Bringing Everyday Mind Reading Into Everyday Life: Assessing Empathic Accuracy With Daily Diary Data

Maryhope Howland[1] and Eshkol Rafaeli[2]

[1]University of Minnesota
[2]Barnard College, Columbia University

ABSTRACT Individual differences in empathic accuracy (EA) can be assessed using daily diary methods as a complement to more commonly used lab-based behavioral observations. Using electronic dyadic diaries, we distinguished among elements of EA (i.e., accuracy in levels, scatter, and pattern, regarding both positive and negative moods) and examined them as phenomena at both the day and the person level. In a 3-week diary study of cohabiting partners, we found support for differentiating these elements. The proposed indices reflect differing aspects of accuracy, with considerable similarity among same-valenced accuracy indices. Overall there was greater accuracy regarding negative target moods than positive target moods. These methods and findings take the phenomenon of "everyday mindreading" (Ickes, 2003) into everyday life. We conclude by discussing empathic *accuracies* as a family of capacities for, or tendencies toward, accurate interpersonal sensitivity. Members of this family may have distinct associations with the perceiver's, target's, and relationship's well-being.

Interpersonal perception serves as the backbone of any social interaction, from detecting another's mere presence to appreciating his or her innermost thoughts and feelings. The degree of accuracy in perceptions that focus specifically on transitory internal states (i.e.,

thoughts, feelings, motives) within another's mind has been referred to as *empathic accuracy* (Ickes, Stinson, Bissonnette, & Garcia, 1990), or more poetically as "everyday mindreading" (EA; Ickes, 2003). This process may be particularly important within close committed relationships (e.g., Kilpatrick, Bissonnette, & Rusbult, 2002).

Imagine a hypothetical couple, Anne and John. If Anne is entirely tone-deaf to changes in John's positive moods, would she also be tone-deaf to changes in the negative ones? If John overestimates Anne's sadness, is he likely to overestimate her other moods? Perhaps when Anne is mildly upset, John perceives her as severely depressed. And would we expect Anne (or anyone else) to be especially adept at noticing some of John's moods—anger? anxiety?—but not others?

We easily imagine the many processes that may be affected by such accuracy or inaccuracy (and we will return to the costs and benefits of EA later). For now, we offer these examples to point to the many different elements that Anne's and John's empathic accuracies comprise (e.g., how far off they may be, how consistent or inconsistent they are over time, and how their accuracy about one mood may or may not relate to their accuracy about another).

We see empathic accuracy as a multifaceted construct that can involve accuracy toward various target variables (i.e., cognitive and affective, positive and negative, intrarelational and extrarelational) over different time spans and in different environments. EA is multifaceted also because it can be (and has been) operationalized in a variety of ways, each speaking to a different aspect of accuracy. When assessed using cross-sectional designs (e.g., Kenny & Acitelli, 2001), it can be thought of as a normative process (in that particular case, a process involving a balance of both bias and accuracy). When assessed using a 10- or 20-minute videotaped dyadic interaction (e.g., Ickes, Stinson, Bissonnette, & Garcia, 1990; Thomas & Fletcher, 2003), it provides an individual difference index of "micro" accuracy, perhaps akin to the *capacity* for accurate inference. In this article, using dyadic diary data, we contribute to a third, more "macro" approach to the study of accuracy.

Assessing EA

Previous research operationalized EA in three ways. The first and oldest of these methods consists of cross-sectional or panel ques-

tionnaire methods, comparing raters' judgments of other persons' traits with those persons' trait self-reports. This approach characterized the early person perception research, reviewed quite critically by Cronbach (1955; cf. Biesanz, West, & Millevoi, 2007). Cronbach criticized the use of one-time measurements of accuracy that relied heavily on simple difference scores, scores that combine components of accuracy and may result in statistical artifacts. He argued for the need to distinguish among separate elements of accuracy, each tapping a different component. At the time of Cronbach's critique, methodological obstacles impeded progress in the area. More recently, with increasingly advanced and appropriate statistical analyses, cross-sectional methods for the study of accuracy have reemerged. For example, Pietromonaco, Rook, and Lewis (1992) investigated accuracy among strangers or friends immediately following a conversation with each other. Their participants completed identical measures for themselves and the target. Accuracy was calculated as both the average absolute differences and as ''shape,'' a correlation measure describing perceivers' ability to detect certain patterns of moods in a target; bias was defined as the signed difference between the mean of a perceiver's estimates about the target and the mean of the target's self-reports. More recently, Kenny and Acitelli (2001) conducted individual interviews with both members of couples, inquiring about their relationship, their jobs, and their closeness to family. The participants also answered the same questions from their partners' perspectives. To distinguish accuracy from assumed similarity, structural equation models of the associations between both partners' feelings and empathic perceptions were used to simultaneously estimate path coefficients.

Cross-sectional data of this sort are less costly and quite convenient to collect, and advancements in statistics (e.g., Biesanz et al., 2007; Kenny, Kashy, & Cook, 2006) have made it possible to mine such data more fully. As Kenny and Acitelli (2001) show, this method allows one to examine accuracy toward several target variables (e.g., accuracy regarding relational closeness vs. job satisfaction). However, it involves assessment at a single time point and uses retrospection or generalization about self or other. As work by Robinson and Clore (2002) indicates, questions about transient states that refer to periods longer than 24–48 hours tend to elicit responses that rely primarily on trait knowledge or long-established schemas instead of on actual retrospection or on some sort of

"moving average" regarding real experiences. As a consequence, they have limited utility for researchers interested in accuracy toward more fleeting states such as moods.

In the early 1990s, a methodological innovation in the form of videotaped dyadic interactions helped energize and shape the study of EA into what it is today. The Unstructured Dyadic Interaction Paradigm, first developed by Ickes and colleagues (e.g., Ickes, 1993), built on previous videotape methodologies (Gottman & Porterfield, 1981), narrowed the definition of EA, and provided some distinct guidelines for measurement and treatment of data (for a thorough description of the paradigm, see Ickes, 2003, pp. 59–83). This method situates accuracy at an online or *micro* level and asks participants to consider their thoughts and feelings and those of their partner in a moment-by-moment manner, as they review interactions that lasted several minutes. Accuracy is determined by raters who compare the targets' own thoughts and feelings with the perceivers' inferences about them. Two major advantages of this method are its highly detailed micro-level analysis and its use of specific (and often thematic) topical conversations as stimuli, rather than the reliance on schema-driven retrospection (as is the case in the cross-sectional questionnaire methods). One concern with this methodology is ecological validity: it relies on a nonspontaneous lab interaction, which takes place outside the context of the dyad's everyday life. Additionally, the thoughts and feelings toward which the perceivers are accurate (or not) are generated by the targets; as a consequence, they are not guaranteed to include any particular array of target variables. For some, they could be mostly thoughts, for others mostly feelings; for some all positive, for others negative.

A third and more recent approach to studying EA is the daily diary approach, used in the present study, which offers a complementary tack to the earlier two. Daily diaries can assess experience and perception in the context and privacy of individuals' everyday lives, thus addressing earlier concerns about the ecological validity of EA methods. By examining these perceptions repeatedly over the course of weeks, they provide a *macro*-level examination of these dynamics, complementary to the more *micro* lab approach. We use the terms *micro* and *macro* to denote the different scope of the targets of accuracy in these different approaches. For micro targets, the time span is short, and the details attended to are limited (e.g.,

a facial expression, a speech turn in a conversation). For macro targets, the time span is longer (1 day in our case), and the scope of material is consequently larger and more global.

The ability to examine perceivers over a greater span of time, as the diary method allows us to do, offers a glimpse into whether they are likely to be accurate in everyday life—that is, whether they tend to be accurate. This is another way in which the macro approach differs from the micro approach: rather than examining the *capacity* for accuracy in a very well-defined setting, it indexes the *tendency* toward accuracy or inaccuracy in daily life.

Additionally, diaries elicit a predetermined set of observations from each perceiver and target at each time point. This allows investigating a wide array of target variables and examining differential EA regarding these targets. In this way, they respond to the appeal made by Cronbach (1955), who noted that where possible, accuracy should be estimated separately for different qualities of the target. Finally, because diary methods generate multilevel data, they lend themselves to the kind of dismantling analysis called for by Cronbach's central critique.

