

Empathic Accuracy and Relationship Satisfaction: A Meta-Analytic Review

Haran Sened¹

Michal Lavidor¹

Gal Lazarus¹

Eran Bar-Kalifa¹

Eshkol Rafaeli^{1,2,3}

William Ickes⁴

Authors Notes

¹ Department of Psychology, Bar-Ilan University.

² Gonda Multidisciplinary Brain Research Center, Bar-Ilan University;

³ Department of Psychology, Barnard College, Columbia University

⁴ Department of Psychology, University of Texas at Arlington

Accepted to the Journal of Family Psychology. This article may not exactly replicate the final version published in the APA journal. It is not the copy of record.

Abstract

Empathic accuracy (EA; Ickes & Hodges, 2013) is the extent to which people accurately perceive their peers' thoughts, feelings and other inner mental states. EA has particularly interested researchers in the context of romantic couples. Reviews of the literature suggest a possible link between romantic partners' EA and their relationship satisfaction (Sillars & Scott, 1983; Ickes & Simpson, 2001). To assess the magnitude of this association and examine possible moderators, we performed a meta-analytic review of 21 studies (total $N = 2739$ participants) which examined the association between EA and satisfaction. We limited our review to studies measuring EA using the dyadic interaction paradigm (Ickes, Stinson, Bissonette, & Garcia, 1990). We found a small, but significant association between the two ($r = .134$; $p < .05$). Subsequent moderation analyses demonstrated that EA for negative emotions (one's accuracy when assessing their partner's negative emotions) was more closely related to satisfaction ($r = .171$; $p < .05$) than EA for positive ones ($r = .068$; $p > .1$). The association was also stronger in relationships of moderate length, suggesting that EA may be more meaningful when relationships are consolidating but before they become stable. Gender and procedural variations on the Dyadic Interaction Paradigm did not moderate the association, and there was no difference depending on whether the association was between EA and perceivers' or targets' satisfaction (i.e., actor or partner effects). We discuss the implications of these findings and offer recommendations for future EA studies.

Keywords: Empathic Accuracy, Relationship Satisfaction, Marital Satisfaction, Person Perception, Empathy

It would be difficult to imagine a functioning society, not to mention close interpersonal relationships, without humans' ability to accurately perceive and assess the mental states of their peers. The term *empathic accuracy* (EA; Ickes & Hodges, 2013) refers to the extent to which human beings use this ability to arrive at accurate inferences.

People attempt to achieve high EA in various contexts—from negotiators trying to assess the other side's intentions, and therapists trying to hone in on their clients' difficulties, to romantic partners trying to understand each other as a way to promote mutual support and relationship satisfaction. The association between empathic accuracy and satisfaction has in fact been studied extensively in this last context, that of romantic relationships. In general, reviews of this literature find accuracy and satisfaction to be positively associated (Sillars & Scott, 1983; Ickes & Simpson, 2001), but also point to some caveats.

One caveat, proposed by Sillars and Scott, is that high accuracy is not required in mundane, everyday interactions. They also refer to research (Sillars, Pike, Jones, & Redmon, 1983) indicating that some couples do not particularly value openness or intimacy in their relationships; couples such as these experience greater satisfaction when they *avoid* discussion and conflict. An additional caveat is the phenomenon termed *motivated inaccuracy* (Ickes & Simpson, 1997, 2001). As Ickes and Simpson (1997, 2001) argue, access to more information regarding one's partner may serve as a threat, rather than a benefit, in certain relationships and situations. They suggest that in these cases, partners might (consciously or not) attempt to be less accurate, thereby protecting themselves and their relationships from this potentially threatening information. For example, Simpson et al. (1995) manipulated relationship threat by asking the members of dating couples to rate aloud the attractiveness and sexual appeal of potentially available partners while their current dating partners were sitting next to them in the same room. They found that higher empathic accuracy regarding the other partner's thoughts and feelings in this relationship-threatening

situation correlated with negative relationship outcomes. In a later study, Simpson, Oriña, and Ickes (2003) demonstrated that empathic accuracy for relationship-threatening thoughts and feelings during a marital conflict discussion was associated with reduced closeness.

