

HOW IS MY PARTNER FEELING IN DIFFERENT DAILY-LIFE
SETTINGS? ACCURACY OF SPOUSES' JUDGEMENTS
ABOUT THEIR PARTNER'S FEELINGS AT WORK AND AT
HOME

ABSTRACT. How accurately do spouses know their partner's feelings when the partner is not physically present, but in a specific setting (at work, or at home)? This question addresses a special kind of empathic inference that relies mainly on content knowledge and projection rather than perception. We answered this question using a computer assisted diary approach. A total of 190 husbands and wives out of 95 couples simultaneously recorded how they were feeling and what they thought their partner was feeling, six times each day during an ordinary week. They also recorded where they were, who they were with, and where they thought their partner was. This enabled us to assess measures of accuracy and assumed similarity under natural conditions in spouses' daily lives. Results showed that the spouses' judgements of the absent partner's feelings relied on their own feelings (assumed similarity). Despite this they were often quite accurate, even when assumed similarity was controlled. In general, our findings indicate that spouses have a basic knowledge about their partner's feelings when they are apart. However, the specific situation, the items being judged, and the gender of the spouses also need to be taken into account.

KEY WORDS: accuracy, ambulatory assessment, assumed similarity, couples, daily life, electronic data collection, empathic inference, interpersonal perception

INTRODUCTION

Spouses do not usually spend much of their normal daily life together. Using Experience Sampling data as a base, Larson and Richards (1994, p. 109) estimated that during an average weekday American middle class spouses spent about a quarter of their waking hours in each other's company. Husbands spent most of this time apart at work, while the wives, depending on their employment status, passed this time either at work or at home.

When they are at work, both men and women feel differently from the way they feel at home. However, while men usually report



feeling more positive and less negative at home (Larson and Richards, 1994; Myrtek et al., 1999; Myrtek and Foerster, 2001), women, at least those who are married and have children, tend to have less negative and more positive emotional experiences at work. According to Larson and Richards (1994), women's different experience might be because they receive social rewards at work that they would not get at home. "Their labor is not taken for granted; they got appreciation from others; many of them also received social support" (p. 65). In summary Larson and Richards concluded that spouses live in different realities, have different emotional experiences and needs, and have different expectations and roles to fulfill. Part of the challenge of living together is to understand and respect each other's realities, and to find ways of bringing these realities closer together.

In this article we examine how well spouses know each other's divergent emotional realities. We are specifically concerned with exploring how accurately spouses can predict their partner's feelings when they are apart, either when one (or both) is at work, or when they are at home but are not together.

In more general terms, our study is concerned with a specific aspect of empathic inference. According to the definition of Ickes (2001), *empathic inference* "is the 'everyday mind reading' that people do whenever they attempt to infer other people's thoughts and feelings" (p. 219), while *empathic accuracy* is "the extent to which such mind reading attempts are successful" (p. 219). Although Ickes' definitions exactly match our research topic, it is important to note that the concepts of empathic inference and empathic accuracy are usually used in a narrower sense, where they are restricted to face to face interactions. Thus, empathic inference in Ickes' sense relies on direct perception of the other person's verbal and nonverbal behaviour as well as the inferences drawn from this perception. However, when judging their absent partner's feelings, spouses cannot rely on perception. Their judgments can only be based on knowledge and inference, and on other components we are going to discuss. As far as we know, the question of whether judgments about absent people's states are accurate has not yet been addressed.

Approaches Towards Investigating the Accuracy of Person Perception

Research on empathic accuracy is one branch of the broad field of research on person perception. According to Ickes (1993, 1997), the development of research on the accuracy of person perception has been characterized by a change from stable to fluid and fast changing constructs (from traits to states).

Since its inception, empirical psychological research has investigated how accurately we can judge other people's personality traits (target accuracy).¹ Later research explored dyad members' perceptions of each other's self conceptions, attitudes and values (meta-accuracy). More recently, activity has focused on the observer's accuracy to perceive other people's emotional states (affective or nonverbal sensitivity), while current research is characterized by Ickes' and his colleagues' own work on accuracy at perceiving other people's thoughts and feelings in an ongoing interaction (empathic accuracy).

Research on target or meta-accuracy has been based on both survey and questionnaire approaches (e.g., Cook and Douglas, 1998; Hoch, 1987; Kenny and Acitelli, 2001) and laboratory experiments (Kenny, 1994; Kenny et al., 2001; Funder, 1999). Research on nonverbal sensitivity (e.g., Ambady et al., 2000; Noller, 2001; Johnstone and Scherer, 2000; Keltner and Ekman, 2000) and empathic accuracy (e.g., Ickes, 2001; Levenson and Ruef, 1992; Neyer et al., 1999) was almost exclusively done in the laboratory, using sophisticated experimental designs.

To study empathic accuracy in "natural interactions" Ickes et al. (1990) developed an experimental paradigm. In this paradigm the interaction of two dyad members was unobtrusively filmed in a waiting room scenario. After some minutes the experimenter informed the participants that their interaction had been filmed, and asked them for permission to continue the experiment. The participants were asked to watch the video separately and to record each specific thought or feeling they remembered having. The video was then shown to their partner, and stopped each time the first participant had indicated having had a thought or feeling. The partner was asked to infer what it was. To get a measure of empathic accuracy external observers rated the agreement between each indi-

cated thought or feeling and the partner's estimate. The first part of this procedure in particular has been adapted in several studies to answer specific questions (e.g., Maragnoni et al., 1995; Simpson et al., 1995; Thomas et al., 1997).

Although this paradigm is a powerful tool for studying how accurately people perceive and infer the thoughts and feelings of their interaction partner, given the laboratory context, the external and ecological validity of the findings are of some concern (for a general discussion of this issue see Bolger et al., 2003; Fahrenberg, 1996; Fahrenberg et al., 2002; Wilhelm and Perrez, 2001). Interactions in studies on empathic accuracy were short (some minutes only), and were influenced by the laboratory context (such as a waiting room situation or an instruction to discuss a marital problem or some other topic, which might be negative, neutral or positive). Thus, empathic accuracy has yet only been studied at the micro level (changing states in a time frame of seconds and minutes). In most of the studies the range of thoughts and feelings was restricted, and the intensity of the feelings was rather low.

To overcome some of the limitations of such laboratory experiments, our research group has developed a new approach to study empathic inference in couples' and families' daily lives (Perrez et al., 2000; Wilhelm, *in press*). With the help of handheld computers, we asked husbands and wives during a normal week how they were feeling, and what they thought their partner was feeling at the same moment (see method section for more details). With this data we were able to calculate measures of accuracy and measures for other basic indicators of dyadic interpersonal perception (see next chapter).

Using Bernieri's (2001) dimensions, the type of accuracy that can be studied with our approach can be characterized as the accuracy of spouses' (or other dyads', e.g., parent-child, therapist-patient) mutual judgments about each other's changing states, in the setting and under the conditions in which these states naturally occur, which are either made from direct perception in face-to-face interactions, or inferred from content information. The time frame is one of hours and days and is situated between a microscopic (changes over seconds or minutes) and a macroscopic perspective (changes over months and years) (Larson and Almeida, 1999).

Basic Aspects of Interpersonal Perception

The basic indicators of dyadic interpersonal perception (Kenny, 1994; Kenny and Acitelli, 2001) are similarity, reciprocity, accuracy and assumed similarity. Further concepts which could also be assessed using our approach include meta-accuracy and assumed reciprocity (Kenny, 1994), but we have not yet done this. Although we have calculated measures for similarity and reciprocity, our main focus has been on the accuracy and assumed similarity of spouses' judgments of each other's feelings.

Similarity (or real similarity) refers to the equality between two partners' self-judgments. In the content of our research, it indicates how similarly husbands and wives feel at the same time.

Reciprocity refers to the equality between the spouses' mutual partner-judgments. It indicates how similarly partners think about each other's feelings.

Accuracy is the equality between the spouses' partner-judgments and their partner's self-judgments. Separate accuracy coefficients can be calculated for husbands and wives. For example, the wives' accuracy is an indicator of how close the wives' judgments of their partner's feelings correspond to the husbands' ratings of their own feelings.

Assumed similarity is the equality between the spouses' partner-judgements and their own self-judgments. The husbands' assumed similarity for example measures how similar the husbands' own feelings are to their judgments of their wife's feelings. The term *projection* has also been used (e.g., Hoch, 1987; Neyer et al., 1999) in addition to the term *assumed similarity* (Cronbach, 1955; Gage and Cronbach, 1955). However, in the context of interpersonal perception research, *projection* is primarily defined operationally, and its meaning is quite different from the psychoanalytical meaning of projection as an unconscious defence mechanism (Kenny, 1994). Kenny and Acitelli (2001) have also used the more general and neutral term *bias* to refer to assumed similarity.

Almost 50 years ago Gage and Cronbach (1955; Cronbach, 1955) pointed out that accuracy scores might be inflated by assumed similarity. They argued that even a lazy judge who simply takes his self-judgments as the other-judgments (full projection) will be very accurate if the real similarity between the judge and the target is

high. The accuracy measure might not therefore be a true indicator of the judge's ability to accurately perceive or infer the target's trait or state. On the other hand, if a judge knows that the similarity between himself and the target is high, as it is often the case in couples, use of projection would even be a sensible strategy for accuracy. Hoch (1987) showed that, in the absence of valid additional information, judges could have improved the accuracy of their judgments through using projection. Thus, overall accuracy coefficients should be separated into the part that is due to correctly-assumed similarity (*indirect accuracy*) and the part that does not rely on assumed similarity (*direct accuracy*) (Kenny and Acitelli, 2001).

A New Approach to Assess Accuracy and Assumed Similarity of Spouses' Judgments of Each Others' Feelings in Their Usual Daily Lives

In the Second Fribourg Family Project (Perrez et al., 2000) 95 husbands and wives participated. They reported their own and estimated their partner's feelings six times a day, over the course of a week. Spouses also gave us information about their current setting (e.g., where they were, who they were with, what they were doing) and their estimate of their partner's local setting. In addition we asked them about other concepts which are not relevant here (see Michel, in preparation; Perrez et al., 2000; Perrez et al., 2001; Schoebi et al., in preparation. Further details are given in the method section; for the results reported below see Wilhelm, in press).

Spouses' accuracy at judging their partner's average feelings (individual mean across the 42 repeated observations) was high (accuracy correlation coefficients for eight different feeling items were between $r = 0.46$ and $r = 0.68$). However, assumed similarity was even higher (ranging from $r = 0.58$ to $r = 0.86$), indicating that spouses rated the average feelings of their partner very similarly to their own feelings. A substantial part of these high coefficients can be attributed to the fact that spouses used the rating scales for judging their partner's feelings in the same way than they used it to judge their own feelings. Another part was due to the fact that the feelings of husbands and wives were actually similar to a certain extend (correlations ranging from $r = 0.30$ to $r = 0.45$), and so they correctly assumed similarity. Therefore, direct accuracy measures

that were controlled for assumed similarity (according to the model of Kenny and Acitelli, 2001, see also Figure 1) were substantially lower, indicating that a good part of the overall accuracy was due to correctly-assumed similarity.

We also looked for differences in accuracy and assumed similarity when judging the partner's emotional states² in situations where husbands and wives were together and where they were not together.³ We could show that both the overall accuracy and the direct accuracy were higher when partners were together. This result was expected because spouses have more information about each other's feelings when they are together. They can directly perceive each other's verbal and nonverbal behavior, and they have an exact knowledge about the situational circumstances. Therefore, they should be more accurate. However, when spouses were apart from each other, accuracy was not zero. They also reached a reasonable level of accuracy even in these situations. This result indicates that spouses have some basic knowledge about how their partner is feeling when he or she is apart.

This raises questions about the nature of this "knowledge". What is it that helps spouses to make reasonably accurate judgements about their partner's feelings without having any possibility of perceiving their partner's behaviour? Is that "knowledge" setting-specific? Do they "know" how their partner is feeling at work? And do they also "know" how he or she is feeling at home when they are not together? Are wives more accurate than husbands? Are spouses who have a more accurate "knowledge" about their partner's feelings in another setting more satisfied with their relationship? These are the questions we address in this paper.

Where Might Accuracy in Judging an Absent Partner's Feelings Come From?

When judging their absent partner's feelings, spouses obviously cannot rely on perceptions about their partner's verbal or nonverbal behavior. An exception are situations in which spouses talk to each other via phone. However, these situations are special because although spouses are not physically together, they are in direct contact and are able to communicate and also derive perceptions about each other. If spouses cannot directly observe their partner,

their judgments about their partner's feelings might rely on knowledge and inference, guessing, projection or stereotypes.

We would expect that accuracy in the judgment of the absent partner's feelings is largely based on knowledge and inference. Knowledge can be acquired through communication. Spouses who frequently talk about their daily experiences with each other know to a certain extent how their partner might usually feel in a certain setting. They might also know something about specific circumstances which could have an influence on their partner's feelings. For example, a wife might know that her husband's colleague is currently on holidays, from which she could infer that her husband has more work to do, and might be experiencing more stress than usual. Conversely, she might know that the relationship between her husband and this colleague is bad, and might therefore assume that her husband feels happier and less angry at work while the colleague is absent than he does in general.

According to Stinson and Ickes (1992), friends (and partners) develop *shared knowledge structures* during their relationship history that should facilitate their mutual understanding of each other's thoughts and feelings. Using the experimental paradigm discussed above, they showed that friends in particular were more accurate than strangers when they inferred each other's thoughts and feelings about events that had occurred at another place or at another time. Although the participants in Stinson and Ickes' study were in direct contact with each other, we assume that such common knowledge structures also facilitate how spouses judge each other's feelings when they are not together. We do not have the means to directly prove this expectation in our present study. However, we are able to exclude alternative explanations.

One might argue, that spouses simply guess how their partner feels. However, if spouses reach a level of accuracy that is higher than would be expected by chance, it is implausible that their judgments are based solely on guessing. Our expectation is that spouses' accuracy at judging their absent partner's feelings at home or at work is higher than chance (Hypothesis H-1).

Spouses might also be more accurate if they use their own actual feelings to judge their partner's feelings. However, this will only work if the husband's and wife's feelings are, to some extent,

similar. There is empirical evidence that it is a common strategy to use one's own experience as an anchor to judge other persons' experiences (Markus et al., 1985; Nickerson, 1999). Various studies have demonstrated strong assumed similarity or projection that was higher than accuracy in most of the cases (Hoch, 1987; Kenny and Acitelli, 2001; Neyer et al., 1999; Thomas et al., 1997). Thus, we expect this also to be true in our study: When spouses judge their absent partner's feelings, assumed similarity will be higher than accuracy (Hypothesis H-2). However, we do not think that accuracy can be completely explained by projection or assumed similarity. Therefore, we expect that accuracy is higher than chance even when assumed similarity is controlled (Hypothesis H-3).

A more subtle process of projection is likely when husbands and wives are not in equivalent settings. When husbands judge their wife's feelings at work, they might think about how they themselves feel at the workplace. Thus, they abstract from their actual feelings in the judgment situation, and use their own emotional experience in an equivalent situation to judge how their partner is feeling in such a situation. This kind of projection is already due to inference. According to Nickerson (1999), there is empirical evidence that judges use their own situation-specific knowledge and experience to judge other people's experience or knowledge in equivalent situations. We therefore expect that the judgment of the partner's feelings at work will be based on the spouse's own emotional experience at work (Hypothesis H-4).

