Unpublished master's thesis, University of Texas at Arlington.
- Phares, E.J. (1975). Internal-external control, interpersonal judgments of others in need, and attribution of responsibility. *Journal of Personality*, 43, 23-28.

- Pringle, R.K., & Hanstad, M. (1971). Estimating WAIS IQ's from progressive matrices and Shipley-Hartford scores. *Journal of Clinical Psychology, 27*, 479-481.
- Roberts, M.D. (1967). *The persistence of interpersonal trust*. Unpublished master's thesis, University of Connecticut, Storrs.
- Rokeach, M. (1956). Political and religious dogmatism: An alternative to the authoritarian personality. *Psychological Monographs, 70*, No. 425.
- Rotter, J.B. (1966). Generalized expectancies for internal vs. external control of reinforcement. *Psychological Monographs: General and Applied, 80*, 1-28.
- Rotter, J.B. (1967). A new scale for the measurement of interpersonal trust. *Journal of Personality, 35*, 651-665.
- Rotter, J.B. (1971). Generalized expectancies for interpersonal trust. *American Psychologist, 26*, 443-452.
- Searle, S.R. (1971). *Linear models*. New York: Wiley.
- Shipley, W.C. (1940). A self-administering scale for measuring intellectual impairment and deterioration. *Journal of Psychology, 9*, 371-377.
- Taft, R. (1955). The ability to judge people. *Psychological Bulletin, 52*, 1-23.
- Thomas, G. (1998). *Empathic accuracy in close relationships: The role of the perceiver, the target, and the perceiver-target relationship*. Abstracts of the 27th Annual Meeting of the Society of Australasian Social Psychologists. Christchurch, New Zealand, April 16-19th.
- Vogt, D.S., & Colvin, C.R. (2000). The good judge of personality: Theory, correlates, and Cronbachian "artifacts." Unpub. man.
- Wood, R.W., Conn, D.C., & Harrison, H.D. (1977). The efficacy of using Shipley-Hartford cores to predict WAIS IQ in a penal population. *Quarterly Journal of Corrections, 1*, 39-41.

(continued from inside back cover)

Chapter in an Edited Book

Parke, R.D. & Tinsley, B. J. (1987). Family interaction in infancy. In J.D. Osofsky (Ed.), *Handbook of Infant Development* (pp. 579-641). New York: John Wiley & Sons Inc. *Appendix* (optional). Use only if needed. May be usefully included for review, even if not appropriate for publication.

Tables. Tables must appear as a set following the references. Each table should be typed double-spaced on a separate sheet, numbered consecutively with an Arabic numeral and given a short title (e.g. Table 3. Marital outcomes for premarital cohabitators and noncohabitators.). All tables must be cited in the text.

Figures. Figures must appear as a set following the tables. Each figure must be numbered consecutively with an Arabic numeral and must include a descriptive legend and a title. Titles must be typed together, double-spaced, on a separate sheet preceding the artwork (e.g. Figure 1. Knowledge structures and the development of relationships.). Figures must be supplied no larger than 8" by 10", be black and white and be ready for photographic reproduction. Diagrams must be professionally rendered or computer generated. All labels and details must be clearly printed and large enough to remain legible at a 50% reduction. Artwork should be identified by figure number and be carefully packaged in a protective envelope. All figures must be cited in the text.

Copyright and Originality

It is a condition of publication that all manuscripts submitted to *Personal Relationships* have not been pub-

lished and will not be simultaneously submitted or published elsewhere. All authors must sign the Transfer of Copyright Agreement, available from the publisher, before an article can be published. Government authors whose articles were created in the course of their employment must certify to that fact in lieu of copyright transfer. Authors are responsible for obtaining written permission from the copyright owners to reprint any previously published material included in their article.

Copyediting and Page Proofs

The publisher reserves the right to copyedit manuscripts to conform to journal style and the copyedited version of the paper will be returned to the lead author for comment. The lead author will also receive page proofs for correction of typographical errors only. No rewriting of the original manuscript is allowed in the page proof stage. Authors must return proofs to CUP within 48 hours of receipt, or approval will be assumed.

Offprints

The lead author will receive 25 free article offprints of his or her article. A form accompanying page proofs allows the lead author to order complete copies of the issue and/or purchase additional offprints. All coauthor offprint requirements must be included on this form. Orders received after the issue is printed are subject to a 50% reprint surcharge.