Wilhelm and Perrez (2004) were the first to introduce diaries into the study of EA and to highlight the benefits of the daily diary method for this purpose. These authors focused primarily on diaries' ability to bring the research out of the laboratory and into the participants' real lives. They examined partners' mutual perception about several emotional dimensions (e.g., sad–happy, tense–relaxed, unsatisfied–satisfied) when in (or out) of one another's presence (e.g., at home vs. at work), and at several times of day—an analysis that could only be conducted with naturalistic experience sampling. Accuracy was calculated as the product-moment correlations between individuals' estimations of their partners' moods and the targets' own mood reports; additional analyses included both accuracy and assumed similarity within a structural equation model. Their primary findings (to which we return later) were of gender differences and contextual effects in accuracy.

Wilhelm and Perrez (2004) begin to show the richness of dyadic data collected with the daily diary method. Here, we extend their work and propose ways in which additional information can be gleaned from such data. Specifically, with nested data, accuracy can be thought of as a characteristic of the day (or moment) or of the individual. At the daily level, accuracy can be thought of as

the minimization of difference between a perceiver's estimate and the target's actual mood—which we term *discrepancy,* calculated as the absolute difference between these scores.

Beyond the analysis of day-level accuracy, we can describe *perceivers* who tend to be more or less accurate in the aggregate *person* level. Perceivers may vary in their accuracy regarding any of several mathematically distinct components: the level, scatter, or pattern of the target's mood. *Level accuracy* can be indexed by the mean of the daily discrepancies. In contrast, accuracies regarding both scatter and pattern are properties that emerge most easily[1] at the person level. *Scatter accuracy* reflects the extent to which the variability in the perceiver's estimates of a given target mood approximates the variability in the target's actual moods, or the degree to which a partner picks up on the extent to which a given mood is fluctuating over time. *Pattern accuracy* reflects the correspondence between the temporal patterns of the target's mood ratings and the perceiver's estimates regarding that mood—that is, the within-dyad correlation between the target's actual mood on each day and the perceiver's corresponding estimation of the target's mood on those days.

Assessing accuracy at both day and person levels, and operationalizing it in these diverse ways, allows an examination of the costly or beneficial role of EA at both the momentary and the aggregated levels. Doing so is important because the findings of existing research on the consequences of empathic accuracy for individuals or relationships have been mixed. Though many studies support the general finding that accuracy helps relationships (e.g., Kilpatrick et al., 2002), a prominent model (cf. Ickes & Simpson, 2004) emphasizes the existence of particular situations and contexts that

---

1. The current study uses a daily process method, assessing moods once daily, in the evening. However, if multiple measures were collected within each day (using ecological momentary assessment, for instance), it would have been possible to calculate within-day accuracy indices. Still, to obtain scatter or pattern indices within a day, one would need to assess both partners simultaneously and repeatedly each day. Apart from the challenge (both technical and conceptual) of obtaining simultaneous responses from both partners, there is a clear burden trade-off between the number of within-day measurements and the overall duration during which participants complete diaries; at its extreme, an intensive within-day approach will begin resembling what we call the micro approach used in lab-based research.

favor motivated *in*accuracy. We return to these mixed findings in our discussion but wish to point out now that a better understanding of *when* accuracy would or would not be useful may require attending to *which* accuracy one is considering (i.e., micro vs. macro, daily vs. person level, level vs. scatter vs. pattern). It would also require specifying which target moods are attended to—that is, *what* one is accurate about.

Recognizing the variety of accuracy types and applying them to a variety of targets generates a set of EA variables for each person. We expect that these empathic accuracies (plural), like moods or emotions themselves, are not all created equal. For one, we expect that the empathy process—which involves emotional awareness at the *interpersonal* level—shares some features with emotional awareness processes in *intrapersonal* mood regulation. Our research on emotional self-awareness or differentiation (Gilboa-Schechtman, Rafaeli, Granat, Kuint, & Feldman, 2009) suggests that (self) awareness of positive emotions is unrelated to that of negative emotions. Moreover, our work (Rafaeli, Mor, Drejet, Ehrlich, & Bodkin, 2009) suggests that awareness of negative emotions has consequences that are different from awareness of positive emotions, the former being adaptive, the latter much less so.

This similarity between intrapersonal emotional awareness and interpersonal empathic accuracy provides some of the motivation for exploring different facets of EA (e.g., accuracy regarding positive and negative moods). Thus, if EA is indeed the relational or interpersonal counterpart to emotional self-awareness and differentiation, we may hypothesize that EA regarding negative target moods would be directly associated with adaptive relationship outcome. In contrast, we may hypothesize that EA regarding positive target moods would have weaker and possibly even inverse associations with these adaptive relationship outcomes. For example, if the target is anxious (has a negative mood), the perceiver may benefit from accurately recognizing the mood as anxiety rather than misreading it as anger or sadness and then incorrectly responding to it. However, if the target is happy (e.g., as a result of something at work), there seems to be less harm should the positive mood be mislabeled—after all, positive moods rarely lead to any narrow action tendencies (Fredrickson, 1998). There may even be some possible bonus should other positive emotions (such as love or excitement) be incorrectly assumed.

Some evidence for a distinction between accuracies toward targets of different valence comes from work by Gaelick, Bodenhausen, and Wyer (1985), who asked couples to discuss a problem in their relationship and found that only negative emotions were perceived accurately, whereas positive and neutral emotions were inaccurately perceived. Similarly, Gable, Reis, and Downey (2003), who used a quasi signal-detection analysis to estimate accuracy of perceptions of positive and negative behaviors among partners in a diary study, found inaccuracy (defined as either ''misses'' or ''false alarms'') to be more than twice as likely for positive than for negative events.

The Present Study

Most of what is currently known about EA (Ickes, 1997, 2003) emerges from studies using the micro level of analysis. The present work demonstrates several innovations that the macro (daily diary) design offers. First, the diary method allows us to examine accuracy about a variety of target variables in a quantifiable and standardized way. Second, the data resulting from this method can be examined at both momentary (day) and aggregated (person) levels, allowing us to differentiate between state versus trait accuracy. Third, it allows for distinguishing among accuracy regarding the level, the scatter, or the pattern of a target mood.

A fourth innovative feature of our study is the inclusion of both positive and negative moods as target variables. Though this feature need not be unique to daily diary data, it does correct an oversight of most studies to date, which tended to measure emotions as unidimensional (i.e., on bipolar scales, e.g., sad–happy; Wilhelm & Perez, 2004). Because positive and negative affect are not polar opposites (Larsen, McGraw, & Cacioppo, 2001; Rafaeli & Revelle, 2006), the current study examines various positive and negative moods as separate scales.

These innovations in the assessment of EA allow us to identify a set of psychologically distinct accuracy indices. The general hypothesis guiding our work is that these distinct indices of EA reflect differing aspects of the construct at both the day and the person levels. Specifically, at the day level, we expected that (1) accuracies for the various target moods, operationalized using discrepancy indices, would not be synonymous with each other (e.g., discrepancy

about vigor will not be strongly related to discrepancy about depressed mood, and so on), but that they will be associated with one another along valence lines (i.e., discrepancies regarding negative moods would be positively associated with each other, as would those regarding positive moods). Similarly, at the person level, we expected that (2) accuracy indices (for either level, scatter, or pattern) for the various target moods would not be synonymous with each other (e.g., scatter accuracy about vigor will not be strongly related to scatter accuracy about depressed mood). Instead, we again expect them to be associated with one another along valence lines (e.g., scatter accuracy indices regarding negative moods would be positively associated with each other, as would those regarding positive moods). We also expected that (3) person-level accuracy indices regarding positive moods would yield lower accuracy scores than those regarding negative moods. Finally, we predicted that (4) level, scatter, and pattern accuracy indices regarding the same moods would be independent from each other (e.g., level accuracy, scatter accuracy, and pattern accuracy about anger will be unrelated to each other).