To the best of our knowledge, no work to date has attempted to quantitatively consolidate these lines of research by estimating the general association between EA and satisfaction across multiple studies. Such work could verify the conclusion of the reviews by Sillars and Scott (1983) and Ickes and Simpson (2001), who argued that the association between EA and satisfaction should, in general, be positive, because the aforementioned caveats apply to relatively rare exceptional cases. Testing the proposed positive relationship is therefore the aim of the present meta-analysis.

An earlier meta-analysis by Fletcher and Kerr (2010) did focus on synthesizing a set of studies examining broadly-defined accuracy within romantic relationships; as a secondary analysis, they reported a null correlation between studies of “tracking accuracy”—i.e., the correlation between self- and partner reports on various constructs—and relationship evaluations or satisfaction. Notably, their analyses relied primarily (85.7%) on studies assessing accuracy in the perception of various stable traits (e.g., trait aggression) and predictions about the future or assessment of partner memories; only a minority of the studies (14.3%) examined empathic accuracy (i.e., assessed accuracy of inferences regarding transient mental states).

Fletcher and Kerr (2010) also found a correlation between positive bias—the extent to which partner’s reports are more positive than self-reports—and higher satisfaction. This finding might be interpreted as a negative association between accuracy and satisfaction because high bias can mean lower accuracy. However, because this finding was based even more heavily on non-EA studies (with only 6.2% of the included studies examining EA), the findings from this study are of only limited relevance to the question at hand.

One obstacle that needs to be surmounted when synthesizing EA studies is that various methods have been used for its assessment. Many studies have relied on questionnaires, particularly the Interpersonal Responsivity Index (IRI; Davis, 1983), which asks participants to report how empathic they think they are. Although self-estimates of empathy obtained from such measures are often correlated with relationship satisfaction (e.g. Long & Andrews, 1990), they suffer from the inherent potential for bias present in any self-report measure (Harvey, Hendrick & Tucker, 1988). Moreover, it is unclear if they tap into perceptions of one's empathic *accuracy* or of one's general *capacity* for empathy (Davis & Kraus, 1997; Gleason, Jensen-Campbell, & Ickes (2009). For example, Davis and Kraus (1997) and Ickes (1993, 2003) noted that self-report measures rarely correlate with performance measures of accuracy, suggesting that people possess low meta-knowledge regarding their own empathic abilities.

A more recent method for studying accuracy, one that circumvents the use of meta-knowledge, involves the use of diary or experience sampling data collected from dyads (e.g., Wilhelm & Perrez, 2004; Howland & Rafaeli, 2010). In this method, both partners report on their own moods or other transitory constructs and infer their partner's levels on these same constructs every day (and in some cases, multiple times per day). Accuracy can then be assessed in regard to both the daily (or within-day) level as well as the aggregate (person or couple) level. Some studies using this method (e.g., Rafaeli et al., 2016) have found associations between accuracy and satisfaction; however, because this method is relatively new, few studies to date have used it, and fewer still have examined the association between accuracy and satisfaction. Therefore, diary studies will not be included in the current review.

The method most closely tied to the surge in theoretical and empirical research regarding EA in the last few decades is the one pioneered by Ickes and his colleagues (1990). In this method, labeled the *dyadic interaction paradigm*, dyads are videotaped having a

conversation, after which each partner is asked to independently review the video twice: first to record their own thoughts and feelings and then to infer their partner's thoughts and feelings during the conversation. In the classic version of the dyadic interaction paradigm (e.g., Ickes et al., 1990), the similarity between the targets' actual thoughts and feelings and the perceivers' inferences of these thoughts and feelings was coded by objective raters who read the corresponding actual and inferred thoughts and feelings that the partners had provided during the video review. Some recent studies have used a variation of the task, asking participants to rate their own and their partners' feelings on some numerical scale, and comparing results mathematically (e.g. Study 1 in Overall, Fletcher, Simpson & Fillo, 2015).