Situation-specific stereotypes or setting-specific cognitive schemata might be a further source of accuracy. If a wife thinks that everyone would be likely to feel worse at work than in a non-work setting, although her judgment might be accurate, it would not be specific for her husband.

However, not only the judgment of another persons' emotional state, but also the judgment of one's own emotional state seems to be influenced by setting-specific schemata, as Myrtek and colleagues showed in various ambulatory psychophysiological monitoring studies (Myrtek et al., 1999, Myrtek and Foerster, 2001; Myrtek et al., 2001). They found that physiological indicators of arousal and emotional strain (an increase in heart rate that is not due to physical activity) were higher at home than at work, but participants rated

their own state at work as more aroused and less comfortable than at home. Thus, their subjective ratings did not correspond to their physiological reactions, but to setting-specific cognitive schemata. Therefore, accuracy might be due to common situation-specific schemata. Because we did not assess setting-specific stereotypes in our study, we were not able to test this hypothesis.

We think that judgments of the absent partner's feelings are based on a mixture of all these components. Spouses could reach a certain amount of accuracy if they drew inferences on the basis of their knowledge of how their partner reacts in specific situations. They could also to be accurate through projection, but only if real similarity exists, and they could be accurate through setting or situation-specific stereotypes, but only if these stereotypes are shared.

Gender Differences in Accuracy

According to widespread gender stereotypes, women should be more accurate in their interpersonal perceptions than men (Brody and Hall, 2000; Graham and Ickes, 1997). In fact, a meta-analysis (Hall, 1978) and a series of later studies found that women were consistently more accurate than men when decoding non-verbal behavior, except for the decoding of anger (Brody and Hall, 2000). These effects were small to moderate. However, gender differences diminished when spontaneous behavior or "leaky" cues were judged, or when participants had more praxis (see Graham and Ickes, 1997). Evidence was mixed in studies on empathic accuracy. In their first studies, Ickes and his colleagues found no gender differences, whereas later studies suggested that women were better judges when they were aware of being evaluated on an empathy relevant task (Ickes et al., 2000; see also Klein and Hodges, 2001). Kenny and Aticelli (2001) did not find consistent gender differences in spouses' accuracy at judging their partner's caring, feelings of job satisfaction or closeness to the partner and the family, or their sexual enjoyment. Funder (1999) reported mixed evidence about accuracy in judging personality traits, while a review on gender differences in empathy and related capacities by Eisenberg and Lennon (1983) concluded that better results for females were closely related to the methods used to assess empathy.

Although most of these judgment abilities are not directly related to the accuracy of judging the absent partner's feelings, we have taken them as a general reference when formulating our hypothesis: Gender differences in accuracy at judging the absent partner's feelings are rather weak. But if gender differences do appear, women should be more accurate than men (Hypothesis H-5).

Is Accuracy Related to Relationship Satisfaction?

It is plausible to expect that accurate judgments of our partner's nonverbal behavior, their thoughts and feelings, their values and ideals of relationship, and their personality traits should, in general, enhance mutual understanding and help us to maintain relationship satisfaction. As Noller and Ruzenne (1991, p. 204) stated: "It is taken for granted that marital harmony is strongly related to effective communication between spouses, and that effective communication, to some optimal degree, involves spouses' understanding of each other's thoughts and feelings, personal experiences of life in general, and the marital relationship in particular". Gottman and Silver (1999, p. 48) similarly argued that couples who are "intimately familiar with each others' world" and therefore "having a richly detailed love map" of their partner are "emotionally intelligent". These spouses know each other's preoccupations and worries as well as their pleasures and delights, and they keep this knowledge up to date. For Gottman and Silver, an elaborated love map is an important factor in satisfying interactions, enabling the partners to give adequate responses and have more reciprocity of positive affects, and therefore to be happier within their relationship.

Empirical support for this line of reasoning is mixed. Various studies have found evidence that accuracy in decoding nonverbal messages is associated with higher marital satisfaction. However, results were not always consistent between partners (Gottman and Porterfield, 1981), or appeared in some but not in all of the different variables that were tested (Noller, 1984; Noller and Ruzzene, 1991; Noller and Feeny, 1998).

In a study on empathic accuracy in early married couples, Bissonnette et al. (1997) found a positive connection between empathic accuracy and features of the relationship such as dyadic adjustment and commitment, but could not replicate it one year later.

In a further study of Thomas et al. (1997), relationship satisfaction was not correlated with empathic accuracy.

Furthermore, other studies have found a negative connection between indicators of relationship satisfaction and accuracy (Ickes and Simpson, 1997; Sillars, 1985). The results of their own experiment led Simpson et al. (1995) to the conclusion that partners tend not to be accurate, especially in situations that threaten their relationship or their self-esteem.

We expect that having an accurate idea of one's partner's feelings at home or at work should, in general, have positive effects. If spouses can accurately infer their partner's feelings when they are apart, it should be easier for them to understand each other's feelings when they are together. This ability to accurately predict how one's absent partner is feeling indicates that spouses have what Gottman and Silver (1999) would describe as a well developed and actualized "love map" of each other. We therefore assume that accuracy in judging an absent partner's feelings is higher in couples whose relationship is highly satisfying (Hypothesis H-6).

METHOD

Assessment of Spouses' Own Feelings and the Judgment of Their Partner's Feelings in Different Daily-Life Settings

To assess the daily life experiences of family members, we developed a computer-assisted self-monitoring system (FASEM-C, Perrez et al., 2000). Six times a day over the course of a week spouses reported their own emotional state, their judgment about their partner's emotional state, and information about the current setting. They also recorded additional information which is not relevant to this article, such as physical complaints, causal attributions and coping with conflicts and stressful events (see also Michel, in preparation; Perrez et al., 2001; Schoebi et al., in preparation).⁴

Each participant was given a handheld computer (Hewlett Packard, HP 360 XL), which presented the questions and the answering options. Other than for the first observation in the morning, when participants were expected to answer the questions immediately after waking up, the computer gave an acoustic signal whenever a record was expected. The interval between each signal

was three hours, plus or minus an additional component between 0 to 30 minutes so that the participants could not anticipate when the signal would be given. The computers used in each individual family were synchronized to ensure that the responses of the different family members were recorded at the same time.

The handheld computer automatically stored the time of the recording and the duration of every assessment. This allows us close control of the participants' compliance and commitment. A total of 62.3 percent of the records were started within five minutes, an additional 18.2 percent within 30 minutes and another 9.9 percent within one hour after the signal, which indicates a reasonably satisfactory compliance for 90 percent of the recordings.

The instrument produces slight *reactivity-effects* over the time of the observation, such as an improvement of the emotional state, a decrease of the within day variability of the emotional state, or a reduction of responses that are followed by further questions. However, these *reactivity-effects* were small (for more details see Perrez et al., 2000).

The Sample

The couples were a sub-sample of the Second Fribourg Family Project (Perrez et al., 2000), that was founded by the Swiss National Science Foundation. In this project Swiss families with mothers, fathers (or step fathers) and at least one adolescent child between 14 and 17 years, living together in the same household were recruited. From the administrations of secondary schools and the town population census we got addresses of households in which adolescent children lived. We wrote to approximately 5200 families explaining the purpose of our study and inviting them to take part. Those families who were interested returned a form which was signed by each participating family member. A total of 101 families were willing to participate. Two families cancelled their participation before the self-observation started, two families stopped participating because of computer problems, and one family had problems understanding the language. The remaining 96 families (314 persons) completed the one-week self-observation.

Data for the self-observation was collected between November 1998 and April 2000. We ensured that the self-observation took

place during a normal family week and not during holidays or other exceptional events. Interviewers visited each participating family at home to train the family members in self-observation with the help of the handheld computer, and also gave them various questionnaires to complete.⁵ The one-week self-observation period started the following day. Adolescent participants received a coupon for a CD with a value of about 30 SFr at the end of their self-monitoring task. All the families took part in a raffle for a holiday prize with a value of about 3500 SFr. In addition, they were given feedback about results during their observation week.

For this article only the data of the parents were used, to answer our research questions. The following analyses are based on subsamples of 95 couples (One couple had to be excluded, because the partners had been separated.)

A total of 55 couples spoke German, 39 couples spoke French, and one couple was bilingual. A total of 84 couples were in their first marriage, and on average had lived together for 20.6 years ($SD = 4.5$). The other couples had on average lived together for 10.3 years ($SD = 6.6$). The husbands were between 33 and 64 years old ($M = 47$, $SD = 5.95$). Their academic level was above average for Switzerland (52% had a College or University degree). Eighty-four husbands were in full time employment, eight worked almost full time (more than 30 hours a week), one worked less than 30 hours, one was unemployed, and one was retired. The wives were between 36 and 64 years old ($M = 44.8$, $SD = 5.46$). Wives' academic level was significantly lower than husbands' (Wilcoxon test: $z = 4.75$; $p < 0.001$), since only 20 percent had a College or University degree. Twelve women were in full time employment, 61 were employed part-time, one was unemployed and 21 were housewives.

Measures

Daily life measures (FASEM-C-Variables)

Six times a day over the course of one week each spouse answered various questions, that were presented by the handheld computer.

Own feelings. The question "How do you feel at the moment?" was asked to measure the spouses' own feelings, and a list of the following bipolar adjective pairs was then presented: (1) physically

not well – physically well, (2) without energy/tired – full of energy, (3) tense – relaxed, (4) dissatisfied – satisfied, (5) stressed – at ease (6) sad/depressed – happy, (7) anxious/concerned – confident, (8) angry – peaceful.⁶ Each item had to be rated on a six-point scale (e.g., very dissatisfied, dissatisfied, rather dissatisfied, rather satisfied, satisfied, very satisfied).

Factor Analysis, repeated for each of the 42 observations, showed a stable two-factor solution: a physical state factor, comprising the items “physically not well – physically well” and “without energy/tired – full of energy”, and an emotional state factor that included all of the other items. These factors were highly correlated. Therefore, a one-factor solution would also be appropriate (see Perrez et al., 2000; Wilhelm, in press). Although the items could be combined into one or two factors, we preferred to present the results at the item level. We did this despite the disadvantages of this procedure, such as a lower reliability of single items and redundancy in the results presented. The advantage of presenting single items is that we capture information about differences in the judgment of the various emotional and physical state qualities.

Own setting. The question “Where are you at the moment?” could be answered by selecting one of the following categories: (1) at home, (2) at work, (3) at someone’s house, (4) in a public place, (5) in transit, (6) somewhere else. As there was only a small number of responses in the last four categories, they were combined into one category.

Presence of other people. If the participants’ responses indicated that they were not alone, they were offered a list of each family member, and a selection of non-family alternatives. The presence of each person present was then rated on a six-point scale from very pleasant to very unpleasant.

Estimates of the partner’s setting. If the spouse did not indicate that the partner was present, the next question asked was “Where do you think your partner is at the moment?”. The following categories were presented: (1) at home, (2) at work, (3) at friends, (4) somewhere else, (5) I don’t know. Categories 3 and 4 were also combined into one category.

Judgment of the partner's feelings. The question "How do you think your partner is feeling at the moment?" was answered from the same menu of items as for the spouses' own feelings.⁷

Questionnaire data

Relationship satisfaction was measured with the *Relationship Assessment Scale* (RAS, Hendrick, 1988).⁸ The RAS consists of seven items that are rated on a five-point scale. Hendrick reported a Cronbach's alpha of 0.86 and a correlation of $r = 0.80$ with the Marital Adjustment Scale. We used a German version of the RAS (Sander and Böcker, 1993) and created a French translation ourselves. For the French and German speaking participants of our sample Cronbach's alpha reached a minimum of 0.86. The scores of husbands and wives were highly correlated ($r = 0.70$), which indicates a high correspondence between the spouses. It also justified the use of the couples' relationship satisfaction score (mean of the spouses RAS-scores). On average, couples were satisfied with their relationship ($M = 4.02$), and the differences between couples were rather small ($SD = 0.57$). French speaking couples had slightly lower RAS scores than German speaking couples ($M = 3.9$ vs. 4.1 , $p = 0.056$).

Data Preparation

We first eliminated stable differences between persons from all the self- and other-judgment variables, using centering on the individual mean. Then we selected observations in which the spouses were either at work or at home. For each person we computed the mean of their own feelings over the selected observations and the corresponding judgments about the feelings of their partner (using the variables centered around the individual mean). Finally we applied a procedure to correct reliability differences in the individual aggregate scores. In the following paragraphs we are going to describe each step in detail.

Eliminating between-person variation through individual mean centering

Each self-report of feelings in our data set is composed of at least three components: (a) stable characteristics of the judge (e.g.,

his or her general mood, personality, response style), (b) unstable characteristics of the current situation (e.g., setting, time of the day) and (c) unsystematic fluctuations (error). The judgments of the partner's feelings can further be divided into four components, as stable characteristics of the partner (e.g., the partner's general mood, personality) must also be taken into account beside the three components described above.

If we only computed setting-specific means for the self- and other-judgments from the raw data, we would be ignoring its composite character. We would also get measures in which the participants' ability to accurately judge the feelings of their partner in specific settings would be confounded with the accuracy of judging their partner's general feelings during the observation week (see Cronbach, 1955; Kenny, 1994).

We therefore felt it necessary to separate effects that are due to differences between situations from effects that are due to stable differences between persons. We did this by centering each variable on the individual mean.⁹ After centering, the self-judgment of the item "sad/depressed – happy", for example, indicates how sad or happy a spouse feels at the precise moment of a certain observation, relative to his or her mean level of happiness during the observation week. Also, after centering, the corresponding other judgment variable indicates the spouse's opinion about how happy or sad his or her partner is at the moment of a certain observation, relative to how happy the spouse thinks his or her partner feels in general during the observation week.

Criteria for the selection of observations

Observations within a particular setting were selected according to the criteria listed in Table I.

s1: Husbands are at work and wives think that their husband is at work. We selected all the observations in which husbands reported that they were at work. For each husband we calculated the mean of these observations for both his own feelings and his judgments about the feelings of his wife (note that all variables were centered around the individual mean). We then selected all the observations in which wives reported that they thought their husband was at work. As before, we calculated the intra-individual means of these

observations. The aggregated self- and other-judgments of the husbands and the wives were then matched together.

s2: Wives are at work and husbands think that their wife is at work. Intra-individual means for the wives were calculated from all observations in which they reported being at work. The corresponding means for the husbands were calculated from all the observations in which the husbands thought that their wife was at work. The reduction in sample size is due to the fact that 22 women were not employed at all and most of the others were employed part-time, some for only a couple of hours per week.