METHOD

Participants

To recruit a diverse sample of community couples in committed relationships, flyers listing details about the study were posted in various locations around New York City. Couples responding to the flyers were screened over the phone for these eligibility criteria: cohabitation (for a minimum of 6 months), age (greater than 18), and availability to participate for the entire length of the study. Forty-one heterosexual couples and one same-sex couple met criteria and agreed to participate. The data of five individuals were lost due to equipment malfunction, resulting in a final sample of 37 couples with usable data.

Participants ranged in age from 19 to 51 ($M = 29$, $SD = 5.57$). Eleven percent were Latino, 13.4% were East Asian/Pacific Islander, 5% were other Asian, 6% were African American, 61% were White, and 3.6% described themselves as "other." Fifty-five percent of couples were married. Mean length of acquaintance was 61 months ($SD = 38.8$); mean length of romantic involvement was 48 months ($SD = 30.5$); mean length of cohabitation was 28 months ($SD = 23.1$). Couples were compensated a total of $90 for participation and were entered in a raffle to win $200.

Procedure

In the initial session, participants completed a background questionnaire and were introduced to the electronic diary and to the study's instructions. Every couple was told the following: "Before we get started on the actual diary, I just want to let you know that your responses are going to be kept anonymous and confidential. I also have to ask you not to discuss your responses with one another." During the training session, each participant was asked to practice responding on a training Palm Pilot, on which a sample diary was installed. This was done to illustrate that the responses, once entered, were irretrievable, both by the participant and by his or her partner. During the practice sessions, the experimenter reviewed the entire diary, item by item, explaining each style of question and allowing participants to ask any questions about word or item meanings. Participants were encouraged to repeat the diary as many times as they needed until they felt comfortable with the device and with the meanings of the items.

Because they were provided with two similar devices, we reminded them of their unique code (which was noted on the device), asked them to mark it with their first names (so they would not confuse the two devices), and asked them to each name a specific place where they would keep the diary, helping them remember to complete it and to keep it charged. Before leaving the lab, the participants were again reminded of the logistics and of the need to refrain from discussing their responses.

Participants all began the study on a Thursday and completed it 3 weeks later; during that time, each participant completed twice-daily questionnaires. Participants initiated the entries (i.e., they were not signaled by the device). Participants were instructed to complete the morning questionnaire within 1 hour of waking and the evening questionnaire within 1 hour of going to bed, times that were determined idiosyncratically by each participant's daily routines.

Participants could contact us anytime and were contacted by staff twice the 1st week and once weekly after that (with a minimum of two completed contacts) to ensure compliance, review the need to refrain from discussing their responses, make sure they were experiencing no technical difficulties, and remind them about their scheduled drop-off time. The daily diary was completed electronically on Palm Zire 21 PDA handheld devices (Palm Inc.). Each participant received a device personalized with his or her first name and labeled with a reminder to charge the device, and with the study coordinator's contact information. The devices were outfitted with the iESP diary program, first developed by Barrett and Barrett (2001) and later adapted by the Intel Corporation. All functions of the Plam Pilot apart from the diary program were inaccessible to participants. Additionally, neither the respondent nor his or her partner was

able to access the responses once entered. Upon conclusion of the 3-week diary period, participants returned the devices to the lab and completed a short follow-up questionnaire.

Measures

These data are taken from a larger study of relationship processes; only relevant measures (which are taken from the evening diary questionnaires) will be reported here. Other results from this study (pertaining to social support and hindrance and their effects on daily moods) have been reported elsewhere (Rafaeli, Cranford, Green, Shrout, & Bolger, 2008, Study 3).

Target Moods and Their perception (Inputs for EA)

Every evening, each participant was asked to estimate both his or her own and his or her partner's current mood; thus, every individual acted as both a perceiver and a target in the study. Mood was assessed using an adapted and shortened daily diary version (Cranford et al., 2006) of Lorr and McNair's Profile of Mood States (*POMS;* 1971). For space reasons, participants were presented with only 11 of the original POMS items, composing brief scales for the following moods: anger (angry), depression (hopeless, sad, blue, discouraged), anxiety (on edge, anxious, uneasy), and vigor (vigorous, cheerful, lively). Because of the preponderance of negative moods, a single item (self-sufficient) was added as an index of an additional positive feeling, efficacy. Items were rated on a 5-point scale, ranging from 1 (*not at all*) to 5 (*extremely*).

We calculated the between- and within-person reliabilities for the multi-item POMS scales using procedures outlined in Cranford et al. (2006), which require randomly selecting one partner from each couple and estimating the reliabilities without within-dyad dependencies. For a given measure, the between-subjects reliability coefficient (R_{1F} in Cranford et al.) is the expected between-subjects reliability estimate for one fixed day. The within-subjects reliability coefficient (R_C in Cranford et al.) is the expected within-subjects reliability of change of persons over days. Respectively, the between-person and within-person reliabilities were .85 and .83 for vigor, .77 and .66 for anxiety, and .67 and .81 for depression. Two of our moods (efficacy and anger) were single-item measures, for which the procedure outlined by Cranford et al. cannot be used. It is, however, possible to estimate the amount of reliable between-subject variance in these two moods. To obtain these, we conducted simple two-factor (Person $\times$ Day) ANOVA models for both moods. In efficacy, 67% of variance was due to person. In anger, on the other hand, only 26% of the variance was due to person, suggesting greater day-to-day variability.

RESULTS

Analyses were conducted only with data from days on which both partners in a couple had responded. Because diary entries were time-stamped, we were able to confirm that both partners' entries for a given day actually occurred on the relevant day. This allowed us to match entries by day and avoid the possibility of mismatching partners' entries in the event that one partner skipped a day. Of a total of 21 possible days, couples fully responded on average 18.5 days (median = 20 days). There were a total of 1,449 usable person-days (days where both partners had data).

We computed one accuracy index (discrepancy) for each possible day and three accuracy indices (level accuracy, scatter accuracy, and pattern accuracy) for each person. Descriptive statistics for the accuracy indices at both the day level and the person level are reported in Table 1.

Before considering the results, it is helpful to review how each of these indices was calculated. Day-level discrepancy (which is smaller when perceivers are more accurate, and can therefore be thought of as an inverse ''level accuracy'') is the absolute value of the difference between a perceiver's estimate of a given mood (e.g., anger) and the target's self-report of that mood on the same day. A higher score on this index reflects inaccuracy on the part of the perceiver (an inaccuracy that could be due to either under- or overestimation). For

Table 1
Descriptive Statistics for Aaccuracy at the Day and Person Levels

| Perceived Mood | Day-Level | | Person-Level | |
| | Discrepancy | Level Accuracy (Average Discrepancy) | Scatter Accuracy | Pattern Accuracy |
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Anger	.22 (.56)	.29 (.39)	− .02 (.67)	.55 (.74)
Depression	.31 (.45)	.36 (.37)	− .00 (.68)	.31 (.73)
Anxiety	.47 (.53)	.50 (.30)	.01 (.63)	.20 (.75)
Vigor	.82 (.65)	.84 (.34)	− .00 (.42)	.37 (.69)
Efficacy	.76 (.77)	.82 (.49)	.06 (.58)	.25 (.56)
n	1,361	36	$19 \leq n \leq 36$	$18 \leq n \leq 36$

example, on a day when the perceiver estimated his or her partner's anger at either a 3 or a 5, whereas his or her partner recorded his or her own anger at a 4, the discrepancy value would be a 1. The mean day-level discrepancy scores reported in Table 1 reflect the average discrepancy across perceivers and days.

Person-level accuracy indices were computed in the following ways. A perceiver's level-accuracy is the average, across all days, of his or her (day-level) discrepancy scores for a particular mood. As with the day-level index, a higher score for this variable signifies greater inaccuracy. For example, a perceiver who was discrepant by an average of 1 scale point every day would have a person-level (discrepancy) score of 1. The mean person-level discrepancies reported in Table 1 reflect the average person-level discrepancies across all perceivers.