Because of the greater ubiquity of studies that have not only used the dyadic interaction paradigm but have also examined the association of empathic accuracy with relationship satisfaction, these studies will be the focus of the present meta-analysis. The dyadic interaction paradigm is unique among the presented methods in that it permits the assessment of moment-to-moment accuracy with high precision but also provides an aggregated summary measure of empathic accuracy across the entire interaction period.

To our knowledge, no previous meta-analysis has examined the associations between EA and relationship satisfaction as examined in these studies. Fletcher and Kerr (2010) did include five of them in their meta-analysis, but as mentioned earlier, they were not analyzed separately with regard to relationship satisfaction but instead were included with over 90 other studies assessing various forms of interpersonal accuracy through diverse methods.

Moderators

Besides a general association between EA and satisfaction, we examined some possible moderators of the association.

Target and perceiver effects. The reviews mentioned earlier (Sillars & Scott, 1983; Ickes & Simpson, 2001) tend to examine the concepts of accuracy and satisfaction at the

couple level. However, data permitting, the association between these concepts can also be examined separately for the *perceiver* (i.e., if a person is accurate, is he or she likely to be more satisfied with the relationship) and for the *target* (i.e., if a person is accurate, is his or her partner likely to be more satisfied?)

Compelling mechanisms could exist for both effects. Concerning target effects, a person whose partner is empathically accurate might feel more validated (Swann, Chang-Schneider, & Angulo, 2007) and that their partner is more aware of, and more responsive to, their wishes and needs (Reis & Clark, 2013). With regard to perceiver effects, greater accuracy might allow the perceiver to manage risks in the relationship better (Murray, Holmes, J. G. & Collins, 2006). Even motivated inaccuracy (Simpson et al., 1995; Simpson, Ickes, & Oriña, 2001) requires partners to be accurate in order to anticipate situations in which relationship-threatening information might be revealed, and pre-emptively lower their accuracy accordingly, because one has to first see a baseball coming towards their face to know when to close one's eyes. Of course, it is likely that target and perceiver effects cross over: heightened satisfaction for one can elevate satisfaction for the other in myriad ways.

Single studies tend to yield mixed results when attempting to differentiate target and perceiver effects. For example, Cohen, Schulz, Weiss and Waldinger (2012) found that women's accuracy for both negative and positive moods (measured using an adaptation of the DIP) was associated with their own and their partners' relationship satisfaction. However, only men's accuracy for positive emotions was associated with their own satisfaction, whereas only their accuracy for negative emotions was associated with their female partners' satisfaction. Kilpatrick et al. (2002) found an association between accuracy and satisfaction for both targets and perceivers (of both genders) when the study began, did not find it a year later, and found a significant association only between men's accuracy and their own satisfaction two years later.

Emotion type. Another possible moderator could be the specific mental state (e.g., emotion or mood) one is trying to be empathically accurate about. Each EA study measures the accuracy of partner reports for different types of mental states (for example, Haugen, Welsh and McNulty, 2008, examined accuracy regarding several “feeling” states such as connection vs. discomfort). Though the specific mental states vary across studies, they can generally be grouped into positive versus negative states. In their seminal review, Baumeister, Bratslavsky, Finkenauer, and Vohs (2008) have demonstrated that, across numerous domains, the effect of negative stimuli is stronger than that of positive stimuli. This general phenomenon was also documented regarding the effects of accuracy: accuracy regarding negative moods has been found to be stronger than accuracy regarding positive moods (e.g., Howland & Rafaeli, 2010) and the effects of accuracy regarding negative moods have also been shown to be stronger than the effects of accuracy regarding positive moods (e.g. Haugen et al. 2008; Cohen et al. 2012; Rafaeli et al. 2016).