As Table I shows, the number of observations that matched the criteria was different for men and women. In the procedure we have used so far, the aggregated scores for the husbands do not rely on exactly the same set of observations as the aggregated scores for the wives. Thus, other variables that also have an influence on the judgment of feelings, such as the time of day (Wilhelm, 2001), are not the same for both partners within a couple. The fewer observations that are available to compute the individual mean scores, the more such other factors might influence the measures. The most probable consequence of this is that the observed interpersonal correlations (accuracy-, real similarity-, reciprocity correlations) are weaker than the true correlations. On the other hand, the advantage of this procedure is that no information had to be dropped.

s3: Husbands are at work and think that their wife is at home; wives are at home and think that their husband is at work. When the husbands are at work (s1), their judgments about their wife's feelings are difficult to interpret. They might think that their wife is in different settings and therefore their judgments about their wife's feelings are not setting-specific. For that reason husbands' observations were chosen when they were at work and thought that their wife was at home. For the wives the corresponding observations were chosen: Wives were at home, thinking that their husband is at work. To hold potential confounding effects constant, only these observations were selected in which all conditions for both wives and husbands were true at the same time. As Table I shows, this reduced the available number of observations

TABLE I
Criteria to select observations for the analysis

Type of situation to be analyzed	Husbands report that:	Wives report that:
(s1) H:w; W(H:w) Husbands are at work; wives think that their husband is at work $n_c = 90$	H:w They are at work $n_p = 92, n_s = 836$	W(H:w) Their husband is at work $n_p = 92, n_s = 1004$
(s2) W:w; H(W:w) Wives are at work; husbands think that their wife is at work $n_c = 59$	H(W:w) Their wife is at work $n_p = 65, n_s = 336$	W:w They are at work $n_p = 64, n_s = 306$
(s3) H:w \cap H(W:h) \cap W:h \cap W(H:w) Husbands are at work <i>and</i> think that their wife is at home; <i>and</i> vice versa $n_c = 80, n_s = 304$	H:w \cap H(W:h) They are at work, <i>and</i> their wife is at home $n_p = 88, n_s = 453$	W:h \cap W(H:w) They are at home, <i>and</i> their husband is at work $n_p = 89, n_s = 628$
(s4) H:h \cap H(W:h) \cap W:h \cap W(H:w) Husbands are at home, <i>and</i> think that their wife is at home, <i>but</i> they are not together; <i>and</i> vice versa $n_c = 67, n_s = 205$	H:h \cap H(W:h) They are at home, <i>and</i> they think that their wife is at home, <i>but</i> they are not together $n_s = 358, n_p = 83$	W:h \cap W(H:h) They are at home, <i>and</i> they think that their husband is at home, <i>but</i> they are not together $n_s = 279, n_p = 74$

W = Wives, H = Husbands, w = at work, h = at home, W(. . .) = wives thinking, H(. . .) = husbands thinking, \cap = logical and, n_s = number of situations, n_p = number of persons, n_c = number of couples.

substantially (from 628 observations for women and 453 observations for men to 304 observations for couples). In this instance, to avoid confounding effects in the interpretation, we accept this loss of information.

s4: Husbands are at home and record that their wife is at home, and wives are at home and record that their husband is at home. In this situation both spouses were at home, but although they were not together they both thought that their partner was also at home. As in situation 3, we only selected observations that fit the conditions for both husbands and wives at the same time.

Reliability adjustment of the aggregated scores

As can be inferred from the figures given in Table I, the number of observations which fit each criterion varies according to the type of setting and for the persons involved. For the first situation (men at work) the individual means are based on the highest number of observations (9 percent of men and 2 percent of women reported less than four observations). In the second situation (wives at work), 40 percent of the spouses had fewer than four observations that matched the selection criteria; in situation three (husbands at work, wives at home) the figure was 59 percent and in situation four (both at home, but not together) 69 percent.

A consequence of the varying number of observations is that the reliability of the individual setting-specific means also varies between persons and between the selected settings. Therefore, we multiplied each participant's setting-specific mean by its reliability estimate to correct it for potential errors. The less reliable the individual's setting-specific mean was, the stronger was the correction. As a result of this correction procedure, the individual's setting-specific mean is placed closer to the individual's overall mean (due to the centering procedure, the overall mean for each person is zero).

The reliability of each aggregated measure can be estimated applying the following formula (Snijders and Bosker, 1999, p. 26):

$$\text{reliability of person } j\text{'s score } (rel_j) = \frac{\sigma^2_{(\text{between persons})}}{\sigma^2_{(\text{between persons})} + \sigma^2_{(\text{within persons})} / n_j (\text{observations per person})}$$

To obtain estimates of the variation within and between persons for each variable we used multilevel analysis.¹⁰ We then applied the formula above to correct each person's aggregated score for measurement error.

If a score is based on only one observation ($n_j = 1$) the reliability is equal to the intra class correlation coefficient (*ICC*), that informs us about how similar two observations within the same person are on average. The *ICC*'s for the eight self-rating variables were the lowest for being tired – full of energy ($ICC = 0.32$) and the highest for being concerned – confident ($ICC = 0.49$; Mean *ICC* over the eight items = 0.40). The *ICCs* for the judgment of the partner's emotional state were slightly higher (Mean *ICC* over the eight items

= 0.44), ranging from 0.37 for being tense – relaxed to 0.51 for being concerned – confident. These coefficients indicate that the larger amount of variation in the variables is due to situational differences. Thus, they demonstrate that our measures are sensitive enough to capture short-term changes. On the other hand, the amount of stable between-person-variation is large enough to reach a satisfying reliability of about 0.80 when the aggregated score is based on four observations, if the *ICC* is high ≈ 0.49 , and up to nine observations if the *ICC* is low ≈ 0.32 .

Statistical Analyses

First we explored intra- and inter-individual mean differences in the aggregated, reliability-adjusted scores with *t*-tests for paired samples. Then, for every variable, we described the relationship between each of the four perspectives (self- and other-judgment of husbands and wives) with conventional product moment correlations. We used Structural Equation Modeling (SEM) to obtain measures of *direct accuracy* that are controlled for assumed similarity. We also used SEM to test the moderating effects of relationship satisfaction on accuracy and assumed similarity. Detailed descriptions of the statistical analyses are given in each of the results sections.

Because we were mainly interested in the pattern of results, we did not adjust the alpha level for the amount of tests computed. A better protection against a false rejection of the null hypothesis (alpha error) is only possible if one increases the risk of neglecting effects that really exist in the population (beta error). Power analysis showed that the risk of neglecting an existing medium effect (beta error) was at least three times higher than the risk of assuming a non-existing effect to be true (alpha error).¹¹

RESULTS

Intra- and Inter-Individual Mean Differences

Before presenting the main results, we will now give a detailed description of the data, while exploring mean differences between situations and mean differences between the spouses' self- and other-judgments.

Do spouses feel different when they are in different settings? Results concerning mean differences in the self-judgments of the spouses

We first looked at situation-specific differences in the feelings of the spouses by computing paired sample *t*-tests to compare the husbands' feelings at work (s1 and s3) with their feelings at home (s4). For the wives we tested differences between situation two (when they were at work) and situations three and four (when they were at home). We also tested the differences between the two situations at home. The sample size changed for each comparison (s1 vs. s4: $n = 64$, s2 vs. s4: $n = 46$, s3 vs. s4: $n = 58$; s2 vs. s3: $n = 50$). In a second analysis, we also used paired sample *t*-tests to look for differences between husbands and wives. We highlight differences if the *t*-test reached the level of significance ($p < 0.05$, two-tailed). As a measure of effect size we report Cohen's d_z (mean difference/*SD* of the differences; Cohen, 1988, p. 48).¹² Means and standard deviations for each variable based on all the available cases are shown in Table II.

When husbands were at work (s1 and s3) most of their feelings were below average. The items indicating stress and tension (stressed – at ease, tense – relaxed) were the most negative ones. In contrast, the energy level (tired – full of energy) was higher than average. *When husbands thought that their wife was at work (s2)*, all the items deviate from zero in the same direction as in situation one. However, the deviations were smaller, especially for tension and stress.

When husbands were at home (s4), the pattern of their feelings changed: Tension and stress returned to average levels, and were significantly better than when they were at work (s1 vs. s4/s3 vs. s4: for tension $d_z = 0.49/0.50$, $p < 0.001$, for stress ($d_z = 0.34/0.38$, $p < 0.01$). On the other hand, the husbands felt more tired at home than at work ($d_z = 0.51/0.35$, $p < 0.01$). This difference might be explained by the fact that husbands were most often at home in the mornings and evenings, when their energy levels were rather low due to their circadian rhythm. Physical well-being was also below the average, but not significantly worse than during work ($d_z = 0.21/0.14$, $p \geq 0.94$). No significant differences were found for the other variables.

Comparing the *wives' feelings at work with their feelings at home* (see Table II), we found that both the energy level and the physical well-being were substantially higher at work (s2 vs. s3/s2 vs. s4: for energy $d_z = 0.91/0.38$, $p < 0.01$; for physical well-being $d_z = 0.31/0.34$, $p < 0.05$). Our results show quite different patterns of feelings between the two situations at home (s3 vs. s4). During times their husband was at work, wives felt less satisfied ($d_z = 0.29$, $p < 0.05$), and reported more anger ($d_z = 0.32$, $p < 0.05$), more stress ($d_z = 0.36$, $p < 0.01$) and more tension ($d_z = 0.47$, $p < 0.001$) than during times their husband was at home, but not with them. One explanation for these differences is that, in situation three, wives were more occupied with working, either for the job, the family or for somebody else.¹³ However, the energy level was lower in situation four than in situation three ($d_z = 0.29$, $p < 0.05$), probably due to the circadian rhythm.

Do husbands feel different from wives?

Compared with their husband, wives had significantly less stress ($d_z = 0.34$, $p < 0.01$), tension ($d_z = 0.36$, $p < 0.001$), anger ($d_z = 0.22$, $p < 0.05$) and concerns ($d_z = 0.25$, $p < 0.05$) in situation one, when their husband was at work. In this situation wives were either at home (61%) or in a non-work setting (18%). The direct comparison of the husbands' feeling at work with the wives' feelings at home (s3) showed that the husbands experienced more stress ($d_z = 0.24$, $p < 0.05$). They also tended to experience more tension ($d_z = 0.22$, $p = 0.054$).

In situation two, when wives were at work and this was where their husband thought them to be, we found no significant differences between the wives' and husbands' feelings ($d_z \leq 0.24$, $p \geq 0.07$). This result might be due to the fact that while wives were at work husbands were usually also at work (59%). A direct comparison of the husbands' and wives' feelings at work did not reveal significant differences.¹⁴ In situation four, when both partners were at home, husbands and wives had quite similar feelings, and we found no significant differences ($d_z \leq 0.21$, $p \geq 0.09$).

TABLE II

Means and standard deviations of the self- and other-judgment of feelings in different settings

		Husbands				Wives			
		Self-judgments		Partner-judgments		Self-judgments		Partner-judgments	
	N_c	M	SD	M	SD	M	SD	M	SD
<i>s1: husbands were at work</i>									
Unsatisfied – satisfied	91	-0.08	0.26	-0.02	0.22	-0.02	0.20	-0.08	0.22
Concerned – confident	91	-0.07	0.25	-0.02	0.23	0.01	0.21	-0.06	0.20
Angry – peaceful	91	-0.08	0.20	-0.05	0.21	-0.03	0.21	-0.11	0.23
Sad – happy	91	-0.04	0.22	-0.02	0.22	-0.02	0.21	-0.06	0.18
Stressed – at ease ^a	86	-0.20	0.33	-0.11	0.26	-0.08	0.22	-0.23	0.31
Tense – relaxed ^a	90	-0.21	0.30	-0.12	0.24	-0.09	0.24	-0.29	0.32
Tired – full of energy ^a	90	0.10	0.27	0.10	0.25	0.12	0.21	0.09	0.28
Physically not well	91	0.04	0.26	-0.02	0.24	0.02	0.21	-0.05	0.20
– ph. well									
<i>s2: wives were at work</i>									
Unsatisfied – satisfied	59	-0.04	0.22	-0.03	0.23	-0.03	0.32	-0.11	0.22
Concerned – confident	59	-0.02	0.24	0.00	0.30	0.00	0.32	-0.04	0.27
Angry – peaceful	59	-0.07	0.24	-0.04	0.22	-0.08	0.38	-0.07	0.21
Sad – happy	59	-0.03	0.23	-0.03	0.23	0.03	0.33	-0.05	0.22
Stressed – at ease ^a	55	-0.11	0.29	-0.15	0.36	-0.08	0.42	-0.14	0.32
Tense – relaxed ^a	58	-0.06	0.26	-0.20	0.31	-0.17	0.42	-0.14	0.28
Tired – full of energy ^a	58	0.09	0.26	0.08	0.31	0.18	0.27	0.07	0.28
Physically not well	59	0.02	0.24	-0.05	0.26	0.05	0.26	-0.06	0.22
– ph. well									
<i>s3: husbands were at work and wives were at home</i>									
Unsatisfied – satisfied	80	-0.09	0.26	-0.05	0.25	-0.08	0.25	-0.08	0.21
Concerned – confident	80	-0.06	0.29	-0.05	0.28	-0.02	0.27	-0.06	0.24
Angry – peaceful	80	-0.08	0.23	-0.06	0.24	-0.06	0.25	-0.11	0.22
Sad – happy	80	-0.04	0.23	-0.08	0.26	-0.06	0.26	-0.07	0.17
Stressed – at ease ^a	76	-0.21	0.37	-0.11	0.33	-0.11	0.29	-0.23	0.39
Tense – relaxed ^a	79	-0.21	0.33	-0.08	0.31	-0.11	0.28	-0.27	0.34
Tired – full of energy ^a	79	0.06	0.31	0.05	0.32	0.03	0.26	0.06	0.30
Physically not well	80	-0.01	0.29	-0.05	0.33	-0.02	0.28	-0.06	0.27
– ph. well									

TABLE II

Continued

	<i>N_c</i>	Husbands				Wives			
		Self-judgments		Partner-judgments		Self-judgments		Partner-judgments	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<i>s4: both were at home</i>									
Unsatisfied – satisfied	67	−0.04	0.33	−0.03	0.28	−0.01	0.25	−0.01	0.35
Concerned – confident	67	−0.02	0.29	−0.07	0.34	−0.07	0.32	0.03	0.38
Angry – peaceful	67	−0.01	0.30	0.02	0.25	0.05	0.28	0.04	0.34
Sad – happy	67	−0.06	0.29	−0.06	0.33	0.01	0.28	−0.02	0.34
Stressed – at ease ^a	63	0.00	0.39	0.00	0.39	0.02	0.32	0.09	0.42
Tense – relaxed ^a	66	0.01	0.30	0.01	0.31	0.03	0.33	0.12	0.37
Tired – full of energy ^a	66	−0.14	0.30	−0.11	0.33	−0.10	0.33	−0.09	0.38
Physically not well	67	−0.09	0.33	−0.04	0.38	−0.03	0.31	−0.06	0.36
– ph. well									

The data for which means and standard deviations are computed are the situation-specific, intra-individual means, from which the general mean of each person has been removed. Scores were adjusted for reliability differences.

N_c = number of couples; ^asee endnote 7 for an explanation of the lower sample size in these items.

Do spouses think that their partner feels better or worse than they feel themselves? Results concerning mean differences in the judgments of spouses about their partner

We computed intra-individual comparisons first for the husbands and then for the wives using paired sample *t*-tests. This enabled us to determine whether spouses think that their partner feels better or worse than they do at any given moment. We also made inter-individual comparisons between the spouses' other-judgments and their partner's self-judgments. This enabled us to discover whether spouses tend to over- or underestimate their partner's feelings.

What did wives think their husband was feeling? In general, wives imagined that while at work their partner felt worse than they did. In situation one, wives thought that, in relation to their own feelings, husbands were less satisfied ($d_z = 0.22$, $p < 0.05$) and less confident ($d_z = 0.28$, $p < 0.01$), and felt more anger ($d_z = 0.30$, $p < 0.01$), stress ($d_z = 0.46$, $p < 0.001$) and tension ($d_z = 0.63$,

$p < 0.001$). They also overestimated the tension ($d_z = 0.23$, $p < 0.05$) and underestimated the physical well-being their partner actually experienced at work ($d_z = 0.36$, $p < 0.001$). In situation three, the wives' judgments of their husband's feelings were almost the same as in situation one. However, the pattern of results changes slightly in relation to their own feelings. Differences between the wives' own feelings and their judgments of their partner's feelings were only significant for stress ($d_z = 0.26$, $p < 0.05$) and tension ($d_z = 0.35$, $p < 0.01$). Furthermore, their estimates did not deviate significantly from their partner's feelings ($d_z \leq 0.17$, $p \geq 0.142$).