A perceiver's scatter accuracy is the log-transformed ratio of the standard deviations of a perceiver's estimate and the target's self-report. For example, to calculate a particular perceiver's scatter accuracy for anger, the standard deviation of the perceiver's estimates of the target's anger across the days is divided by the standard deviation of the target's self-reports of anger across the diary period. This value is then log-transformed in order to correct for the inevitable skew that results from calculating a ratio (centered, as it is, on 1, with a fixed lower bound at 0 but no fixed upper bound).[2] With the transformed values, a score of 0 reflects perfect accuracy; high scores indicate overestimation and low scores indicate underestimation on the part of the perceiver. The mean person-level scatter accuracies reported in Table 1 reflect the average scatter accuracy across all perceivers.

A perceiver's pattern accuracy is the Fisher-transformed correlation between a perceiver's estimates of a target's given mood and the target's self-reports over the course of the diary period (i.e., 21 days). For example, to calculate a particular perceiver's pattern accuracy of anger, his or her estimates of the target's anger on all of the days are correlated with the target's self-reports of anger on all of the days. These scores range between − 1.0 and 1.0. Higher positive scores reflect greater accuracy; a score of 0 reflects no association between a perceiver's estimates and the target's self-reports; negative scores

---

2. Because ratios are bounded by 0 on one side but unbounded on the other side, they are positively skewed. A log transformation addresses this problem. This skewness is a simple artifact of using ratios when the size of the denominator may just as likely be bigger as smaller than the size of the numerator.

reflect an inverse relationship between perceiver and target (e.g., perceiver reports high amounts of anger when the target reports low amounts). Before submitting it to further analyses, the obtained correlation coefficients are Fisher-transformed (to normalize their distribution). The mean person-level pattern accuracies reported in Table 1 reflect the average pattern accuracy across all perceivers.

Hypothesis 1: Day-Level Indices

We expected that the discrepancy indices for the various target moods would not be synonymous with each other. However, we did expect them to be associated with one another along valence lines. We can think about these predictions in the context of our hypothetical couple, Anne and John. If they are correct, then on days when Anne was discrepant in her perceptions of John's anger she is also likely to be discrepant in her perceptions of his anxiety, but not necessarily discrepant in her perceptions of his vigor.

To account for the nonindependence in our data, we used multilevel models, with days nested within persons, and persons nested within couple. The discrepancy regarding a given mood (e.g., anger) was predicted from discrepancy toward another mood (e.g., anxiety). Below is an example of the Level-1 (within-person) model for testing this association for individual i in couple j on day k:

$$\text{discrepancy(anger)}_{ijk} = \beta_{0ij} + \beta_{1ij}\text{discrepancy(anxiety)}_{ijk} + r_{ijk}$$

The model includes an intercept term (β_{0ij}) that estimates the dependent variable (discrepancy toward anger) when the predictor variable (discrepancy toward anxiety) is equal to zero (i.e., when there is no discrepancy toward anxiety). It also includes a slope term (β_{1ij}) that estimates the association for that person between the level-1 dependent and predictor variables. We ran each pair of target moods in both directions and report the lower coefficient for each pair in Table 2. Associations within valence were significant and positive. Slope estimates among the negative moods ranged from $b = .27$ to .30. The estimate of the association between discrepancies in the two positive moods was lower, but still positive and significant ($b = .07$). Across valence, however, only one of six associations was significant (discrepant perceptions of depression predicted those of efficacy, $b = .05$, $p < .05$). These results indicate that individuals who were inaccurate (as indexed by discrepancy at the daily level) in their perceptions of one negative mood were likely to be inaccurate in

Table 2
Day-Level Prediction of Discrepancy Across Emotions

Accuracy Regarding ...	Predicted by Accuracy Toward ...	Discrepancy Estimate		
		b	SE	df
Within Valence: Negative Moods				
Depression	Anger	.27*	.04	30
Depression	Anxiety	.28*	.04	36
Anxiety	Anger	.30*	.04	30
Within Valence: Positive Moods				
Vigor	Efficacy	.07†	.03	36
Across Valence				
Anger	Vigor	.01	.03	36
Depression	Vigor	.04	.04	36
Anxiety	Vigor	.03	.03	36
Anger	Efficacy	.05	.03	36
Depression	Efficacy	.05*	.02	36
Anxiety	Efficacy	.03	.02	36

Note. Values are unstandardized multilevel regression coefficients.
$^{\dagger}p < .05$. $*p < .001$.

their perceptions of other negative moods; those inaccurate in perceiving one positive mood were likely to be inaccurate in their perception of the other positive mood; however, individuals' accuracy in the negative moods was, for the most part, not associated with their accuracy in perceiving positive moods.

As an additional test of whether accuracy is segregated along valence lines, we ran a P-technique confirmatory factor analysis (Rausch, 2009). In this analysis, we examined a model in which one latent factor reflects accuracy about negative moods (anger, depressed mood, and anxiety) and another reflects accuracy about positive moods (vigor and efficacy). The factors were allowed to covary. Daily accuracy scores were centered around both the couple and individual mean to account for nonindependence and to remove

any random effects due to person or couple. Results from the analysis revealed a good fit for our model. The root mean square error of approximation (RMSEA) statistic for this model was .011, less than the typical requirement of .05. The negative accuracies were all strongly associated with the negative accuracy factor (with loadings of .64, .71, and .56 for anger, depressed mood, and anxiety, respectively). The accuracy index for efficacy was strongly associated with the positive factor (loading = .71). The accuracy index for vigor failed to load as strongly (.21). In addition, a small but positive correlation ($r = .24$, $p < .01$) was found between the positive accuracy and negative accuracy factors, suggesting that accuracy toward positive and negative moods was somewhat (although not strongly) related. Thus, accuracy indices about negative partner moods represent a distinct tendency or skill, whereas those about positive moods may be less cohesive and may operate in a more mood-specific way.

Hypotheses 2–4: Person-Level Indices

The remaining three hypotheses involved person-level, rather than day-level, predictions. To examine these, we computed for each perceiver a set of accuracy scores regarding the level, scatter, and pattern of each mood. The discrepancy indices regarding the target's moods (level accuracies) are averages of the perceivers' daily discrepancies described earlier. In other words, a perceiver's accuracy regarding the target's mood level reflects an average, taken across the entire diary period, of the day-level discrepancy indices. A perceiver's scatter accuracy indices reflect the extent to which the variability in his or her estimates of the target's moods approximates the variability in the target's actual moods. Finally, a perceiver's pattern-accuracy indices are the Fisher-transformed Pearson correlations between the perceiver's and the target's reports of the target's mood over the course of the diary period.

Hypothesis 2

We expected that at the person level, the accuracy indices (for either level, scatter, or pattern) for the various target moods would not be synonymous with each other (e.g., scatter accuracy about vigor will not be strongly related to scatter accuracy about depression, and so on). Instead, we again expect them to be associated with one another along valence lines (e.g., scatter accuracy indices regarding negative moods would be positively associated with each other, as would those regard-

ing positive moods). For example, if Anne is typically discrepant in her perceptions of John's anger, we expected her to also be discrepant in her perceptions of his depression. Similarly, if she tracks John's vigor accurately (as indexed by the pattern-accuracy score), we expected her to track his efficacy accurately as well. However, her accuracy regarding level, scatter, or patterns of negative moods was expected to be unrelated to these accuracy indices regarding positive moods.

Confirmatory factor analyses similar to those used for Hypothesis 1 (regarding the day-level results) are not suitable for the person-level analyses due both to the size of our sample and to the double nesting (of indicators within moods or vice versa, as well as of partners within couples). Thus, we took a different approach to testing our person-level hypotheses. Intercorrelations among all person-level accuracy indices were computed separately for male perceivers and for female perceivers. Results are reported in Table 3 (with men below and women above the diagonal, respectively). The patterns of results were essentially identical, with the few gender differences noted below.

Associations among subjects' level-accuracy indices regarding negative target moods were uniformly strong, positive, and significant (for women, average $r = .66$, all $ps < .001$; for men, average $r = .62$, all $ps < .05$). The association between level-accuracy indices regarding the two positive target moods was also positive and significant, though weaker (for women, $r = .33$, $p < .01$; for men, $r = .32$, $p < .05$). Interestingly, the index for one positive mood, efficacy, was positively associated with the indices regarding negative target moods (for men, $r = .52$, $p < .001$ for efficacy and anger and $r = .48$, $p < .05$ for efficacy and depression; for women, average correlation for efficacy with the three negative moods was $r = .46$, $p < .05$). Overall, while associations within valence were strong and positive, the index for one positive mood crossed valence lines.