Relationship length. Kilpatrick, Bissonnette, and Rusbult (2002), who examined empathic accuracy in newlyweds using the DIP over three years, found the association between accuracy and satisfaction to diminish with time. They suggest that at the beginning of a relationship, partners have much to learn about each other, and there is potential for EA to play a larger role than later, when there are few surprises and partner's behavior is guided much more by habit. Ickes and Simpson (2001) suggested that this reduction might explain why another study (Thomas, Fletcher & Lange, 1997), examining empathic accuracy among couples married for 15 years (on average), did not find an association with satisfaction.

Gender. Several authors have considered gender as a possible moderator of EA itself or of its association with satisfaction. A meta-analysis by Ickes, Gesn and Graham (2000) examined gender differences in EA itself as measured by the DIP or by the Standard Stimulus Paradigm (SSP), a paradigm which measures accuracy when viewing standard recorded

discussions. Gender differences in EA were found to occur only in studies in which participants were asked about the accuracy of each of their empathic inferences (e.g., "How accurate were you?"). Ickes and his colleagues argued that such questions activate a motivation for women (but not for men) to be more accurate in concordance with the prescriptive stereotype of "women's intuition," which states that women are supposed to be more empathic than men. This increased motivation caused the women to actually be more accurate in the subsequent inferences they made. Thus, gender differences in EA, when they occur, are differences in motivation and not in ability. With regard to gender as a moderator of the association between accuracy and satisfaction, no conclusive results emerge from the extant literature (e.g., in Kilpatrick et al. 2002, Cohen et al. 2012, as discussed above).

Employed paradigm. Finally, we examined the role of the paradigm/procedure used to obtain accuracy scores as a possible moderator of the studies' results. As mentioned earlier, the classic version of the DIP paradigm developed by Ickes et al. (1990) uses observers' subjective ratings to quantify EA. Specifically, trained coders rate the similarity between pairs of actual and inferred thoughts and feelings on a 3-point scale: 0 (essentially different content), 1 (similar, but not the same, content) or 2 (essentially the same content; Ickes, 1993). The coders' ratings are then averaged for each participant and divided by the total possible score to get a "percent correct" accuracy rating.

Some studies have used variations on the classic version of the DIP. One variation involves adjusting the accuracy score by subtracting a baseline accuracy level obtained by applying the same coding scheme to randomly selected pairs of statements (i.e., partner A's thought or feeling at moment X matched with the inference made by partner B regarding A's thought or feeling at moment Y; Simpson, Ickes & Blackstone, 1995). A second variation on the DIP involves using researcher-determined (rather than participant-determined) time points for stopping the recording and eliciting responses.

A markedly different approach, which can be thought of as a separate paradigm, eschews the use of verbal reports (and their subsequent coding) and instead elicits numerical ratings (and inferences) of mental states from the interaction partners. EA is calculated by computing some index of congruence between the ratings and the inferences (e.g., absolute difference [e.g., Blackmon, 1999]; correlation coefficient [e.g., Cohen et al. 2012]; or mixed model indices of both average difference and association [e.g. Overall et al. 2015]).

The difference between the content coding and the numeric calculation of EA can be significant in at least two ways. Theoretically, content coding examines participants' ability to understand their partners' broad mental state (all thoughts and feelings) whereas numeric methods examine understanding of specific (and pre-specified) emotions. Methodologically, content coding studies are more taxing, both during the lab sessions (as participants need to answer open-ended questions) and afterwards (as coders must first be carefully trained and then review inputs from all participants). Any difference in the predictive validity of the accuracy scores generated by these methods could have theoretical implications for the significance of different types of accuracy, and could have practical implications for the prioritization of research resources. To our knowledge, only one study so far has used both methods on the same sample (Drwal, 2003). Although it found associations with satisfaction only for content-coded accuracy, it did not attempt to directly compare the coding types.