When the wives were at work (s2), they thought that their husband had less energy ($d_z = 0.38$, $p < 0.01$) and felt physically less well ($d_z = 0.34$, $p < 0.05$) in comparison with themselves, and they underestimated the men's physical well-being ($d_z = 0.27$, $p < 0.05$). In situation four, when both partners were at home, wives assumed that their husband felt more confident ($d_z = 0.25$, $p < 0.05$) and more relaxed ($d_z = 0.29$, $p < 0.05$) than they did. However, wives overestimated the level of relaxation their partner felt in this situation ($d_z = 0.28$, $p < 0.05$).

What did husbands estimate their wife's feelings to be? In situation one, when the husbands were at work, they thought that their partner felt more satisfied ($d_z = 0.24$, $p < 0.05$), less stressed ($d_z = 0.27$, $p < 0.05$) and less tensed ($d_z = 0.34$, $p < 0.01$) than the husbands did at the same moment. In situation three, when wives were at home while husbands were at work, husbands thought that their wife felt more relaxed than they did ($d_z = 0.35$, $p < 0.01$). In contrast, husbands thought that their wife felt more tension at work (s2) than they felt at the same time ($d_z = 0.36$, $p < 0.01$). When both partners were at home (s4), husbands believed that they and their wife felt about the same.

The husbands' judgments about the wives' feelings were quite close to the feelings the wives reported. The only exception was the husbands' believe that their wife felt physically worse at work than the women actually felt ($d_z = 0.37$, $p < 0.01$).

How Accurate Are Spouses in Judging Their Absent Partner's Feelings? How Much Similarity Do They Assume?

Results based on product moment correlations

To obtain measures of the basic indicators of interpersonal perception and thus get a first answer to our research questions, the four perspectives (self- and other-judgment of husbands and wives) were correlated with each other. The correlation coefficients and the medians of the coefficients which belong to one situation (*median r*) are reported in Table III.

Real similarity. Similarity in the spouses' feelings was rather low (Table III, column 1) and differences between the situations were small, as can be seen by the medians of the correlation coefficients. However, in situation four, feelings of tension, stress and happiness were correlated substantially ($r \geq 0.33$). Some similarity in stress levels also existed consistently across all the situations.

Reciprocity. Reciprocity correlations (Table III, column 2) differed from situation to situation. When wives were at work the coefficients were the highest, indicating that in this situation husbands and wives thought in quite similar ways about each other's feelings (median $r = 0.32$). In situation three, when husbands were at work and wives were at home, the reciprocity coefficients of the different items were the lowest and varied around zero (median $r = -0.02$), indicating that they thought differently about each other's feelings.

Assumed similarity. In general, husbands assumed a moderate to high amount of similarity between their own and their partner's feelings (Table III, column 3), and consequently they overestimated the actual similarity. A consistent exception was in their judgment of physical well-being, in which they did not assume similarity between their own and their wife's state. For husbands, the assumed similarity was the highest when both partners were at home (s4). It was rather low when they thought that their wife was either at home (s3) or at work (s2).

Situation-specific differences in the amount of assumed similarity were much larger for wives than for husbands. As Table III (column 4) shows, wives did not assume any similarity between

TABLE III
Product moment correlation coefficients between the measures

	Real similarity	Reciprocity	Assumed similarity		Accuracy	
			Husbands	Wives	Husbands	Wives
<i>s1: husbands were at work</i>						
Unsatisfied – satisfied	0.00	0.19	0.46***	0.24*	0.09	0.24*
Concerned – confident	0.10	0.13	0.44***	0.30**	−0.07	0.44***
Angry – peaceful	0.16	0.06	0.32**	0.20	0.02	0.36***
Sad – happy	0.19	0.16	0.41***	0.25*	0.17	0.43***
Stressed – at ease	0.23*	−0.03	0.34**	0.32**	0.09	0.44***
Tense – relaxed	0.15	0.27**	0.50***	0.37***	0.26*	0.41***
Tired – full of energy	0.14	−0.07	0.34***	0.03	−0.11	0.19
Physically not well – ph. well	−0.15	−0.13	0.12	−0.30**	0.21*	0.47***
Median <i>r</i>	0.15	0.09	0.38	0.25	0.09	0.42
<i>s2: wives were at work</i>						
Unsatisfied – satisfied	0.02	0.22	0.19	0.26*	0.22	0.18
Concerned – confident	0.04	0.40**	0.28*	0.27*	0.28*	0.13
Angry – peaceful	0.16	0.35**	0.33*	0.29*	0.21	0.10
Sad – happy	0.27*	0.40**	0.39**	0.23	0.15	0.20
Stressed – at ease	0.27*	0.29*	0.45***	0.33*	0.49***	0.34***
Tense – relaxed	0.10	0.45***	0.11	0.26*	0.40**	0.26*
Tired – full of energy	0.22	0.16	0.38**	0.50***	0.16	0.27*
Physically not well – ph. well	−0.03	−0.03	0.13	0.05	0.35**	0.20
Median <i>r</i>	0.13	0.32	0.30	0.27	0.25	0.20

TABLE III
Continued

	Real similarity	Reciprocity	Assumed similarity		Accuracy	
			Husbands	Wives	Husbands	Wives
<i>s3: husbands were at work and wives were at home</i>						
Unsatisfied – satisfied	–0.02	0.02	0.42***	–0.10	0.07	0.23*
Concerned – confident	–0.08	–0.04	0.20	0.02	–0.01	0.21
Angry – peaceful	0.23*	–0.07	0.28*	0.02	0.11	0.17
Sad – happy	0.16	0.01	0.30**	0.00	0.16	0.17
Stressed – at ease	0.23*	0.09	0.19	0.11	0.19	0.26*
Tense – relaxed	0.08	0.21	0.41***	0.01	0.09	0.27*
tired – full of energy	0.19	–0.05	0.37***	0.02	0.21	0.30**
Physically not well – ph. well	0.05	–0.20	0.13	–0.01	0.32**	0.26*
Median <i>r</i>	0.12	–0.02	0.29	0.01	0.13	0.25
<i>s4: both were at home</i>						
Unsatisfied – satisfied	0.10	0.17	0.57***	0.43***	0.16	0.31**
Concerned – confident	–0.01	0.10	0.44***	0.44***	0.12	0.19**
Angry – peaceful	0.19	0.25*	0.53***	0.45***	0.26*	0.25*
Sad – happy	0.33**	0.18	0.52***	0.41***	0.36**	0.29*
Stressed – at ease	0.40***	0.17	0.46***	0.42***	0.33**	0.52***
Tense – relaxed	0.47***	0.14	0.37**	0.56***	0.18	0.35**
Tired – full of energy	0.19	0.11	0.39***	0.44***	0.25*	0.38***
Physically not well – ph. well	0.10	–0.04	0.07	0.26*	0.45***	0.60***
Median <i>r</i>	0.19	0.16	0.45	0.43	0.26	0.33

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$, two-tailed.

their own and their partner's feelings when they were at home and their husband was at work (s3). When they were at work (s2) and when they thought their husband was at work (s1), they assumed a modesty similarity. Assumed similarity was rather high when both spouses were at home, but not together (s4). However, the wives' judgment of their husband's physical well-being shows a deviation from this pattern.

Accuracy. The accuracy of the husbands' judgments of their partner's feelings varied over situations (Table III, column 5). When husbands were at work and wives were at home (s3) or in different settings (s1) accuracy was very low, except for the judgement of the wives' physical well-being, for which accuracy was small to moderate. In situation one a moderate and significant accuracy correlation could also be found for the judgment of tension. When their wife was at work (s2), husbands achieved moderate levels of accuracy in four out of eight items. This was the situation in which they were most accurate in judging their wife's level of stress and tension. Also when both spouses were at home (s4) husbands had moderate levels of accuracy in most of the items.

Wives had higher accuracy scores than husbands, except in situation two (Table III, column 6). When wives were at work and were asked to judge their partner's feelings, they did not know exactly where he was. They could not, therefore, take full advantage of setting-specific knowledge, which resulted in lower accuracy coefficients. In contrast, they were quite accurate at judging how their husband felt at work. Accuracy was the highest in situation one, in which for most items at least a moderate level of accuracy was achieved (median of $r = 0.42$). It was somewhat lower in situation three, but still five out of eight accuracy coefficients were significant. Wives also achieved moderate to high levels of accuracy when both partners were at home, but not actually with each other.

In summary these results provide some evidence for hypothesis H-1: Accuracy in judging the partner's feelings, when he or she was either at work or at home, was often higher than mere chance. This was especially true for wives, while for the husbands it was mainly true in two situations: Only when their wife was at work (s2) or when they were both at home but not together (s4) did they reach

accuracy levels which were higher than chance in most of the items. Thus, results also indicate different levels of accuracy for different situations.

In addition, results seem to provide some evidence for hypothesis H-5 that women tend to be more accurate than men. Because a direct comparison of husbands' and wives' accuracy correlations requires them not to be standardized separately (Kenny and Acitelli, 2001), the validity of these gender differences might be questioned. However, accuracy correlation coefficients that are standardized on the base of a pooled estimate of the partners' standard deviations are not much different (see Table V).

Hypothesis H-2, which predicts higher coefficients for assumed similarity than for accuracy, seemed to be compatible with the results for husbands, except for the judgment of physical well-being. For the wives, H-2 was only compatible with the results when they were at work (s2) or when both partners were at home (s4). Results contradict H-2 when wives judged their husband's feelings at work.

The conclusions concerning H-2, and H-5 were based only on descriptive comparisons. The next section presents significance tests which, after accuracy and assumed similarity coefficients are adjusted for each other, directly prove H-2 and H-5.

Results Based on Structural Equation Modeling

Almost 50 years ago, Cronbach (1955) and Gage and Cronbach (1955) criticized that accuracy scores might be inflated due to assumed similarity. The best way to deal with this problem is to separate the total accuracy correlation into that part which is due to assumed similarity (*indirect accuracy*) and that which does not rely on assumed similarity (*direct accuracy*). The assumed similarity coefficient can similarly be divided into two components, namely the *correctly-assumed similarity*, which is due to accurate perceptions of real existing similarity between spouses, and the *falsely-assumed similarity* that does not rely on a correct estimation of whatever similarity between spouses actually exists.

We used a model proposed by Kenny and Cook (1999, see also Kenny and Acitelli, 2001) to control the accuracy and assumed similarity effects for each other (see Figure 1). Two regression equations were computed. In one equation the husbands' judgment of

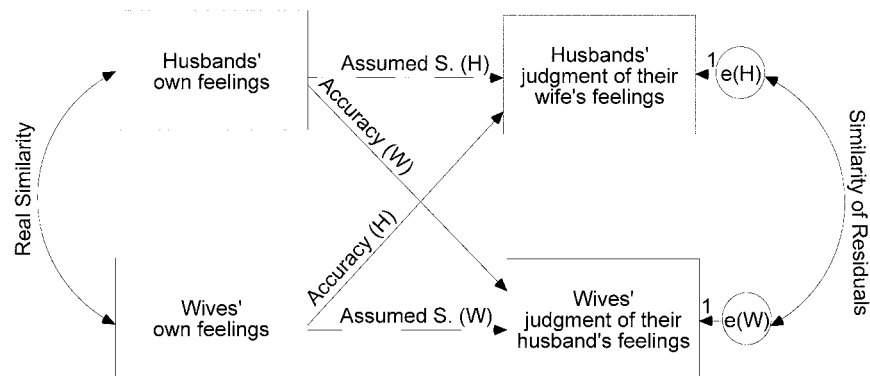


Figure 1. Model to obtain coefficients of *direct accuracy* and *falsely-assumed similarity* within the couple (Kenny and Acitelli, 2001). Notes: Assumed S = Assumed Similarity, e = Residual, (H) = Husbands, (W) = Wives.

their wife's emotional state was the criterion variable which was regressed to the wives' and husbands' own emotional state. Thus, the first coefficient is a measure of the *husbands' direct accuracy*, and the second coefficient is a measure of the *husbands' falsely-assumed similarity*. In the other equation the wives' judgment of their husband's emotional state was predicted with the husbands' emotional state (measure of the *wives' direct accuracy*) and with the wives' own emotional state (measure of the *wives' falsely-assumed similarity*).

To further model the interdependence between the spouses, both equations were estimated within one single analysis. Therefore, in addition to the four regression coefficients described above, correlations were estimated between the self-perceived emotional state of husbands and wives, and between the unexplained variances in the two outcome variables. Thus, the complete model shown in Figure 1 was saturated. This model can be estimated within a Structural Equation Modeling approach (SEM). We used the program AMOS 4.01.

The SEM approach also allowed us to test whether there are differences in the direct accuracy between husbands and wives. We compared the fit of a model in which the accuracy coefficients of husbands and wives were restricted to being equal to each other to the fit of another model in which both coefficients were free to vary. The difference in the fits of the two models follows a χ^2 distribution

with one degree of freedom. In the same way, we could test whether gender differences in the assumed similarity paths exist. To further test whether the accuracy coefficients were significantly different from assumed similarity coefficients, we compared the fit of a model in which accuracy paths of wives and husbands were set equal to their assumed similarity paths with the fit of the saturated model. This test had 2 *df*.

How accurate are partners when assumed similarity is controlled?

In general, the unstandardized falsely-assumed similarity coefficients (Table IV, column 1 and 2) were slightly smaller than the assumed similarity correlation coefficients reported in Table III. Only in situation four women's unstandardized falsely assumed similarity coefficients were bigger than the corresponding correlation coefficients. However, the pattern of results in Table III and Table IV was very similar.

Direct accuracy coefficient (Table IV, column 4 and 5) were not much different from the correlation coefficients reported in Table III (columns 5 and 6). However, in situation four, when both were at home the direct accuracy of the husbands' judgments of their wife's level of stress, tension and energy decreased below the significance level. A decrease of direct accuracy coefficients below the significance level could also be observed for the wives' judgment of their partner's tension, anger and sadness. These results indicate that a part of the accuracy observed in situation four was due to correctly-assumed similarity. We will examine this interpretation in more detail later.

Gender differences in the accuracy coefficients were only significant in situation one (see Table IV, column 6). When husbands were at work (s1), they were not at all accurate in judging their wife's concern, anger, stress or energy level (coefficients were close to zero or negative), whereas wives reached moderate levels of accuracy in judging the same items for their husband. Although direct significance tests of gender effects did not become significant in the other situations, accuracy coefficients for women were generally higher than for men in situation one, three and four (as the medians of the coefficients show), and were more often higher than would be expected through chance. Taken together, these results are in accordance with hypothesis H-5.