Individual differences in the scatter-accuracy indices were somewhat, though not entirely, valence dependent: The indicators for the three negative moods (anger, anxiety, and depression) were positively associated (for men, average $r = .58$; for women, anger and depression $r = .58$, and anger and anxiety $r = .60$; all $ps < .05$; note that for women, the association of the depression and anxiety indicators was nonsignificant). In contrast, there was no association between the indicators for the two positive moods, and though the indicators for the negative moods were unrelated to that for vigor (average $r = -.02$, ns), the indicators for the three negative moods

Table 3

Perceiver Intercorrelations Among Indices (Hypotheses 2 and 4)

Target Mood	Discrepancy (Level Accuracy)					Scatter					Pattern				
	Ang.	Dep.	Anx.	Vig.	Eff.	Ang.	Dep.	Anx.	Vig.	Eff.	Ang.	Dep.	Anx.	Vig.	Eff.
Discrepancy															
Anger	—	.85*	.55*	.28	.44†	**.28**[a]	.04	.07	−.09	.30	**.15**[a]	−.77*	−.71*	−.42†	.63*
Depression	.88*	—	.59*	.20	.38†	.12[a]	**.20**	.06	−.13	.28	.05[a]	**−.70***	−.71*	−.44†	.38†
Anxiety	.42†	.55*	—	.15	.59*	.00[a]	−.13	**−.10**	.27	.27	.51†[a]	−.54†	**−.54***	.16	.66*
Vigor	.18	.19	.29	—	.33†	.28[a]	.08	.26	**.07**	.43†	−.03[a]	−.17	−.12	**−.27**	−.09
Efficacy	.52*	.48†	.12	.32†	—	.23[a]	−.07	.05	.35†	**.37†**	.58†[a]	−.53†	−.37†	.14	**.49†**
Scatter															
Anger	**.02**	.05	.05	−.11	.03	—	.58†[a]	.60†[a]	−.12[a]	.59†[a]	**−.06**[a]	.15[a]	.06[a]	−.11[a]	−.29[a]
Depression	.04	**.11**	.10	−.05	.02	.57†	—	.25	−.33	.40†	−.04[a]	**.00**	−.03	−.33	−.62*
Anxiety	.14	.13	**.01**	−.15	.47†	.57†	.61*	—	.15	.50†	−.08[a]	.01	**−.02**	−.15	−.34
Vigor	.21	.04	−.08	**−.02**	.28	.08	−.10	.15	—	.13	.63†[a]	−.18	−.22	**.57***	.49†
Efficacy	−.10	−.18	−.11	−.09	**.02**	.31	.30	.22	.14	—	.04[a]	−.01	−.01	−.24	**−.20**
Pattern															
Anger	**−.37**	−.36	−.07	−.24	−.22	**.43†**	.27	.28	−.02	.01	—	−.55†[a]	−.56†[a]	.75*[a]	.57†[a]
Depression	.09	**−.09**	−.03	−.21	−.06	.25	**.17**	.26	.25	.32	.30	—	.88*	.15	−.75*
Anxiety	−.55*	−.61*	**−.07**	−.33†	−.35†	.04	−.03	**−.01**	.17	.23	.36	.34†	—	.07	−.68*
Vigor	.41†	.51†	.14	**−.09**	.29	.26	.42†	.48†	**−.15**	−.11	.42†	.05	−.54*	—	.77*
Efficacy	−.09	−.10	.08	.13	**−.32**	.18	.09	−.15	−.07	**−.11**	.47†	.55†	−.08	−.11	—

Note. Results for male and female perceivers reported below and above the main diagonal, respectively. Bold-faced values are same-mood, cross-index correlations.

$n = 20$, except where noted ([a]$15 \leq n \leq 19$).

†$p < .05$. *$p < .001$.

were positively associated with that for efficacy for women (average $r = .50$, $ps < .05$; the corresponding associations for men were also positive though nonsignificant). Thus, as was found with level-accuracy (discrepancy) indices, the scatter-accuracy index for efficacy stood out as being associated with negative target moods.

Intercorrelations among the pattern-accuracy indices did not generally cohere to valence lines. While the pattern-accuracy indices for depression and anxiety were positively associated for both men and women ($r = .34$, $p < .05$, and $r = .88$, $p < .001$, respectively), the indicators for the positive target moods were only associated for the women ($r = .77$, $p < .001$). The pattern accuracy toward anger was not associated with the indices for the other negative target moods for the men and was significantly negatively associated with the indices for depression and anxiety for the women ($r = -.55$ and $r = -.56$, $ps < .05$, respectively). As with the discrepancy index, pattern accuracy toward anger was strongly and positively associated with the indices for vigor (for men, $r = .42$, $p < .05$; for women, $r = .75$, $p < .001$) and efficacy (for men, $r = .47$, $p < .05$; for women, $r = .57$, $p < .05$). The indicator for efficacy was also associated with the indicator for depression (for men, $r = .55$, $p < .05$; for women, $r = -.75$, $p < .001$) and anxiety (for women only, $r = -.68$, $p < .001$). Additionally, for men, the pattern-accuracy index for anxiety was negatively associated with that for vigor ($r = -.54$, $p < .001$). Generally, there were considerable cross-valence associations with the pattern-accuracy index, particularly for the anger and efficacy indices.

Overall, results for Hypothesis 2 were mixed, with stronger support for valence dependence in level accuracy than in scatter or pattern accuracy. While indicators of negative moods seem to cohere, the indicators for one of the positive moods (efficacy) and one of the negative moods (anger) are also associated with the accuracy in perceptions of cross-valence target moods.

Hypothesis 3

We expected that person-level accuracy indices regarding positive moods would yield lower accuracy scores than those regarding negative moods. To examine relative accuracy in the different target moods, a series of paired t tests with Holm-Bonferroni corrections was conducted for each accuracy index. We report results for all perceivers (men and women combined) and note any gender differences.

Results for the person-level discrepancy index revealed that all but 1 of the 10 pairs were significantly different, average $t(73) = -6.26$, all $ps < .005$; the single exception was the nonsignificant difference in the mean discrepancies of the two positive emotions, vigor and efficacy, $t(73) = .36$, *ns*. Discrepancies (inaccuracies) regarding negative emotions were smaller than those regarding positive emotions (see Table 1 for means). The lowest discrepancy (greatest amount of accuracy) was that regarding anger, followed by depressed mood and then anxiety. When examined separately by gender, the same pattern held, though for the women, two of the comparisons among negative moods were no longer revealed as significantly different (anger and depression, $t(35) = -1.94$, *ns*; depression and anxiety, $t(35) = -2.41$, *ns*). Nonetheless, for both the total sample and for the genders separately, the discrepancies for all negative moods were lower than those for the positive moods.

Results for the scatter index reveal that while the perceivers were the least accurate in regard to their perceptions of efficacy ($M = .06$) and the most accurate in regard to their perceptions of anger ($M = -.02$), none of the comparisons were significant; average t (*df* ranged from 40 to 72) $= .27$. The same was true when these indices were examined separately by gender.

Results for the pattern-accuracy index (the Fisher-transformed correlation coefficient) ranged from .20 to .55, but both the greatest and the lowest estimates were regarding negative moods (pattern accuracy regarding anger being highest, and pattern accuracy regarding anxiety being lowest). None of these comparisons was significant, average t (*df* ranged from 35 to 72) $= 1.28$. The same was true when these indices were examined separately by gender.

Overall, these results suggest that greater accuracy regarding negative moods may predominantly occur in the level accuracy (discrepancy) domain, and to a lesser extent in the scatter- and pattern-accuracy domains.

Hypothesis 4

Our last hypothesis was that the three conceptually distinct components of accuracy (level, scatter, and pattern) regarding the same moods would be independent from each other. For example, we expect Anne's overall tendency to underestimate, say, the degree to which John's depressed mood fluctuates (i.e., low scatter accuracy)

will not necessarily suggest a reduced ability to track his depressed mood over time (low pattern accuracy) or a high average discrepancy between her inferences and his actual depressed mood (low level accuracy).