Predictions

Following our literature review, we made the following predictions:

Main Effect. We expected EA to be positively linked to relationship satisfaction, based on earlier (narrative) reviews (Sillars & Scott, 1983; Ickes & Simpson, 2001).

Target and Perceiver Effects. Because plausible mechanisms for both target and perceiver effects exist, and the results of single studies are often inconsistent with each other, we decided to examine the differences between these effects with no specific hypothesis.

Emotion Type. As discussed above, various lines of research have found negative emotions and moods to have more significant effects than positive ones. Thus, we predicted that the meta-analytic results would follow a similar pattern— that EA regarding negative moods would have a stronger association with satisfaction than EA regarding positive moods.

Relationship Length. Following the review by Ickes & Simpson (2001), and findings by Kilpatrick et al. (2002) and Thomas et al. (1997), we expected the association between accuracy and satisfaction to weaken as relationship length increases.

Gender. As previous results concerning gender effects on EA are conflicted, we examined gender as a possible moderator without a specific directional hypothesis.

Employed Paradigm. To our knowledge, no previous studies have directly compared scoring paradigms within the DIP. Thus, we examined the possible moderating role played by the measurement paradigm/procedure that was employed in an exploratory manner.

Method

Sample of Studies

Our meta-analysis followed the PRISMA (Moher, Liberati, Tetzlaff, Altman, and the PRISMA Group, 2009) guidelines for conducting and reporting meta-analyses. We searched PsycINFO, Medline and Google Scholar with the keywords “empathic accuracy” and “relationship satisfaction” or “marital satisfaction.” The search yielded 853 results. We also sent a call for published and unpublished papers on the subject to the International Association of Relationship Research (IARR) newsletter, and examined article abstracts to make sure that studies met the following criteria:

- a) Measured empathic accuracy among couples with the DIP method: Asked couples to have an interaction in the lab, showed them their videotaped interactions, asked them to provide reports/inferences of their own and their partner’s mental states, and compared these in some way.

- b) Measured relationship satisfaction using a self-report questionnaire.

This led us to download or obtain library access to 37 articles and two unpublished datasets which we thoroughly examined. Some of the articles met the criteria but did not report the association between relationship satisfaction and empathic accuracy. We attempted to contact the authors of such papers for the required association data. We also asked those authors to suggest further papers. Study collection ended on May 1st, 2016, except for ongoing correspondence with authors who provided association data which were not reported in the published versions of their studies. There was no stated start date for study collection; the DIP was developed around 1990 (for the first DIP study, see Stinson & Ickes, 1992), and no earlier studies were found. We were left with 21 articles reporting on 20 samples, for a total of 2739 individuals. The list of qualified studies is presented in Table 1. As we analyzed existing studies with no new participants, IRB approval was not required.

Study Coding

Two coders (the first and third authors) coded a range of variables from each study separately and then discussed any differences until reaching consensus. Age, gender, relationship type (married, cohabiting or dating), sample size, average empathic accuracy score, and relationship length were coded. Emotion type was coded using the specific mental states that served as the accuracy targets (e.g. positive regard, Overall, Fletcher & Kenny, 2012; discomfort, Haugen, Welsh & McNulty, 2008). Emotions were coded either “positive” or “negative,” whereas when all thoughts and feelings were measured, the assigned code was “overall”. Data relating to perceiver effects—the association between the perceiver’s accuracy and their own satisfaction—were coded as “self,” whereas data relating to target effects—the association between their accuracy and their partner’s satisfaction—were coded as “other.” Accuracy coding type was coded as “content coding” when the coding system developed by Ickes et al. (1990) was used, and as “numeric” when the similarity between

self- and partner-reported moods was quantified, by using correlations (e.g. Cohen et al. 2012), regressions (e.g. Overall et al. 2015), or difference scores (e.g. Bentley, 2010).