TABLE IV
Unstandardized regression coefficients and significance tests based on model comparisons

	Assumed similarity (AS)		Accuracy (Acc)		AS, Vs Acc. $\chi^2(2)^b$
	Husbands	Wives	Husbands	Wives	
<i>s1: husbands were at work</i>					
Unsatisfied – satisfied	0.39***	0.27*	0.10	0.21*	5.40
Concerned – confident	0.42***	0.25**	-0.13	0.34***	14.01***
Angry – peaceful	0.34***	0.15	-0.03	0.39***	7.79*
Sad – happy	0.40***	0.15	0.10	0.33***	6.10*
Stressed – at ease	0.26**	0.33*	0.01	0.37***	2.65
Tense – relaxed	0.39***	0.43***	0.20*	0.39***	2.25
Tired – full of energy	0.33***	0.01	-0.19	0.20	12.51**
Physically not well – ph. well	0.14	-0.23**	0.27*	0.34***	25.69***
Median <i>b</i>	0.36	0.20	0.05	0.34	
<i>s2: wives were at work</i>					
Unsatisfied – satisfied	0.19	0.18*	0.15	0.17	0.08
Concerned – confident	0.34*	0.22*	0.25*	0.13	0.73
Angry – peaceful	0.28*	0.15*	0.10	0.05	2.92
Sad – happy	0.37***	0.12	0.04	0.14	4.13
Stressed – at ease	0.42**	0.19*	0.35***	0.30*	0.41
Tense – relaxed	0.08	0.16	0.30***	0.25	2.69
Tired – full of energy	0.42***	0.49***	0.10	0.18	4.68
Physically not well – ph. well	0.16	0.05	0.35**	0.18	1.79
Median <i>b</i>	0.31	0.17	0.20	0.17	

TABLE IV
Continued

	Assumed similarity (AS)		Accuracy (Acc)		AS, Vs Acc. $\chi^2(2)^b$
	Husbands	Wives	Husbands	Wives	
<i>s3: husbands were at work and wives were at home</i>			$\chi^2(1)^a$	$\chi^2(1)^a$	
Unsatisfied – satisfied	0.41***	-0.08	12.70***	0.08	0.19*
Concerned – confident	0.19	0.03	1.22	0.01	0.18
Angry – peaceful	0.28*	-0.02	3.71	0.04	0.17
Sad – happy	0.32**	-0.02	5.57*	0.12	0.13
Stressed – at ease	0.13	0.08	0.10	0.17	0.26*
Tense – relaxed	0.37***	-0.02	5.57*	0.07	0.28*
Tired – full of energy	0.35***	-0.04	5.63*	0.18	0.29**
Physically not well – ph. well	0.13	-0.02	0.89	0.37**	0.24*
Median <i>b</i>	0.30	-0.02		0.10	0.21
<i>s4: both were at home</i>					
Unsatisfied – satisfied	0.47***	0.56***	0.27	0.12	0.29*
Concerned – confident	0.53***	0.52***	0.01	0.14	0.25
Angry – peaceful	0.43***	0.50***	0.24	0.15	0.19
Sad – happy	0.51***	0.42**	0.24	0.26*	0.20
Stressed – at ease	0.40***	0.34*	0.10	0.21	0.46***
Tense – relaxed	0.38**	0.57***	1.00	0.01	1.09
Tired – full of energy	0.39**	0.45***	0.13	0.19	0.50
Physically not well – ph. well	0.04	0.24*	1.38	0.54***	0.39**
Median <i>b</i>	0.41	0.48		0.17	0.64***

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$, two tailed.^aTest that the coefficients for men and women are the same.^bTest that the assumed similarity coefficients are equal to the accuracy coefficients.

Significant gender differences only appeared in situation one and three for the falsely-assumed similarity coefficients, indicating that coefficients were lower for women than for men (see Table IV, column 3). These results in addition to the medians of the coefficients in situation one, two and three, suggest that women tend to rely less on their own feelings when judging the feelings of their partner. The exception was situation four, when both partners were at home.

The tests for differences in magnitude of the falsely-assumed similarity- and the direct accuracy coefficients became significant for most items in situation one and four, and for a single item in situation three (see Table IV, column 7). However, the meaning of the significant results is quite different. When husbands were at work (s1 and s3), the tests for equality of the paths were significant, because for men, accuracy was not as high as assumed similarity, while for women it was the other way around. In situation four, when both were at home, the significant results indicated that assumed similarity was higher than accuracy for both partners. However, when judging the physical well-being of their partner, both husbands and wives were quite accurate without relying on their own physical state. This was also true in situation one.

In summary these results, plus the medians of the coefficients, suggest a modification of hypothesis H-2. Whether assumed similarity is higher than accuracy seems to depend on the situation, the item and the gender of the spouse.

Separation of accuracy correlations into direct accuracy and indirect accuracy

According to Kenny and Acitelli (2001), the standardized accuracy coefficients cannot be compared between partners if they were standardized separately for men and women. Therefore, while computing the basic model (Figure 1), we restricted the variance of the husbands' and wives' own feelings, and the variances of the two error terms, to be equal. Since the variances are the same, the standardized coefficients provided by this model are comparable across partners.

The overall accuracy correlation can be separated into direct accuracy and indirect accuracy which is due to correctly-assumed

similarity. The *indirect accuracy* coefficient is the product of the standardized assumed similarity coefficient and the actual real similarity correlation. When we add together the direct and indirect accuracy coefficient of a given variable, we obtain its overall accuracy correlation coefficient, which is comparable across partners. As seen in Table V, the overall accuracy coefficients are quite similar to the accuracy correlations reported in Table III, but they are not exactly the same.

Table V shows that substantial overall accuracy correlations were mainly due to direct accuracy. An exception was the judgment of tension when both partners were at home (s4), for which the indirect accuracy coefficients were higher than the direct accuracy coefficients for both spouses. This results was due to the high amount of real similarity in the partners' level of tension in this particular situation. In general, indirect accuracy coefficients greater 0.10 were rare, and could only be observed when real similarity was substantial (Table III).

Do Spouses Rely on Their Own Emotional Experience at Work When They Judge How Their Partner Feels at Work?

To judge the subjective experience of another person in a certain situation, we use our own experience in that situation as the reference point (Nickerson, 1999). If this is also true for judgments of the partner's feelings at work, these judgments should be substantially correlated to the judging spouse's own feelings when at work. Moreover, these correlations should even remain substantial when the judging spouse's own feelings in the judgment situation, in which he or she could be in any setting, are controlled.

To test this hypothesis (H-4), we performed simple regression analyses separately for husbands and wives. The husbands' judgments of their wife's feelings at work (s2) were a) predicted by the husbands' own feelings when their wife was at work (to control the assumed similarity in s2) and b) by the husbands' own feelings when they themselves were at work (s1). In the same way the wives' judgments of their husband's feelings at work (s1) were a) predicted by their own feelings when their partner was at work (s1) and b) by their own feelings when they were at work themselves (s2).

TABLE V

Separation of the overall accuracy correlation into direct accuracy and indirect accuracy

	Husbands			Wives		
	Overall accuracy	Direct accuracy	Indirect accuracy	Overall accuracy	Direct accuracy	Indirect accuracy
<i>s1: husbands were at work</i>						
Unsatisfied – satisfied	0.10	0.10	0.00	0.22	0.22	0.00
Concerned – confident	−0.09	−0.13	0.05	0.39	0.36	0.03
Angry – peaceful	0.02	−0.03	0.05	0.38	0.36	0.02
Sad – happy	0.18	0.10	0.08	0.39	0.36	0.03
Stressed – at ease	0.08	0.01	0.06	0.42	0.35	0.07
Tense – relaxed	0.25	0.19	0.06	0.42	0.36	0.06
Tired – full of energy	−0.13	−0.17	0.04	0.19	0.19	0.00
Physically not well	0.27	0.29	−0.02	0.38	0.35	0.03
– ph. well						
Median r/β	0.09	0.06	0.05	0.39	0.36	0.03
<i>s2: wives were at work</i>						
Unsatisfied – satisfied	0.19	0.19	0.00	0.21	0.21	0.00
Concerned – confident	0.26	0.24	0.01	0.14	0.13	0.01
Angry – peaceful	0.20	0.14	0.06	0.11	0.07	0.04
Sad – happy	0.16	0.05	0.12	0.23	0.19	0.04
Stressed – at ease	0.46	0.35	0.11	0.39	0.33	0.06
Tense – relaxed	0.36	0.35	0.01	0.32	0.30	0.02
Tired – full of energy	0.18	0.09	0.09	0.25	0.16	0.09
Physically not well	0.35	0.36	−0.01	0.20	0.20	0.00
– ph. well						
Median r/β	0.23	0.22	0.04	0.22	0.20	0.03
<i>s3: husbands were at work and wives were at home</i>						
Unsatisfied – satisfied	0.08	0.08	−0.01	0.22	0.22	0.00
Concerned – confident	−0.01	0.01	−0.02	0.19	0.20	0.00
Angry – peaceful	0.11	0.05	0.07	0.17	0.18	0.00
Sad – happy	0.18	0.13	0.05	0.15	0.15	0.00
Stressed – at ease	0.19	0.16	0.03	0.25	0.24	0.02
Tense – relaxed	0.09	0.07	0.03	0.27	0.27	0.00
Tired – full of energy	0.23	0.17	0.06	0.27	0.28	−0.01
Physically not well	0.35	0.34	0.01	0.23	0.23	0.00
– ph. well						
Median r/β	0.15	0.11	0.03	0.23	0.23	0.00

TABLE V

Continued

	Husbands			Wives		
	Overall accuracy	Direct accuracy	Indirect accuracy	Overall accuracy	Direct accuracy	Indirect accuracy
<i>s4: both were at home</i>						
Unsatisfied – satisfied	0.16	0.12	0.05	0.30	0.26	0.05
Concerned – confident	0.11	0.12	–0.01	0.21	0.21	–0.01
Angry – peaceful	0.23	0.15	0.08	0.27	0.18	0.09
Sad – happy	0.36	0.22	0.14	0.30	0.18	0.12
Stressed – at ease	0.34	0.20	0.14	0.50	0.39	0.11
Tense – relaxed	0.18	0.01	0.17	0.35	0.12	0.23
Tired – full of energy	0.24	0.17	0.07	0.40	0.33	0.07
Physically not well – ph. well	0.48	0.48	0.00	0.55	0.53	0.02
Median r/β	0.24	0.16	0.08	0.33	0.24	0.08

Coefficients of these analyses are presented in Table VI. They indicate that even when their own current feelings in the judgment situation (which are the actual assumed similarity correlations) were controlled, both spouses relied substantially on their own emotional experience at the workplace when they judged how stressed, concerned, and sad their partner felt at work. This was also true for the husbands' judgment of their wife's satisfaction at work, and for the wives' judgment of their husband's level of tension at work. These results were compatible with H-4.

However, the judgment of their partner's anger, energy level and physical well-being could not be predicted by the spouses' own feelings at work. The results concerning the judgment of the partner's anger at work suggest that one's own current anger plays a more important role. For husbands, results concerning energy levels can be interpreted in the same way. Results concerning the judgment of the partner's physical well-being at work suggest that neither the current own physical well-being nor the own physical well-being at the work place had a substantial influence.

TABLE VI

Prediction of the husbands' and wives' judgments of their partner's feelings at work with their own feelings when their partner was at work ($b1$), and their own feelings when they were at work themselves ($b2$)

	Husbands				Wives			
	Partner was at work		Themselves were at work		Partner was at work		Themselves were at work	
	b_1	β_1	b_2	β_2	b_1	β_1	b_2	β_2
Unsatisfied – satisfied	0.14	0.13	0.37	0.33*	0.33	0.32*	0.05	0.07
Concerned – confident	0.34	0.27*	0.45	0.31*	0.04	0.04	0.16	0.28*
Angry – peaceful	0.26	0.27 *	0.19	0.17	0.25	0.26	0.02	0.03
Sad – happy	0.37	0.34**	0.30	0.26*	0.08	0.09	0.18	0.32*
Stressed – at ease	0.22	0.19	0.46	0.40**	-0.06	-0.05	0.33	0.47**
Tense – relaxed	0.22	0.21	0.00	0.00	0.10	0.09	0.31	0.46***
Tired – full of energy	0.28	0.26*	0.15	0.11	0.01	0.00	0.16	0.16
Physically not well – ph. well	0.27	0.20	-0.11	-0.08	-0.19	-0.21	0.04	0.06
Median of b and β	0.26	0.24	0.25	0.21	0.06	0.07	0.16	0.22

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; two-tailed.

N was 62 for husbands and 61 for wives (except for the items stressed – at ease, tense – relaxed, tired – full of energy, in which it was lower, see endnote 7). N was slightly larger than reported for situation two in Table II because all available data were taken into account here, whereas for the analysis reported in Tables II to V data could only be used when both partners had records in the same kind of situation.

Looking at the overall pattern of results, there seems to be a difference in the judgment processes of husbands and wives, which is related to the level of current assumed similarity. For husbands, the standardized current assumed similarity coefficients (Table VI) only changed slightly compared to the corresponding correlation coefficients (Table III, s2, column 3), even when their own feelings at work were significant predictors. In contrast, wives' current assumed similarity coefficients almost disappeared when their own feelings at work were significant predictors (compare assumed similarity correlation coefficients in table III, s1, column 4, and standardized regression coefficients in Table VI).¹⁵

We performed a second set of regression equations to try and strengthen the interpretation that spouses use their own emotional experience in an equivalent situation when judging their partner's feelings at work. Instead of taking the judging spouses' feelings when they were at work, we used their feelings when they were at home (s4) as the second predictor. Therefore, if the spouses' own feelings at home are not related to their judgments of their partner's feelings at work, but their own feelings at work are, we would have an additional discriminant proof of our hypothesis.

Results for women were in line with our expectation that spouses' feelings at home did not predict the judgment of their partner's feelings at work (unstandardized coefficients were between -0.14 and 0.03 , none reached significance). Also, for men, most of the coefficients had low values (six of the unstandardized coefficients were between 0.02 and 0.12 and were not significant). However, two results were unexpected: Husbands' judgments of their wife's physical well being and their level of concern could be predicted with the husbands' own physical well being respectively concern at home ($b = 0.36, p < 0.01$ / $b = 0.30, p < 0.05$).

Although these two significant results do not confirm our expectations, the overall pattern of results does provide evidence that spouses' judgments of their partner's feelings at work are at least partly based on an inference process. In addition to their own feelings in the moment of judgments, men also take into account their specific emotional experience in the work setting (but not their feelings in the home setting) as a reference point to estimate their partner's levels of satisfaction, happiness and stress. For women, results more strongly confirm this situation-specific inference process (H-4). When they estimated how happy, concerned, stressed, and tensed their husband feels at work, they relied only on their own emotional experience at work. All in all a great part of the results are compatible with H-4.

Are Satisfied Couples More Accurate? Moderator Analysis

We extended the basic model (Figure 1) to ascertain whether satisfied couples are more accurate than less satisfied couples, or whether they tend to assume more similarity than less satisfied ones. Three further exogenous variables were introduced: The couple's relation-

ship satisfaction (mean of both spouses RAS-scores, z -transformed) and the products of the couple's relationship satisfaction with each spouse's own feelings. Additional paths were drawn from each of these variables to both outcome variables.

The moderating effect of relationship satisfaction on the husbands' direct accuracy is represented by the path from the wives' product term (couples RAS-score*wives' own feelings) to the husbands' judgment of their wife's feelings. A positive coefficient of this path would suggest that, in satisfied couples, husbands judge their wife's feelings more accurately than in non-satisfied couples. The moderating effect of relationship satisfaction on wives' direct accuracy corresponds to the path from the husbands' product term (couples RAS-score*husbands' own feelings) to the wives' judgment of their husband's feelings.