To determine whether these indices measured similar or divergent aspects of accuracy, we correlated the individuals' level-, scatter-, and pattern-accuracy indices for each target mood. These results are displayed along the diagonals of the across-methods sections of Table 3. Of the 30 intercorrelations (15 for each gender), only 6 proved to be significant—all but one for the women. The discrepancy index (which reflects low level accuracy) was associated positively with the scatter and pattern indices for efficacy ($r = .37$, $p < .05$, and $r = .49$, $p < .05$, respectively) and negatively with the pattern indices for depression ($r = -.70$, $p < .001$) and anxiety ($r = -.54$, $p < .001$). The scatter and pattern indices were associated for vigor ($r = .57$, $p < .001$). For male perceivers, the scatter and pattern indices were associated for anger ($r = .43$, $p < .05$). In other words, there is some evidence that, particularly among the women sampled, greater accuracy regarding levels was associated with lower scatter or pattern accuracy for positive moods but to higher pattern accuracy for negative moods. However, for the most part (i.e., in 24 out of 30 cases), the different components of accuracy were unrelated.

In summary, our data offer considerable (though not full) support for Hypotheses 1–4. The proposed EA indices reflect differing aspects of accuracy at both the day and the person levels, with considerable similarity for accuracy indices within valence and considerable dissimilarity across valence. Results using the discrepancy index, in particular, suggest that accuracy regarding negative target moods may be more pronounced than accuracy regarding positive target moods.

DISCUSSION

Our aims for this study were to help develop a macro approach to the assessment of EA in everyday life, using daily diary methods, and to clarify the nature of EA at both the day and the person levels. We proposed examining empathic accuracy at these two levels with a multifaceted set of indices because of the potential for distinct psychological implications for these indices, and we specified four

hypotheses (the first at the day level, the remaining ones at the person level) to examine our proposal.

In support of Hypothesis 1, day-level accuracy indices (discrepancies regarding various moods) were found to be valence dependent. Days marked by inaccuracy regarding any particular mood were likely to be marked by inaccuracy toward other moods of the same valence, but with very few exceptions, not of the opposite valence. Results of confirmatory factor analyses supported the existence of a latent factor of accuracy toward negative moods. The evidence for a latent factor of accuracies toward positive moods was weaker, but it was also obtained.

Next, in partial support of Hypothesis 2, individuals characterized by inaccuracy regarding any particular mood were again likely to be characterized by inaccuracy toward other moods of the same valence, but there were exceptions to this rule (e.g., pattern accuracy regarding anger was positively associated with pattern accuracy regarding vigor and efficacy; in contrast, scatter accuracies for the two positive moods were unrelated). In partial support of Hypothesis 3, accuracy as indexed by person-level discrepancy was more pronounced for the negative target moods than for the positive target moods. However, this was not the case for scatter or pattern accuracy. Finally, in support of Hypothesis 4, different person-level accuracy indices (discrepancy, scatter, and pattern accuracy) regarding the same target moods were mostly unrelated to each other (the few exceptions occurring mostly among female participants).

Empathic Accuracies (Plural)

Our findings suggest that these various indices of EA are largely empirically distinct from one another. We believe that there are also meaningful psychological differences between these various forms of EA: namely, between day- and person-level accuracy; between level, scatter, and pattern; and between accuracies regarding positive and negative target moods.

For one, accuracies at the day level (and within-person fluctuations in such accuracy) are likely to be associated in a more fine-tuned manner with fluctuating relationship and individual outcomes, more so than global (person-level) indices that tell us how accurate a person is *on average*. Daily (or even momentary) outcomes—such as feelings of intimacy, of perceived partner

responsiveness, or of positive or negative feelings within the relationship—are important proximal components of relational well-being. They may also accumulate to create more global outcomes (e.g., relationship satisfaction). However, for global outcomes, we expect aggregated (person-level) accuracy levels to be better predictors.

Second, note that we found the three person-level indices examined (level, scatter, and pattern) to be distinct from one another. Consequently, we expect them to speak to distinct psychological phenomena. Gauging the level of another person's target mood well (i.e., with little discrepancy) seems to require a particular kind of sensitivity, namely, getting the ballpark right. This sensitivity differs from the one required for gauging scatter correctly, which involves recognizing the variability, range, or extremes in a particular mood, regardless of their mean levels. In turn, the sensitivity to either level or scatter differs from that needed for tracking patterns accurately over time. The latter requires an awareness of day-to-day fluctuation patterns and an ability to compare the target's mood on a particular day to his or her mean levels, or at least to his or her mood on preceding days.

Third, our results support the importance of considering not only *how* one is accurate or inaccurate, but also *what* one is accurate about, or at least what valence the mood is in about which one is (or is not) accurate. Specifically, the findings of segregation by valence and of greater accuracy regarding negative moods than positive ones (at least with some indices) suggest that accuracy regarding positive moods may function differently than accuracy regarding negative moods. This difference between ''positive accuracy'' and ''negative accuracy'' may reflect the possibility that the perception of another person's moods carries similar consequences to the perception of moods in the self. Some of our earlier work (Rafaeli et al., 2009) suggests that heightened awareness or differentiation of (one's own) negative emotions is more adaptive than the awareness or differentiation of positive emotions. It may be that there are similar costs to inaccuracies regarding a partner's negative moods, costs that are not incurred by inaccuracies regarding positive moods.

The finding of a valence gap in level accuracy (more so than in scatter or pattern accuracy) may mean that there is something specifically adaptive or functional about being accurate regarding the levels of negative moods. It is possible that great over- or

underestimation of a target's negative moods (but not, say, the mis-perception of their fluctuations, as would be reflected by the pattern index) negatively impacts the quality of support provision (Ver-hofstadt, Buysse, Ickes, Davis, & Devoldre, 2008) or behavioral accommodations made in such situations. In other words, knowing the ballpark estimate of how upset John is (e.g., his level of anxiety) may be more important than realizing that he is relatively more or less anxious than he was yesterday when it comes to reacting ap-propriately to his mood. The way to examine this prediction is by comparing the predictive validity of the various EA indices, which is one clear direction for future research.

What Do We Learn About Empathic Accuracy From These Results?

If there is a single take-home message from the numerous analyses we report here, it is that students of empathic accuracy should al-ways specify "accuracy about what?" and "accuracy in what way?" We do not argue that the question "about what?" must be answered with the particular set of positive and negative moods used in this study. But at a minimum, we have shown that accuracy about pos-itive and negative states is not a unitary construct and that both should be examined. Similarly, we would not insist that the question "in what way?" must necessarily or exclusively be answered by the set of indices presented here. Indeed, different methods may tap profitably into other elements of empathic accuracy (e.g., Kenny & Winquist, 2001). But at a minimum, we have shown that the accu-racy that reflects attunement to momentary states may differ from the accuracy that reflects more global patterns or average tendencies.

In other words, empathic accuracies may really be a family of capacities or tendencies toward attunement of various kinds toward various targets. The same can be (and has been) said for the broader construct of interpersonal sensitivity. For example, in noting that research on interpersonal perception progresses like the findings of the six blind men groping an elephant, Zebrowitz (2001) argued that in order to "capture the beast," our theories must be general enough to include multiple domains of interpersonal sensitivity. Similarly, Bernieri (2001) pointed out that "the skill with which people judge various psychological constructs seems to be as much a function of the thing being judged as it is a function of the person doing the judging" (p. 8; see also Hall & Bernieri, 2001, for a thorough review

of the diversity of topics that fall under the umbrella of interpersonal sensitivity). The aim of this work was to refine our understanding of one set of such sensitivities—the sensitivity to moods that occur in daily life—and to present evidence that both the valence of the moods (positive vs. negative) and the parameter toward which we may be sensitive (i.e., level, scatter, or pattern) matter.