Analyses

All statistics were converted to Pearson correlation coefficients (r). When they were not reported, we appropriately converted existing statistics (e.g., F values). To allow for additive calculations, because correlation coefficients are non-additive, r coefficients were transformed to z' units using fisher's z' transformation (Lipsey & Wilson, 2001).

In each analysis, when multiple effect sizes originated from the same participants (e.g. accuracy for both positive and negative emotions) they were averaged, and when they originated from different participants (e.g., men and women) they were treated as separate samples. The only exception was the study by Kilpatrick et al. (2002) which examined the same participants three times, with one year passing between tests. We treated these as separate samples. In the moderation analyses, when different effect sizes were reported for different values/conditions of the moderator variable, we considered each effect size separately. Final results were converted to r values for reading simplicity.

Homogeneity tests were performed by calculating the Q statistic, which tests for effect size equality (Hedges & Olkin, 1985). The Q statistic's distribution is a chi-square distribution with $N-1$ degrees of freedom, with N being the number of effect sizes examined. Additionally, we used the I^2 statistic (Higgins, Thompson, Deeks & Altman, 2003) to indicate the magnitude of heterogeneity. The I^2 statistic can also be compared between meta-analyses because it is standardized by analysis size. (Lipsey & Wilson, 2001).

Moderation analyses were performed in a manner outlined in Lipsey and Wilson (2001), using SPSS macros provided by Wilson (2011). Categorical moderators were evaluated by performing an ANOVA analog using Q statistics to estimate variance between and within groups, and continuous moderators were evaluated using multiple regression.

When groups of effect sizes were found to be homogenous (Q not significant), we used a fixed-effects model for aggregating effect sizes. Otherwise, we used a random-effects model.

Results

Description of Studies

Our review included 21 studies, encompassing 2739 participants (1370 couples; For one couple the data for one of the partners was missing). Of these studies, 13 are peer-reviewed, published studies, of which 6 cite the association between EA and satisfaction in the published manuscript, and for 7 we retrieved the specific statistical data via contact with the authors. 6 additional studies are unpublished dissertations (parts of some have been published without citing the relevant associations) and two are manuscripts under preparation. The average age of participants was 28.09 (SD = 9.624). 33% of studies listed relationship length as marriage length, with couples married for 9.1 years on average (SD = 6.3). 19% of studies listed relationship length as cohabitation length, with couples cohabiting for 7.7 (SD = 4.7) years on average. The other 48% of studies listed total relationship length, with an average relationship length of 1.8 years (SD = 1.1). Participant numbers and effect sizes for each study can be found in Table 1.

Overall Association

We found a significant association, $r = .134$ ($p = .016$), between EA and relationship satisfaction (95% CI: 0.031,0.235). The test for heterogeneity of effect sizes was not significant ($Q(39) = 35.912$), and the heterogeneity size was low ($I^2 = 0\%$), suggesting that when averaging for moderator values within each study, there is a consistent effect. However, when not averaging for moderator values, the test for heterogeneity was significant ($Q(115) = 141.406$) and the heterogeneity size was moderate ($I^2 = 18\%$), suggesting that the moderators do influence effect size. Therefore, we analyzed effect sizes separately by moderator.

Publication Bias

Figure 1 depicts a funnel plot of effect and sample sizes across the analysis.

Publication bias due to underrepresentation of small studies with low effect sizes would manifest in a lack of studies in the bottom left of the funnel plot. We calculated relevant test statistics. Upon reviewing the funnel plot (where the larger studies are plotted at the top and the smaller studies are plotted at the bottom) and the associated Kendall's tau ($\tau = -0.0758$, $p = 0.4981$) test statistics, it appears that there is no evidence for a publication bias.