The moderating effect of relationship satisfaction on husbands' falsely-assumed similarity is represented by the path from the husbands' product term (couples RAS-score*husbands' own feelings) to the husbands' judgment of their wife's feelings. A positive coefficient of this path would suggest that, in satisfied couples, husbands tend to assume more similarity between their own and their wife's feelings. The corresponding moderating effect of relationship satisfaction on wives' assumed similarity is represented by the path from the wives' product term (couples RAS-score*wives' own feelings) to the wives' judgment of their husband's feelings.

Because all the exogenous variables were allowed to correlate with each other, our moderator effects model was a saturated one. A moderator effect was indicated when the coefficient of its corresponding path was significant ($p < 0.05$, two-tailed).

Results of the moderator analysis are presented in Table VII. For husbands, we found five significant moderating effects of relationship satisfaction on accuracy in situation one (Table VII, s1, column 3). These moderator effects were positive, indicating that an increase in the couples' relationship satisfaction was associated with an increase in the husbands' accuracy at predicting their wife's feelings. Since the couples' relationship satisfaction was z -transformed, the interpretation of the coefficients is straightforward. In many of these variables the moderating effects of relationship satisfaction on assumed similarity were also significant, but turned out to be

negative (Table VII, s1, column 1). Taken together, these effects indicate that in satisfied couples husbands assumed less false similarity and this might be why they were more accurate when judging their wife's feelings. These results were in line with H-6.

However, this interpretation only holds true for husbands in situation one. Results could not be replicated in the most similar situation, which is situation three, where we even observed a negative moderating effect on husbands' accuracy in judging their partner's tension. When husbands judged their wife's feelings at work (s2), all the moderator coefficients of relationship satisfaction on accuracy were negative, and one of these coefficients was significant: The husbands' accuracy in judging their wife's level of stress at work was lower when the relationship satisfaction was above average. These effects contradict H-6. There was no significant moderator effect of relationship satisfaction on husbands' accuracy in situation four.

Moderating effects of relationship satisfaction on husbands' falsely-assumed similarity were significantly negative for the judgment of their wife's energy level in situation two and four. In situation four, the judgment of the physical well-being was also significantly moderated.

For the wives, significant moderating effects indicate that wives in satisfied couples were more accurate in judging their husband's feelings at work, especially their level of stress (s1 and s3), tension (s1), and sadness (s3). Also, in situation four, when both partners were at home, the wives' judgment of their husband's tension and energy levels was higher when they were more satisfied. At the same time, satisfied wives relied less on their own feelings when they judged their husband's level of tension (significant negative moderating effects of relationship satisfaction on falsely-assumed similarity). One further moderator effect indicating a negative association between relationship satisfaction and the wives' accuracy was found for the judgment of their husband's anger in situation two. A negative moderator effect of relationship satisfaction on falsely-assumed similarity was also found for this item.

Summarizing the results, there seems to be some weak evidence for hypothesis H-6, which claimed that accuracy in judging the partner's feelings is related to marital satisfaction. However, results

TABLE VII

Unstandardized moderator effects of relationship satisfaction (RAS) on falsely-assumed similarity and direct accuracy

	Moderating effects of RAS on assumed similarity		Moderating effects of RAS on accuracy	
	Husbands	Wives	Husbands	Wives
<i>s1: husbands were at work</i>				
Unsatisfied – satisfied	−0.13	0.01	0.19*	0.11
Concerned – confident	−0.27**	−0.12	0.29**	0.08
Angry – peaceful	−0.27*	−0.16	0.24*	0.12
Sad – happy	−0.23*	−0.09	0.20*	0.06
Stressed – at ease	−0.18*	−0.22	0.28*	0.27**
Tense – relaxed	−0.06	−0.13	0.03	0.21*
Tired – full of energy	−0.26**	0.27	0.00	0.11
Physically not well – ph. well	−0.11	0.03	0.03	0.04
Median <i>b</i>	−0.21	−0.11	0.20	0.11
<i>s2: wives were at work</i>				
Unsatisfied – satisfied	−0.02	0.05	0.2	−0.21
Concerned – confident	−0.15	−0.02	−0.16	0.05
Angry – peaceful	−0.02	−0.16*	−0.11	−0.18*
Sad – happy	−0.14	0.07	−0.12	−0.10
Stressed – at ease	−0.07	0.01	−0.29**	0.10
Tense – relaxed	0.03	−0.08	−0.14	−0.06
Tired – full of energy	−0.45**	−0.04	−0.07	0.10
Physically not well – ph. well	0.08	0.04	−0.17	−0.06
Median <i>b</i>	−0.05	−0.01	−0.13	−0.06
<i>s3: husbands were at work and wives were at home</i>				
Unsatisfied – satisfied	−0.04	−0.05	0.10	0.08
Concerned – confident	−0.07	−0.03	0.07	0.10
Angry – peaceful	0.01	−0.11	−0.09	−0.01
Sad – happy	−0.10	−0.01	0.05	0.18*
Stressed – at ease	0.06	−0.13	0.05	0.25*
Tense – relaxed	−0.01	−0.04	−0.26*	0.14
Tired – full of energy	−0.01	0.19	−0.10	0.09
Physically not well – ph. well	−0.22	0.16	−0.01	−0.10
Median <i>b</i>	−0.03	−0.04	0.02	0.10

TABLE VII
Continued

	Moderating effects of RAS on assumed similarity		Moderating effects of RAS on accuracy	
	Husbands	Wives	Husbands	Wives
<i>s4: both were at home</i>				
Unsatisfied – satisfied	0.03	–0.11	–0.02	–0.12
Concerned – confident	0.06	0.03	0.12	0.15
Angry – peaceful	0.14	0.03	0.01	–0.02
Sad – happy	0.03	–0.09	0.00	0.20
Stressed – at ease	0.09	–0.18	–0.01	0.05
Tense – relaxed	–0.01	–0.26*	–0.05	0.47**
Tired – full of energy	–0.40**	0.08	0.16	0.33*
Physically not well – ph. well	–0.27*	–0.05	0.19	0.03
Median <i>b</i>	0.03	–0.07	0.01	0.10

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; two-tailed.

are not definitive, since there are also some moderating effects indicating the opposite relationship. Results confirmed hypothesis H-6 when wives judged their husband's feelings of stress and tension at work, and feelings of tension and energy at home. For the husbands, support for hypothesis H-6 was restricted to the accuracy of judging their wife's feelings in situation one, when husbands themselves were at work. Significant moderating effects of marital satisfaction on falsely-assumed similarity were all negative, indicating that falsely-assumed similarity decreases when marital satisfaction increases.

DISCUSSION

The aim of this paper was to examine how accurately spouses can judge their partner's feelings in everyday situations when they are apart, which we believe is a new topic in the field of research on accuracy of interpersonal perception and interpersonal sensitivity (Bernieri, 2001; Funder, 1999; Kenny, 1994).

Our focus was to answer the following key questions: How accurately can spouses estimate their partner's feelings, both at home

and at work? How far do spouses assume similarity between their own feelings and those of their partner? When they are estimating how their partner feels at work, do they simply project how they are feeling at that moment onto their partner, or do they rely instead on their own emotional experiences from similar situations? Are there gender differences in accuracy or assumed similarity? Are spouses more accurate, or do they assume more similarity, when they are satisfied with their marital relationship? While these may be basic questions in this field their specific application here, and the results we achieved, are unique.

Assessment of Spouses' Own Feelings and Their Judgment of Their Partner's Feelings in Normal Everyday Life

We used a computer-assisted self-monitoring procedure to answer our research questions (FASEM-C, Perrez et al., 2000). Our approach enabled us to assess how spouses were feeling, and how they thought their partner was feeling, in natural settings both at home and at work, during their normal everyday lives. We could therefore gather data which was not affected by recall-errors, giving our measures a high ecological validity (Bolger et al., 2003; Fahrenberg, 1996; Wilhelm and Perrez, 2001). Such recall-errors have been shown to occur when participants are asked to report retrospectively on their feelings (see Bolger et al., 2003; Fahrenberg et al., 2001; K  ppler and Rieder, 2001; Hank et al., 2001, for a more extensive discussion of mood-congruent memory effects, negative retrospective effects, and hindsight bias). This method, whereby data is collected directly from the participants' daily lives, is another means by which our study differs from previous accuracy research, which has been based either on questionnaires (e.g., Kenny and Acitelli, 2001) or on laboratory experiments (e.g., Ickes, 2001; Noller, 2001).

Selection of Settings to be Compared

We used different criteria to select observations for each participant, which were a compromise between theoretical considerations and the data available. First we chose observations where the partner was at work, and their spouse correctly assumed that this was so. This enabled us to ascertain how well spouses knew what their

partner was feeling at work. Then we selected observations where both spouses were at home at the same time, but were not together, to compare participants' knowledge about their partner at work and at home. Finally we compared situations where husbands were at work and wives were at home.

It would have also been interesting to choose other situations, such as when both partners were at work at the same time, when wives were at work and husbands were at home, or when spouses did not know where their partner was. Unfortunately, not enough observations matched these criteria for both spouses to justify including them.

Preparing the Data to Provide Pure Setting-Specific Measures

Separating measures into clearly interpretable components has been an essential prerequisite for accuracy research (Bernieri, 2001; Kenny, 1994; Kenny and Winquist, 2001) ever since the work of Gage and Cronbach (1955; also Cronbach, 1955).

For each item, we separated our data into a stable component and a situation-specific component by computing each spouse's mean over the whole observation period, and subtracting this mean from the corresponding raw data. This procedure of centering around each individual's mean eliminates the stable component that is due to either rating bias (the way a spouse uses the rating scale), longer lasting mood trends, or personality traits (e.g., neuroticism). As a result, the centered variables mainly contain situation-specific information.

Setting-specific means were then computed for each spouse. Because the number of observations which were reported in each setting varied from person to person, we adjusted the individual mean scores according to their reliability. All the results that are reported in this article are based on these separated, setting-specific, error-corrected, mean scores. We used these scores to compute accuracy coefficients which are indicators of the spouses' ability to predict how much their partner's feelings in a specific situation differ from how their partner feels on average.

Our data preparation follows a nomothetic analytic approach (Kenny and Winquist, 2001) where we measure how accurately spouses can estimate the intensity of their partner's feelings (e.g.,

sad–happy, stressed–at ease) in various settings. Centering allowed us to control for elevation accuracy as well as differential elevation accuracy in terms of Cronbach (1955).¹⁶ Because we did the analysis item by item, Cronbach's stereotype accuracy component is not relevant here. Thus, we obtained setting-specific measures of differential accuracy.

An alternative approach would have been an ideographic analysis, that would provide individual accuracy and assumed similarity measures for each spouse (Kenny and Winquist, 2001).

Mean Differences in the Spouses' Own Feelings Between Settings

In addition to our main research questions, we explored the mean differences between settings and between persons. Results concerning the feelings of the spouses at home and at work were in line with results of our First Family Project (Perrez et al., 1998; Wilhelm et al., 2000; Zbinden, 2003) and with other studies based on experience sampling (Larson and Richards, 1994) and ambulatory psychophysiological assessment (e.g., Myrtek et al., 1999). Husbands felt worse at work than they felt on average. Stress and tension were the most negative feelings, and differed significantly from the levels they experienced at home. Energy levels, which were higher at work than at home, were an exception that can partly be explained by the circadian rhythm. Larson and Richards (1994, pp. 238–239) reported very similar findings for American middle class husbands.

Women's feelings at work were not much different from their feelings at home, except that they felt more energetic and physically stronger when at work. Also these findings were quite similar to the findings of Larson and Richards (1994). In addition, we found more differences in women's feelings in the two different situations at home. Women felt worse when their husband was at work than when he was at home, indicating that they mainly have to deal with rather less pleasant events at home, when their husband is not present.

Mean Differences in Spouses' Judgments of Their Partner's Feelings

What did spouses think about each other's feelings at work or at home? Women believed that they felt better and were less stressed than their husband felt when he was working, and in fact tended to overestimate how tense and stressed their husband actually was at work. Conversely, when their husband was at home, women tended to overestimate how much he recovered from tension and stress.

Husbands also tended to think that their wife felt more stress and tension at work, and underestimated her actual physical well-being. Husbands also thought that they experienced higher levels of stress and tension when they were at work than their wife did at home.

Although these differences were rather small, they show how spouses tend to over- or underestimate the setting-specific feelings of their partner. To summarize, these results might be reflecting cultural and in-group stereotypes (Prentice and Miller, 2002) as well as the setting-specific stereotypes or cognitive schemata that the demands of work cause stronger negative feelings of stress and tension than the demands of other settings (e.g., Myrtek and Foerster, 2001; Myrtek et al., 2001). However, there is probably a certain amount of truth in these stereotypes.

The differences between husbands and wives may as well be an indicator of different cognitive schemata about each others' emotional realities, which seem to be divergent (Larson and Richards, 1994).

Comparing Different Components of Accuracy

We used two distinct analytical methods, from which we could develop different measures for assumed similarity and accuracy.

We applied a correlation approach to calculate coefficients of the overall assumed similarity and the overall accuracy. Cronbach (1955) and Gage and Cronbach (1955) criticized such overall accuracy correlation coefficients. They argued that the accuracy correlation might have been inflated by assumed similarity and might not therefore be a true indicator of a good ability to make judgments.

We used a Structural Equation Model, proposed by Kenny and Cook (1999) and Kenny and Acitelli (2001), to separate the overall accuracy correlation into the part that was due to correctly-assumed similarity (indirect accuracy) and the direct accuracy that did not rely on assumed similarity. Because the amount of indirect accuracy depends on the amount of real similarity, which was low in our data, indirect accuracy coefficients were also low for most of the variables. Thus, the pattern of the results based on the overall accuracy correlations was quite similar to the pattern of results based on the direct accuracy coefficients.

Accuracy in Judging the Absent Partner's Feelings

In general our results provide evidence that supports the hypothesis that accuracy in judging the partner's feelings, whether at work or at home, was higher than mere chance (H-1). It also supports in part the prediction of hypothesis H-3, that even when assumed similarity was controlled the resulting direct accuracy coefficients would remain significant. However, our results suggest that hypothesis H-1 and H-3 must be differentiated according to the gender of the spouse, the type of situation, and the particular feelings being judged.

Women were more accurate at judging their partner's feelings at work (s1), reaching impressive overall accuracy coefficients of $r = 0.40$ and higher for most of the items. In the more restricted situation, when women were at home and estimated their husband's feelings at work (s3), accuracy coefficients were lower, but most of them remained significant. When both partners were at home but not together (s4), women were also quite accurate in judging their partner's feelings.

Husbands' judgments of their wife's feelings were also significant when both partners were at home (s4). They were reasonably good at judging their wife's feelings of stress and tension at work (s2) and reached moderate to small levels of accuracy for the other feelings. However, husbands were not very good at judging how their wife felt at home while they themselves were at work (s3), which suggests that husbands were better at empathizing with their wife's feelings at work than with their wife's feelings at home during the day.

The size of many of the coefficients was remarkably high, even after controlling for assumed similarity, bearing in mind that our accuracy coefficients are based on direct situation-specific measures in which variance due to more stable mood trends, response styles and personality traits have been eliminated. Women have a particularly accurate idea about how their partner feels both at work and at home, which is a striking result since their judgments are not based on any perception of their partner's feelings, but on estimates and inferences.