Using These Indices as Predictors: The Costs and Benefits of EA in Everyday Life

Ultimately, a refinement of the EA concept will be considered useful to the extent that it helps us predict meaningful and relevant individual and interpersonal outcomes (e.g., satisfaction, intimacy, trust, and stability). Our understanding of how everyday empathic accuracy functions within close relationships (i.e., what costs and benefits it confers) is still developing. Wilhelm and Perrez's (2004) diary study began exploring the benefits of daily EA but yielded equivocal results. Specifically, though some positive associations were found between accuracy and relationship satisfaction, these depended on the perceiver's gender and on the co-location of the perceiver and the target at the time the diaries were completed. Under some conditions, husbands' accuracy was actually negatively associated with relationship satisfaction. Wives' accuracy seemed to be more beneficial on the whole, but it too varied depending on which emotional dimension was perceived.

Indeed, several authors (beginning with Sillars & Scott, 1983; cf. Ickes & Simpson, 2004) have noted conditions in which accuracy may be bad for the relationship, and which may therefore lead to motivated inaccuracy. These typically involve instances in which the accuracy may reveal a target's negative feelings about the relationship or in which inaccuracy may lead to positive (if illusory) thoughts about the relationship. For example, Simpson, Oriña, and Ickes (2003) used the unstructured dyadic interaction paradigm and found that when thoughts or feelings were threatening to the relationship, less accuracy was related to greater feeling of relationship closeness and satisfaction. Such findings are typically understood to mean that perceivers protect themselves from the effects of their partners' negative emotions by not attending to them.

The motivated inaccuracy view focuses on situations in which the target of empathic inference is a relationship-related—indeed, a

relationship-threatening—emotion or thought. In contrast, most instances of everyday empathic attempts are likely to occur in non-threatening situations. In such situations, greater accuracy is akin to greater understanding—and as such, it should be one hallmark (along with caring and validation) of partner responsiveness (and as important, of *perceived* responsiveness on the part of the target; cf. Reis, Clark, & Holmes, 2004). Prominent recent theories of dyadic relationships highlight the role of such perceived responsiveness in the development of intimacy and in the enhancement of relationships (Reis et al., 2004). In addition, accuracy regarding transitory internal states may play a part in the adaptive relational process of verification: perceivers who are more empathically accurate can serve as a better source of verification for their partners (cf. Swann, Chang-Schneider, & Angulo, 2007). Thus, it is surprising that stronger evidence has not emerged for the benefits of empathic accuracy in nonthreatening situations.

One possibility for this lack of consistent findings may be that most studies have disregarded the multifaceted nature of EA. The present study provides some evidence for the need to distinguish, at the very least, between accuracy regarding positiveversus negative moods, and between components of accuracy (level, scatter, and pattern) that are both conceptually and empirically distinct even for the same mood.

Accuracy regarding two target moods, anger and efficacy, yielded some unexpected results at the person level. For one, pattern accuracy regarding anger was unrelated to pattern accuracies regarding anxiety or depression but was related to pattern accuracies regarding vigor and efficacy. An intriguing possibility that emerges from our data is that detecting anger accurately (i.e., with good temporal tracking) involves mechanisms similar to those used in detecting more prototypical positively activated states like vigor and excitement, but not to those used in detecting negatively activated (and often more inward focused) states such as depressed or anxious mood. This may be best understood by considering recent research revealing that anger may be strongly tied to the behavioral activation system at both a trait and a state level (e.g., Harmon-Jones, 2003).

Accuracy regarding a partner's efficacy also crossed valence lines in several instances. Recall that the efficacy scale (really, a single item) was created in an attempt to balance out the list of mood items, which otherwise tapped only one positive mood (vigor) versus three

negative ones. The considerable similarity between accuracy indices regarding efficacy and those regarding anxiety and depression raises the possibility that this item reflects more deactivation of the behavioral inhibition system (i.e., low negative affect; see Rafaeli & Revelle, 2006). Indeed, if we consider it from the perspective of regulatory focus theory (Higgins, Roney, Crowe, & Hymes, 1994), the experience of efficacy may emerge from successful attempts at prevention-focused self-regulation and thus could be considered akin to low anxiety. However, we could have had greater confidence about this (and other issues) with a more balanced mood measure, including equal representation for positive and negative mood scales.

When using accuracy indices as predictors of individual or relational outcomes, it is important to consider relevant theoretical and statistical issues. For example, since accuracy indices contain two sources of information (namely, the perceiver's *and* the target's reports), it is crucial to control for these components to ensure that whatever predictive power is ascribed to the accuracy index is not due simply to one of the components. For example, if we find that accuracy about depressed mood predicts relationship satisfaction, we would want to show that this holds above and beyond the effects of the target's actual depressed mood and the perceiver's inference about the target's mood.

Empathic Accuracy as a Component of Interpersonal Perception

Two elements of interpersonal perception, related to EA but beyond the scope of the current analyses, are *real* and *assumed similarity* (Wilhelm & Perrez, 2004). The combination of these elements provides one possible source of information for empathically accurate perceivers. Specifically, some (though not all; cf. Wilhelm & Perrez, 2004) empathic accuracy can be due to *indirect* accuracy: correctly assuming similarity between the partners' self-reported emotional states. As Wilhelm and Perrez demonstrated, dyadic diary data can be a useful tool for estimating these direct and indirect elements. We believe this tool could prove even more useful once we know which indices and which target states are characterized by greater or lesser accuracy. In the analyses reported here, we were primarily concerned with developing accuracy indices and determining their utility for the study of EA. Future research should help answer whether these indices are equally attributable to assumed similarity, and whether

we use deviations from our own mood patterns to signal deviations from our partners' patterns.

A concern in EA research is that we resort to using the targets' responses as truth criteria; a mismatch between perceiver and target is then labeled as the *perceiver's* inaccuracy. Yet how can we know it is the perceiver who is inaccurate? Might it not be the target? This question really has two parts. One is whether the target's self-reports are trustworthy, the other whether inaccuracy might actually reflect poor expressivity on the part of the target, rather than inaccuracy on the part of the perceiver. With regard to the first question, we are in agreement with Barrett (2006, p. 24), who reviewed the issue of self-reported emotions by saying, "Verbal report, even with all of its failings, may be the only means of assessing the *experience* of emotion. If we want to know whether a person is experiencing an emotion, we have to ask them." The same is obviously true for thoughts as well.

The second question involves the role of *sending accuracy* (i.e., the target's readability or diagnosticity) in partners' daily accuracy. Even if we accept the target's reports as the truth criterion, placing the onus of (in)accuracy entirely on the perceiver disregards a major point in social interaction research: namely, that dyads bring together actor effects, partner effects, and interaction effects; in this case, the sending accuracy of target partners may exert quite a strong effect on the EA scores of the perceivers, a phenomenon well demonstrated in micro EA research (e.g., Ickes et al., 1990; Thomas & Fletcher, 2003; Zaki, Bolger, & Ochsner, 2008).

This is where the main strength of dyadic diary data, ecological validity, is also their main limitation—they do not provide as ready a check on the targets' readability. Short of conducting a large-scale round-robin study in which participants swap partners (so that X_1, X_2, X_3, ... X_n end up providing perceptions of Y_1, Y_2, Y_3, ... Y_n for several weeks each, and vice versa), we can address this problem in two ways. One is to simply acknowledge that what we call EA is a relationship-specific characteristic of each individual; change the relationship, and EA might change. Still, the relationship in this case is quite an important one, not a randomly created pairing. Duly contextualized, we should still be curious about these relationship-specific EA levels, if the relationship is of such importance.

Another way of addressing the topic of sending accuracy or diagnosticity is by combining macro (dyadic diary data–based) and

micro (lab-based, behavioral observation–based) EA in one study. While diary-based EA indices prove useful in looking at daily dynamics (which are macro, but still fluctuate), we may consider them as complementary to EA findings from research employing more micro moment-to-moment approaches and more stable trait perceptions. If we find that macro EA (or more likely, particular macro EA indices) and micro EA levels are strongly associated, we might be justified in using standard methods for assessing micro (lab-based) diagnosticity (e.g., Thomas & Fletcher, 2003) as a proxy for macro (at-home) diagnosticity.

One methodological challenge faced in the collection of daily diary data is that of compliance and the possibility that participants may communicate with one another about the content of the diary without our knowledge. With electronic diaries, we are able to address the first of these concerns with a time-stamp feature, which confirms the exact time and date of the completion of the diaries). However, we acknowledge that with such data we are unable to monitor participants' compliance with the request that they refrain from discussing the diary or their responses.

CONCLUSION

Previous research may have been somewhat limited in revealing the dynamic role of EA in close relationships, specifically by excluding certain relationship outcomes and by focusing on some types of target transient states (e.g., micro thoughts and feelings in a lab setting) and not others (macro, positively and negatively valenced moods in daily life). By taking empathic accuracy into everyday life, the current study addresses some of these limitations and suggests that EA is a multifaceted construct consisting of multiple components, each assessing a unique aspect of person perception.

REFERENCES

Barrett, L. F. (2006). Solving the emotion paradox: Categorization and the experience of emotion. *Personality and Social Psychology Review*, **10**, 20–46.

Barrett, L. F., & Barrett, D. J. (2001). Computerized experience-sampling: How technology facilitates the study of conscious experience. *Social Science Computer Review*, **19**, 175–185.

Bernieri, F. J. (2001). Toward a taxonomy of interpersonal sensitivity. In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal sensitivity*. Mahwah, NJ: Erlbaum.

Biesanz, J. C., West, S. G., & Millevoi, A. (2007). What do you learn about someone over time? The relationship between length of acquaintance and consensus and self-other agreement in judgments of personality. *Journal of Personality and Social Psychology*, **92**, 119–135.

Cranford, J. A., Shrout, P. E., Iida, M., Rafaeli, E., Yip, T., & Bolger, N. (2006). A procedure for evaluating sensitivity to within-person change: Can mood measures in diary studies detect change reliably? *Personality and Social Psychology Bulletin*, **32**, 917–929.

Cronbach, L. J. (1955). Processes affecting scores on "understanding of others" and "assumed similarity." *Psychological Bulletin*, **52**, 177–193.

Fredrickson, B. (1998). What good are positive emotions? *Review of General Psychology*, **2**, 300–319.

Gable, S. L., Reis, H. T., & Downey, G. (2003). He said, she said: A quasi-signal detection analysis of daily interactions between close relationship partners. *Psychological Science*, **14**, 100–105.

Gaelick, L., Bodenhausen, G. V., & Wyer, R. S. (1985). Emotional communication in close relationships. *Journal of Personality and Social Psychology*, **49**, 1246–1265.

Gilboa-Schechtman, E., Rafaeli, E., Granat, A., Kuint, J., & Feldman, R. (2009). *Beyond alexithymia: The association of self-reported and experience-based emotion awareness with pre-partum somatic experiences, labor difficulty, and post-partum distress*. Unpublished manuscript.

Gottman, J. M., & Porterfield, A. L. (1981). Communicative competence in the nonverbal behavior of married couples. *Journal of Marriage and the Family*, **34**, 817–824.

Hall, J. A., & Bernieri, F. J. (Eds.). (2001). *Interpersonal sensitivity*. Mahwah, NJ: Erlbaum.

Harmon-Jones, E. (2003). Anger and the behavioral approach system. *Personality and Individual Differences*, **35**, 995–1005.

Higgins, E. T., Roney, C., Crowe, E., & Hymes, C. (1994). Ideal versus ought predilections for approach and avoidance: Distinct self-regulatory systems. *Journal of Personality and Social Psychology*, **66**, 276–286.

Ickes, W. (1993). Empathic accuracy. *Journal of Personality*, **61**, 587–610.

Ickes, W. (Ed.). (1997). *Empathic accuracy*. New York: Guilford Press.

Ickes, W. (2003). *Everyday mind reading*. New York: Prometheus.

Ickes, W., & Simpson, J. (2004). Managing empathic accuracy in close relationships. In H. T. Reis & C. E. Rusbult (Eds.), *Close relationships: Key readings* (pp. 345–363). New York: Psychology Press.

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.

Kenny, D. A., & Acitelli, L. K. (2001). Accuracy and bias in the perception of the partner in a close relationship. *Journal of Personality and Social Psychology*, **80**, 439–448.

Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis*. New York: Guilford Press.

Kenny, D. A., & Winquist, L. (2001). The measurement of interpersonal sensitivity: Consideration of design, components, and unit of analysis. In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal sensitivity*. Mahwah, NJ: Erlbaum.

Kilpatrick, S. D., Bissonnette, V. L., & Rusbult, C. E. (2002). Empathic accuracy and accommodative behavior among newly married couples. *Personal Relationships*, **9**, 369–393.

Larsen, J. T., McGraw, A. P., & Cacioppo, J. T. (2001). Can people feel happy and sad at the same time? *Journal of Personality and Social Psychology*, **81**, 684–696.

Lorr, M., & McNair, D. M. (1971). *The Profile of Mood States manual*. San Diego, CA: Educational and Industrial Testing Service.

Pietromonaco, P. R., Rook, K. S., & Lewis, M. A. (1992). Accuracy in perceptions of interpersonal interactions effects of dysphoria, friendship, and similarity. *Journal of Personality and Social Psychology*, **63**, 247–259.

Rafaeli, E., Cranford, J. A., Green, A. S., Shrout, P. E., & Bolger, N. (2008). The good and bad of relationships: How social hindrance and social support affect relationship moods in daily life. *Personality and Social Psychology Bulletin*, **34**, 1703–1718.

Rafaeli, E., Mor, N., Drejet, R., Ehrlich, D., & Bodkin, J. A. (2009). *When details matter and when they do not: Emotion differentiation in non-clinical and emotionally-disordered population*. Unpublished manuscript.

Rafaeli, E., & Revelle, W. (2006). A premature consensus: Are happiness and sadness truly opposite affects? *Motivation and Emotion*, **30**, 1–12.

Rausch, J. R. (2009). Investigating change in intraindividual factor structure over time. *Applied Psychological Measurement*, **33**, 266–284.

Reis, H. T., Clark, M. S., & Holmes, J. G. (2004). Perceived partner responsiveness as an organizing construct in the study of intimacy and closeness. In D. J. Mashek & A. Aron (Eds.), *Handbook of closeness and intimacy* (pp. 201–225). Mahwah, NJ: Erlbaum.

Robinson, M. D., & Clore, G. L. (2002). Belief and feeling: An accessibility model of emotional self-report. *Psychological Bulletin*, **128**, 934–960.

Sillars, A. L., & Scott, M. D. (1983). Interpersonal perception between intimates: An integrative review. *Human Communication Research*, **10**, 153–176.

Simpson, J. A., Oriña, M. M., & Ickes, W. (2003). When accuracy hurts, and when it helps: A test of the empathic accuracy model in marital interactions. *Journal of Personality and Social Psychology*, **85**, 881–893.

Swann, W. B. Jr., Chang-Schneider, C., & Angulo, S. (2007). Self-verification in relationships as an adaptive process. In J. Wood, A. Tesser, & J. Holmes (Eds.), *Self and relationships* (pp. 49–72). New York: Psychology Press.

Thomas, G., & Fletcher, G. J. O. (2003). Mind-reading accuracy in intimate relationship: Assessing the roles of the relationship, the target, and the judge. *Journal of Personality and Social Psychology*, **85**, 1079–1094.

Verhofstadt, L. L., Buysse, A., Ickes, W., Davis, M., & Devoldre, I. (2008). Support provision in marriage: The role of emotional similarity and empathic accuracy. *Emotion*, **8**, 792–802.

Wilhelm, P., & Perrez, M. (2004). How is my partner feeling in different daily-life settings? Accuracy of spouses' judgements about their partner's feelings at work and at home. *Social Indicators Research*, **67**, 183–246.

Zaki, J., Bolger, N., & Ochsner, K. (2008). It takes two: The interpersonal nature of empathic accuracy. *Psychological Science*, **19**, 399–404.

Zebrowitz, L. A. (2001). Groping for the elephant of interpersonal sensitivity. In J. A. Hall & F. J. Bernieri (Eds.), *Interpersonal sensitivity*. Mahwah, NJ: Erlbaum.