Assumed Similarity in Judging the Absent Partner's Feelings

In line with the literature (Hoch, 1987; Kenny and Acitelli, 2001; Neyer et al., 1999; Thomas et al., 1997) we expected higher coefficients for assumed similarity than for accuracy (H-2). We found some evidence for H-2, but our results suggest that gender, the type of situation and the particular feelings being judged need also to be taken into account.

For the husbands, assumed similarity was generally higher than accuracy, except for judgments about their partner's physical well-being. For the wives this was only true when they were at work (s2) or when both partners were at home (s4). When wives estimated their partner's feelings at work while they were at home, they did not assume similarity at all.

These results indicate that husbands rely extensively on their own feelings when judging their wife's feelings. Situation-specific differences in the amount of assumed similarity were not large for husbands, but they were large for wives. Wives seem to assume similarity when it is appropriate, such as in situation four, where both spouses were at home, and a substantial amount of real similarity in feelings existed. However, the effects of correctly-inferred similarity on accuracy were rather small and quite similar for men and for women (see Table V).

Looking more closely at the spouses' judgments of their partner's feelings at work revealed that they are at least partly based on a process of inference, in which spouses develop estimates of how their partner feels at work from their own situation-specific work-based emotional experience. This result confirms a basic assumption of Nickerson's (1999) "working model of specific others' know-

ledge”, namely that we use our own experience in any specific situation as an anchor for judging the subjective experience of another person in the same situation.

However, Kenny and Acitelli (2001) found that assumed similarity for the spouses’ judgment of their partner’s job satisfaction was quite low and even much lower than accuracy. They argued that because husbands and wives usually have quite different jobs there would be little reason to believe the partner to have the same satisfaction with his or her job and therefore assumed similarity should be low.

The discrepancy between ours and the results of Kenny and Acitelli might be explained by the distinction between job satisfaction and feelings at work. Whereas job satisfaction is a more general and stable construct, the feelings at work are more changeable and therefore, harder to judge. If it is more difficult to infer another person’s disposition or state it is also more probable that people then rely on their own experience.

Differences Regarding the Outcome and Process of Judgments Between Genders

Direct significance test of gender differences in the accuracy coefficients only reached significance when husbands were at work (s1 and s3). However, except for situation two, when women did not know exactly where the husband was, accuracy coefficients for women were generally higher than for men. Taken together, these results accord with hypothesis H-5, which predicts that women would be more accurate than men. Meta-analysis on nonverbal sensitivity research has shown slightly higher accuracy scores for women than for men (Hall, 1978; Brody and Hall, 2000). Ickes et al. (2000) also found that women were more accurate at inferring the thoughts and feelings of their interaction partner. However, this gender difference was only observed in the later studies of Ickes and colleagues, in which participants had to additionally evaluate how accurate their own judgments had been. Ickes et al. (2000) speculated that this self-evaluation of their own empathic accuracy would motivate women, but not men, to be more sensitive to the thoughts and feelings of their interaction partner. Therefore these gender differences reflect the

differential motivation, rather than the differential ability, of men and women. This speculation was in line with conclusions about gender differences in empathy that were reached by Eisenberg and Lennon (Eisenberg and Lennon, 1983; Lennon and Eisenberg, 1987), and was further confirmed in the studies of Klein and Hodges (2001).

Our study cannot contribute to the discussion on whether the observed gender differences are due to motivation or ability. However, the results on assumed similarity suggest that there are some differences in the judgment process between men and women. They indicate that husbands in general tend to use their own current feelings to judge their partner's feelings, whereas wives seem to assume similarity in a more situation-specific way. A closer examination of judgments made when spouses were at work also demonstrates that women tend to abstract more from their current feelings than men do, and rely more on their situation-specific experience to judge how their partner feels at work. Taken together, these results might suggest that, when judging their absent partner's feelings, women are more empathetic and men are more egocentric.

Testing for Moderating Effects of Marital Satisfaction on Accuracy and Assumed Similarity

There was some weak evidence to suggest a relation between accuracy in judging the partner's feelings and marital satisfaction (H-6). However, results were not clear-cut as there were also moderating effects which indicated the opposite relationship.

Significant moderating effects of marital satisfaction on falsely-assumed similarity were all negative, indicating that satisfied spouses assume less similarity. This finding contradicts the recent results of Murray et al. (2002), who found that "egocentrism" (their label for what we called falsely-assumed similarity) in judging their partner's traits, values and day-to-day feelings was associated with a higher satisfaction in the relationship.

We must be cautious when interpreting our results, as they might be affected by the fact that most of our couples were quite satisfied with their relationship. Variance in relationship satisfaction was consequently low, making potential moderating effects hard to detect. In addition, our sample size was probably too small to obtain

reliable moderator effect estimates. Therefore, our results might not be very robust and have to be replicated.

Limitations of Our Study

The data analytic approach we took to answer our questions was based on the aggregation of data from selected situations. The major problem in applying this approach to our data was that many of the spouses made only a small number of observations that matched the selection criteria (at least in three of four situations). It can therefore be argued that the feelings which spouses might in general have at work or in another setting are not very well represented if they are based on a sample of only one, two or three observations. To cope with this problem, we applied a reliability adjustment procedure whereby the less reliable a score was, the closer it was placed to the individual mean. However, this was a rather conservative procedure which tends to underestimate how far situation-specific scores deviate from the individual's average, when only a few observations are available.

One could reduce this problem by extending the observation period, but aside from being costly, this would lead to other problems, such as a decrease in the motivation to participate.

Another problem was the loss of couples when one of the spouses had missing data in the specific situations of interest. This led to different sub-samples for each of the situations and means that situational differences in the results might also be due to selection effects. To control against this possibility, we recomputed all the correlation coefficients on the subset of $N = 36$ couples in which both spouses provided data in each of the four situations. Although the size of some coefficients was larger or smaller than reported in Table III, the medians of the coefficients and the general pattern of result only changed slightly.¹⁷ We therefore think that while single results might be affected, our main conclusions are not.

These two limitations are not specific to our study, but are common problems in intensive time sampling research, and often occur when situation- or setting-specific results are being explored.

Another criticism might be that a specification of the setting which relies only on the local dimension (whether at work or home) is a very crude category. Many other setting-specific factors, such

as the kind of activity in which one is engaged (whether work or leisure) and the social environment (whether alone or with other people), would have an influence on one's feelings (e.g., Larson and Richards, 1994; Wilhelm, 2001). In particular home situations can be very heterogeneous depending on what the participants are doing, which might be housework, child care, relaxation, sleeping, eating, leisure pursuits, or even working. The low accuracy of the husbands' judgments of what their wife feels at home during the day (s3) might partly be because they may not know exactly what their wife is doing. It would therefore be a good idea if future studies did not simply ask where the partner is, but went on to find out what the partner might be doing, and who he or she might be with.

The generalization of our results is limited to well educated middle class couples in which spouses have been living together in a stable, mutually satisfying relationship for a long time (many of them 20 years and more). It is plausible to assume that older spouses benefit from an accumulation of knowledge of their partner's reactions when they try to assess how their partner is feeling. They might therefore be more accurate than couples who are at the beginning of their relationship. However, there are potential negative effects for older couples which might balance these potential advantages. For instance, these spouses might believe that they already know their partner so intimately that they might start to rely on this accumulated knowledge, and be less interested in their partner's actual daily life experiences (Colvin et al., 1997).

While most of the men were in full-time employment, most of the women were employed part time (64%), some were not employed at all (22%) and only a few worked full time (13%). Employment status in the female sample, while typical for Switzerland, is therefore fairly heterogeneous, and the degree to which this fact is associated with the husbands' (in)accuracy in judging their wife's feelings, both at home and at work, must remain an open question.

CONCLUSION

Despite its limitations, our study is an original contribution to accuracy research. We used a computer-assisted diary approach to assess accuracy and assumed similarity under natural conditions in

spouses' daily lives. We found the judgment of the absent partner's feelings to be quite accurate. However, accuracy levels depended on the type of situation and on the particular feelings which were being judged. Projection of the judges' own current feelings was often, but not always, higher than accuracy. Direct gender differences were rare, but the pattern of results suggest that wives are more sensitive to their husband's feelings than vice versa.

ACKNOWLEDGEMENTS

Firstly, we would like to thank the Swiss National Science Foundation for financially supporting our research project on "Individual and social regulation of emotions in families" (Nr. 5004-047773), which formed part of a larger interdisciplinary project on Family Life in Switzerland.

We have to thank many people helping us to improve this article. Joe McGrath and Richard Moreland for their comments on a very early version of this paper, and Ian Law and Anthony Clark for improving the English of the first version we submitted. The anonymous reviewers for their constructive criticism and suggestions that helped us to further refine the paper. Judith Baessler and Tony Arthur for their great help and support in revising the English of our final version. David Kenny and Gisela Michel for their additional comments and corrections on this version. And finally, Dorothee Aebischer for her organizational support.

NOTES

¹ Accuracy labels in brackets are from Kenny and Winkquist (2001, p. 266)

² For these analyses we used the average over six 'feeling-items' which could be seen as indicators of the actual emotional state or mood.

³ Since Cronbach (1955), separating measures into clearly interpretable components of accuracy is a central request of any research on accuracy (Bernieri, 2001; Davis and Kraus, 1997; Kenny, 1994).

We applied two data analytic strategies. The first was similar to the approach we have used in this article: centering data around the individual's mean, aggregating the centered data over the observations when partners were together and when they were apart, adjusting the aggregated scores for reliability and then computing correlation coefficients or regression models. These coefficients tell us

about spouses' accuracy at judging their partner's mean state when the partner was either present or absent.

The second strategy was based on multilevel models to control the multiple dependency in our data, and to calculate coefficients for different components of accuracy and assumed similarity. In these models we separated the spouses' accuracy at judging their partner's current emotional state (at a single observation) from their accuracy at judging their partner's daily emotional state (average over six observations per day) and their accuracy at judging their partner's general emotional state (average over all the 42 observations). The coefficients of these analyses tell us about the spouses' accuracy at judging their partner's current emotional state when the partner was present or absent.

Although accuracy measures were not the same in the two analyses, the main results reported in this paper were quite similar (see Wilhelm, in press).

⁴ Adolescents answered the same questions, except for the judgment of another family member's emotional state.

⁵ In addition, family members completed questionnaires at the end of the self-observation-week and 6 to 12 months later.

⁶ The items 'excited/nervous – quiet'; 'unsafe/uncertain – safe/certain'; 'hungry – full' were also presented to rate one's own state, but not to rate the state of the partner.

⁷ The first participating couple in our study did not get the item 'without energy/ tired – full of energy', 'tense – relaxed' to estimate their partner's feelings, and the first four participating couples did not get the item 'at ease – stressed'. Consequently the sample size for these items is always lower.

⁸ We used additional scales beside the RAS, but they were not relevant to the content of this article.

⁹ For each person j , we computed the mean (\bar{x}_j) across all the observations i given by that person, and subtracted this mean from each single observation (x_{ij}) given by that person: $x_{ij}' = x_{ij} - \bar{x}_j$

¹⁰ For each emotional state variable a four level 'empty model' was computed, in which only the intercept and the variation around the intercept at each level was estimated (Wilhelm, 2001). These levels were variation between couples (level 4), variation between spouses within couples (level 3), variation between days within spouses, within couples (level 2), and variation between single observations within days, within spouses, within couples (level 1). Adding together the level 3 variation and the level 4 variation gives the amount of variation between persons. Adding together the variances at level 1 and level 2 gives the amount of variation within persons.

¹¹ We used the program GPOWER (Faul and Erdfelder, 1992) to calculate the power for our conditions of testing. The power to detect a medium correlation of $r = 0.30$, setting alpha 0.05, two-tailed, was 0.84 when 90 couples were available (s1) and 0.66 when 59 couples were available (s2). The power to detect a mean difference of $d_z = 0.25$ under the same conditions of testing was 0.65 to 0.47 (see also next endnote).

¹² The relationship between the effect size measure for independent *t*-tests d and d_z depends on the correlation between the two samples compared: if the correlation is zero, $d_z = 0.5 d$, if the correlation is $r = 0.50$, $d_z = d$. Thus, a d_z of 0.25 corresponds to a medium effect if there is no correlation between the samples compared, and to a small effect if there is a high correlation between these samples.

¹³ When husbands were at work (s3) wives were about twice as much occupied with work for the job, for the household or for other persons than when husbands were at home (s4) (Medians: 67% vs. 33%; Wilcoxon test: $z = 3.87$; $p < 0.001$)

¹⁴ Husbands' self-judgments in situation one were compared with wives' self-judgments in situation two using paired sample *t*-tests.

¹⁵ Because for this analysis, there were only women available who reported their feelings at work, we computed the assumed similarity correlation coefficients in s1 again for this subsample ($n = 61$). Compared to the coefficients reported in Table III (s1, column 4), that are based on 91 women, the coefficients for the subsample were only slightly different. The largest difference between the corresponding coefficients was -0.14 .

¹⁶ It's important to note that Cronbach's (1955) separation refers to a data set in which a perceiver judges different targets on a number of traits. Both judgment and criterion scores are separated into four different components: constant, trait, target and uniqueness. The components therefore correspond to a two-way analysis of variance with the factors (target*trait) (Kenny and Winkquist, 2001). The link between each of the four corresponding components in the judgment and the criterion is a measure for a specific type of accuracy.

The *elevation accuracy* is the degree of correspondence between the constant of the criterion and the constant of the judgment. 'It reflects the discrepancy between the perceiver's mean judgment across targets and traits and the overall judgment across targets and traits on the criterion' (Kenny and Winkquist, 2001, p. 267).

The *differential elevation accuracy* is the correlation between the target effects of the criterion and the target effects of the judgment. 'This form of accuracy reflects the correspondence between the mean of a perceiver's judgment of each target and the overall mean level for the target on the criterion' (Kenny and Winkquist, 2001, p. 267).

The *stereotype accuracy* is the correlation between the trait effects of the criterion and the trait effects of the judgment. 'This type of accuracy concerns the degree of correspondence between the mean of a perceiver's judgments of each trait across targets and the overall mean level of the trait on the criterion' (Kenny and Winkquist, 2001, p. 267).

The *differential accuracy* is the correlation between the uniqueness component of the criterion and the uniqueness component of the judgment. 'This form of accuracy reflects the degree of correspondence between the perceiver's judgment of each trait for each target and the criterion scores of each trait for each target' (Kenny and Winkquist, 2001, p. 267).

Kenny and Winqvist (2001) apply Cronbach's separation approach to other data sets. The design in which each perceiver judges a target on a set of measures, and in which each perceiver is also judged by that target (roles of perceiver and target are reciprocal) is similar to the design of our study. However, we have an additional component that is due to the repeated observations. We did not take each judge's and each target's mean across items to get the differential elevation components, as in the analysis of Kenny and Winqvist (2001). Instead, we took each judge's mean across repeated observations to construct a differential elevation component for every item. Therefore, our differential elevation component is not exactly the same as the one reported by Kenny and Winqvist.

¹⁷ The maximal difference between the coefficients in Table III and the corresponding recomputed coefficients was 0.32, and the maximal difference between the medians of the coefficients reported in Table III and the corresponding medians of the recomputed coefficients was 0.14).

REFERENCES

- Ambady, N., F. J. Bernieri and J. A. Richeson: 2000, 'Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream', in M. P. Zanna (ed.), *Advances in Experimental Social Psychology*, Vol. 32 (Academic Press, San Diego, CA), pp. 201–271.
- Bernieri, F. J.: 2001, 'Toward a taxonomy of interpersonal sensitivity', in J. A. Hall and F. J. Bernieri (eds.), *Interpersonal Sensitivity. Theory and Measurement* (Lawrence Erlbaum, Mahwah, NJ/London), pp. 3–20.
- Bissonnette, V. L., C. E. Rusbult and S. D. Kilpatrick: 1997, 'Empathic accuracy and marital conflict resolution', in W. Ickes (ed.), *Empathic Accuracy* (Guilford Press, New York), pp. 251–281.
- Bolger, N., A. Davis and E. Rafaeli: 2003, 'Diary methods: Capturing life as it is lived', *Annual Review of Psychology* 54, pp. 579–616.
- Brody, L. R. and J. A. Hall: 2000, 'Gender, emotion, and expression', in M. Lewis and J. M. Haviland-Jones (eds.), *Handbook of Emotions*, 2nd edn. (Guilford Press, New York), pp. 338–349.
- Cohen, J.: 1988, *Statistical Power Analysis for the Behavioral Sciences*, 2nd. rev. edn. (Lawrence Erlbaum Associates, Inc., Hillsdale, NJ).
- Colvin, C. R., D. Vogt and W. Ickes: 1997, 'Why do friends understand each other better than strangers do?', in W. Ickes (ed.), *Empathic Accuracy* (Guilford Press, New York), pp. 169–193.
- Cook, W. L. and E. M. Douglas: 1998, 'The looking-glass self in family context: A social relations analysis', *Journal of Family Psychology* 12, pp. 299–309.
- Cronbach, L.: 1955, 'Processes affecting scores on "understanding of others" and "assumed similarity"', *Psychological Bulletin* 52, pp. 177–193.
- Davis, M. H. and L. A. Kraus: 1997, 'Personality and accuracy', in W. Ickes (ed.), *Empathic Accuracy* (Guilford Press, New York), pp. 144–168.

- Eisenberg, N. and R. Lennon: 1983, 'Sex differences in empathy and related capacities', *Psychological Bulletin* 94, pp. 100–131.
- Fahrenberg, J.: 1996, 'Ambulatory assessment: Issues and perspectives', in J. Fahrenberg and M. Myrtek (eds.), *Ambulatory Assessment: Computer-Assisted Psychological and Psychophysiological Methods in Monitoring and Field Studies* (Hogrefe and Huber Publishers, Kirkland, WA), pp. 3–20.
- Fahrenberg, J., P. Hüttner and R. Leonhart: 2001, 'MONITOR: Acquisition of psychological data by a hand-held PC', in J. Fahrenberg and M. Myrtek (eds.), *Progress in Ambulatory Assessment* (Hogrefe and Huber Publishers, Kirkland, WA), pp. 93–112.
- Fahrenberg, J., R. Leonhart and F. Foerster: 2002, *Alltagsnahe Psychologie. Datenerhebung im Feld mit Hand-held PC und physiologischem Mess System* (Huber, Bern, CH).
- Faul, F. and E. Erdfelder: 1992, *GPOWER: A priori, Post-hoc, and Compromise Power Analyses for MS-DOS (Computer-Programm)* (Psychologisches Institut der Universität, Bonn).
- Funder, D. C.: 1999, *Personality Judgment. A Realistic Approach to Person Perception* (Academic Press, San Diego).
- Gage, N. L. and L. Cronbach: 1955, 'Conceptual and methodological problems in interpersonal perception', *Psychological Review* 62, pp. 411–422.
- Gottman, J. M. and A. L. Porterfield: 1981, 'Communicative competence in the nonverbal behavior of married couples', *Journal of Marriage and the Family* 43, pp. 817–824.
- Gottman, J. and N. Silver: 1999, *The Seven Principles for Making Marriage Work* (Orion, London).
- Graham, T. and W. Ickes: 1997, 'When women's intuition isn't greater than men's', in W. Ickes (ed.), *Empathic Accuracy* (Guilford Press, New York), pp. 117–143.
- Hall, J. A.: 1978, 'Gender effects in decoding nonverbal cues', *Psychological Bulletin* 85, pp. 845–857.
- Hank, P., P. Schwenkmezger and J. Schumann: 2001, 'Daily mood reports in hindsight: Results of a computer-assisted time sampling study', in J. Fahrenberg and M. Myrtek (eds.), *Process in Ambulatory Assessment* (Hogrefe and Huber Publishers, Kirkland, WA), pp. 143–156.
- Hendrick, S. S.: 1988, 'A generic measure of relationship satisfaction', *Journal of Marriage and the Family* 50, pp. 93–98.
- Hoch, S. J.: 1987, 'Perceived consensus and the predictive accuracy: The pros and cons of projection', *Journal of Personality and Social Psychology* 53, pp. 221–234.
- Ickes, W.: 1993, 'Empathic accuracy', *Journal of Personality* 61, pp. 587–610.
- Ickes, W.: 1997, 'Introduction', in W. Ickes (ed.), *Empathic Accuracy* (Guilford Press, New York), pp. 1–16.
- Ickes, W.: 2001, 'Measuring empathic accuracy' in J. A. Hall and F. J. Bernieri (eds.), *Interpersonal Sensitivity. Theory and Measurement* (Lawrence Erlbaum, Mahwah, NJ/London), pp. 219–241.

- Ickes, W., P. R. Gesn and T. Graham: 2000, 'Gender differences in empathic accuracy: Differential ability or differential motivation?', *Personal Relationships* 7, pp. 95–109.
- Ickes, W. and J. A. Simpson: 1997, 'Managing empathic accuracy in close relationships', in W. Ickes (ed.), *Empathic Accuracy* (Guilford Press, New York), pp. 218–250.
- Ickes, W., L. Stinson, V. Bissonette and S. Garcia: 1990, 'Naturalistic social cognition: Empathic accuracy in mixed-sex dyads', *Journal of Personality and Social Psychology* 59, pp. 730–742.
- Johnstone, T. and K. R. Scherer: 2000, 'Vocal communication of emotion', in M. Lewis and J. M. Haviland-Jones (eds.), *Handbook of Emotions*, 2nd edn. (Guilford Press, New York/London), pp. 220–235.
- Käppler, C. and St. Rieder: 2001, 'Does the retrospection effect hold as a stable phenomenon? – First results from a trans-cultural self-monitoring study of mood and cognitive states in Brazil and Germany', in J. Fahrenberg and M. Myrtek (eds.), *Progress in Ambulatory Assessment* (Hogrefe and Huber Publishers, Kirkland, WA), pp. 113–122.
- Keltner, D. and P. Ekman: 2000, 'Facial expression of emotion', in M. Lewis and J. M. Haviland-Jones (eds.), *Handbook of Emotions*, 2nd edn. (Guilford Press, New York/London), pp. 236–249.
- Kenny, D. A.: 1994, *Interpersonal Perception: A Social Relations Analysis* (Guilford Press, New York).
- Kenny, D. A. and L. K. Acitelli: 2001, 'Accuracy and bias in the perception of the partner in a close relationship', *Journal of Personality and Social Psychology* 80, pp. 439–448.
- Kenny, D. A. and W. Cook: 1999, 'Partner effects in relationship research: Conceptual issues, analytical difficulties and illustrations', *Personal Relationships* 6, pp. 433–448.
- Kenny, D. A., C. D. Mohr and M. J. Levesque: 2001, 'A social relations variance partitioning of dyadic behavior', *Psychological Bulletin* 127, pp. 128–141.
- Kenny, D. A. and L. Winkler: 2001, 'The measurement of interpersonal sensitivity: Consideration of design, components, and unit of analysis', in J. A. Hall and F. J. Bernieri (eds.), *Interpersonal Sensitivity. Theory and Measurement* (Lawrence Erlbaum, Mahwah, NJ/London), pp. 265–302.
- Klein, K. J. and D. S. Hodges: 2001, 'Gender differences, motivation and empathic accuracy: When it pays to understand', *Personality and Social Psychology Bulletin* 27, pp. 720–730.
- Larson, R. W. and D. M. Almeida: 1999, 'Emotional transmission in the daily lives of families: A new paradigm for studying family process', *Journal of Marriage and the Family* 61, pp. 5–20.
- Larson, R. W. and M. Richards: 1994, *Divergent Realities. The Emotional Lives of Mothers, Fathers, and Adolescents* (Basic Books, New York).
- Lennon, R. and N. Eisenberg: 1987, 'Gender and age differences in empathy and sympathy', in N. Eisenberg and J. Strayer (eds.), *Empathy and Its Development* (Cambridge University Press, Cambridge), pp. 195–217.

- Levenson, R. W. and A. M. Ruef: 1992, 'Empathy: A psychophysiological substrate', *Journal of Personality and Social Psychology* 63, pp. 234–246.
- Marangoni, C., S. Garcia, W. Ickes and G. Teng: 1995, 'Empathic accuracy in a clinically relevant setting', *Journal of Personality and Social Psychology* 68, pp. 854–869.
- Markus, H., J. Smith and R. L. Moreland: 1985, 'Role of the self-concept in the perception of others', *Journal of Personality and Social Psychology* 49, pp. 1494–1512.
- Michel, G.: in preparation, 'Daily rhythms of symptom reporting'.
- Murray, S. L., G. Bellavia, J. G. Holmes, D. W. Griffin and D. Dolderman: 2002, 'Kindred spirits? The benefits of egocentrism in close relationships', *Journal of Personality and Social Psychology* 82, pp. 563–581.
- Myrtek, M., A. Fichtler, M. Strittmatter, G. Brügger: 1999, 'Stress and strain of blue and white collar workers during work and leisure time: Results of psychophysiological and behavioral monitoring', *Applied Ergonomics* 30, pp. 341–351.
- Myrtek, M. and F. Foerster: 2001 'On-line measurement of additional heart rate. Methodology and applications', in J. Fahrenberg and M. Myrtek (eds.), *Progress in Ambulatory Assessment* (Hogrefe and Huber Publishers, Kirkland, WA), pp. 399–414.
- Myrtek, M., D. Zanda and E. Aschenbrenner: 2001 'Interactive psychophysiological monitoring of emotions in student's everyday life. A replication study', in J. Fahrenberg and M. Myrtek (eds.), *Progress in Ambulatory Assessment* (Hogrefe and Huber Publishers, Kirkland, WA), pp. 415–434.
- Neyer, F. J., R. Banse and J. B. Asendorpf: 1999, 'The role of projection and empathic accuracy in dyadic perception between older twins', *Journal of Social and Personal Relationships* 16, pp. 419–422.
- Nickerson, R. S.: 1999, 'How we know – and sometimes misjudge – what others know: Imputing one's own knowledge to others', *Psychological Bulletin* 125, pp. 737–759.
- Noller, P.: 1984, *Nonverbal Communication and Marital Interaction* (Pergamon Press, Oxford).
- Noller, P.: 2001, 'Using standard content methodology to assess nonverbal sensitivity in dyads', in J. A. Hall and F. J. Bernieri (eds.), *Interpersonal Sensitivity. Theory and Measurement* (Lawrence Erlbaum, Mahwah, NJ/London), pp. 243–264.
- Noller, P. and J. A. Feeney: 1998, 'Communication in early marriage: Responses to conflict, nonverbal accuracy, and conversational patterns', in T. N. Bradbury (ed.), *The Developmental Course of Marital Dysfunction* (Cambridge University Press, New York), pp. 11–43.
- Noller, P. and M. Ruzzene: 1991, 'Communication in marriage; The influence of affect and cognition', in G. J. O. Fletcher and F. D. Fincham (eds.), *Cognition in Close Relationships* (Lawrence Erlbaum Associates, Hillsdale, NJ), pp. 203–233.

- Perrez, M., R. Berger and P. Wilhelm: 1998, 'Die Erfassung von Belastungserleben und Belastungsverarbeitung in der Familie: Self-monitoring als neuer Ansatz', *Psychologie in Erziehung und Unterricht* 45, pp. 19–35.
- Perrez, M., D. Schoebi and P. Wilhelm: 2000, 'How to assess social regulation of stress and emotions in daily family life? A computer-assisted family self-monitoring system (FASEM-C)', *Clinical Psychology and Psychotherapy* 7, pp. 326–339.
- Perrez, M., P. Wilhelm, D. Schoebi and M. Horner: 2001, 'Simultaneous computer assisted assessment of causal attribution and social coping in families', in J. Fahrenberg and M. Myrtek (eds.), *Progress in Ambulatory Assessment* (Hogrefe and Huber Publishers, Kirkland, WA), pp. 25–43.
- Prentice, D. A. and D. T. Miller: 2002, 'The emergence of homegrown stereotypes', *American Psychologist* 57, pp. 352–359.
- Sander, J. and S. Böcker: 1993, 'Die Deutsche Form der Relationship Assessment Scale (RAS): Eine kurze Skala zur Messung der Partnerschaftszufriedenheit', *Diagnostica* 39, pp. 55–62.
- Schoebi, D., P. Wilhelm and M. Perrez: in preparation, 'Coping with conflicts in families: Do perspectives matter?'
- Sillars, A. L.: 1985, 'Interpersonal perception in relationships', in W. Ickes (ed.), *Compatible and Incompatible Relationships* (Springer, New York), pp. 277–305.
- Simpson, J. A., W. Ickes and T. Blackstone: 1995, 'When the head protects the heart: Empathic accuracy in dating relationships', *Journal of Personality and Social Psychology* 69, pp. 629–641.
- Snijders, T. A. B. and R. J. Bosker: 1999, *Multilevel Analysis. An Introduction to Basic and Advanced Multilevel Modelling* (Sage, London).
- Stinson, L. and W. Ickes: 1992, 'Empathic accuracy in the interactions of male friends versus male strangers', *Journal of Personality and Social Psychology* 62, pp. 787–797.
- Thomas, G., G. J. O. Fletcher and C. Lange: 1997, 'On-line empathic accuracy in marital interaction', *Journal of Personality and Social Psychology* 72, pp. 839–850.
- Wilhelm, P.: 2001, 'A multilevel approach to analyze ambulatory assessment data: An examination of family members' emotional states in daily life', in J. Fahrenberg and M. Myrtek (eds.), *Progress in Ambulatory Assessment* (Hogrefe and Huber Publishers, Kirkland, WA), pp. 173–189.
- Wilhelm, P.: in press, *Empathie im Alltag von Paaren* (Huber, Bern).
- Wilhelm, P., M. Horner and M. Perrez: 2000, 'Humeur et gestion du stress dans les interactions familiales: résultats à partir de la méthode FASEM', in J.-P. Pourtois and H. Desmet (eds.), *Relation familiale et résilience* (L'Harmattan, Paris), pp. 255–274.
- Wilhelm, P. and M. Perrez: 2001, 'Felddiagnostik', in R.-D. Stieglitz, U. Baumann and H. J. Freyberger (eds.), *Psychodiagnostik in der Klinischen Psychologie, Psychotherapie und Psychiatrie* (Thieme, Stuttgart), pp. 169–182.

Zbinden, M.: 2003, Angstbewältigung und Selbstbeobachtung. Der Einfluss von Bewältigungsdispositionen auf Stimmung, Belastungserleben und dessen Bewältigung unter natürlichen Bedingungen (Peter Lang, Bern).

Department of Psychology

University Fribourg

Rue de Foucigny 2

CH-1200, Fribourg

Switzerland

E-mails: Peter.Wilhelm@unifr.ch and Meinrad.Perrez@unifr.ch