The current meta-analysis attempts to account for publication bias by including unpublished results in our analysis; indeed, 8 out of 21 studies (38%), accounting for 27% of participants included in the meta-analysis, were unpublished. To further verify that the effects found are not artifacts created by a small number of large, published studies, we reexamined the total effect ignoring sample sizes for each effect and found a statistically significant effect of $r = 0.126$. Finally, we calculated fail-safe N values. A "classic fail-safe N " (Rosenthal, 1991) calculation indicated that 257 additional non-significant studies would be required to make the analysis results non-significant. Using Orwin's (1963) fail-safe N , we determined that 8 additional studies with an effect size of $r = 0$ would be required to bring the pooled effect size beneath $r = 0.1$. Under a more relaxed assumption, that additional unpublished studies would have the same average effect size of the sampled unpublished studies ($r = 0.07$), 26 more studies would be required.

Moderators

For each categorical moderator, the number of effects, effect sizes, variances, and Q and I^2 statistics for each moderator level are presented in Table 2.

Target and perceiver effects. An ANOVA analog analysis was used to compare the association between perceiver EA and perceiver's relationship satisfaction and the association between perceiver EA and target's relationship satisfaction. The associations were not found to be different ($Q_{\text{between}}(1) = 0.664$, $p = 0.415$).

Emotion type. An ANOVA analog analysis was used to compare associations between relationship satisfaction and EA for positive, negative or overall mental states. Effect sizes differed significantly by emotion type ($Q_{\text{between}(2)} = 7.392, p = 0.025$). Effect sizes for positive EA were significantly smaller than for negative EA ($Q_{\text{between}(1)} = 5.93, p = 0.015$) and smaller than effect sizes for overall EA on a trend level ($Q_{\text{between}(1)} = 3.357, p = 0.067$). Effect sizes for negative EA were larger than for overall EA on a trend level ($Q_{\text{between}(1)} = 2.966, p = 0.085$).

Relationship length. We examined the association between study effect size and average relationship length. In studies in which relationship length was given as time since the couple started dating, a positive association was found as a non-significant trend, indicating that couples who dated longer had a stronger association between EA and relationship satisfaction ($\beta = 0.448, Q_{\text{between}(1)} = 3.522, p = 0.061$). In studies in which relationship length was given as time since the couple started cohabiting, no significant association was found between relationship length and effect size ($\beta = 0.109, Q_{\text{between}(1)} = 0.067, p = 0.795$). In studies in which relationship length was given as time since the couple married, a significant negative association was found; couples married longer had a *weaker* association between EA and satisfaction ($\beta = -0.597, Q_{\text{between}(1)} = 6.821, p = 0.009$).

In our attempt to interpret these results, we noted that under the assumption that couples first date, then begin cohabiting, then marry, the results would indicate a quadratic relation between relationship length and effect size, with effect size rising with relationship length in short (dating) relationships, reaching a plateau in medium-length (cohabiting) relationships, and reducing as relationship length grows in long (married) relationships.

We attempted to examine this interpretation by using two proxies for total relationship length which could be calculated across all studies. We looked for quadratic effects of these proxies, i.e. effects of these relationship length proxies *squared* on the association between

EA and satisfaction. Specifically, we looked for negative quadratic terms, meaning that very high *and* very low values of the proxy – corresponding with short *and* long relationship lengths – would lead to a lower association between EA and satisfaction than medium values of the proxy, corresponding with medium-length relationships.

First, because relationship length was closely associated with age ($\beta = 0.734, p < .0001$), we used age as a proxy. The quadratic relation between age and effect size was found to approach significance ($Q_{\text{between}(2)} = 5.52, p = 0.063, R^2 = 0.149$), with significant linear ($\beta = 2.687, p = 0.021$) and quadratic ($\beta = -2.585, p = 0.027$) terms.

Second, we estimated total relationship length by calculating the average age at which relationships of each type began. Dating relationships began at age 20.116 ($SE = 0.978$). Cohabiting ones began at ages higher by 4.644 ($SE = 1.955$) years, and married couples' ones began at ages higher by 8.874 ($SE = 1.45$) years than dating relationships. Thus, our second proxy was an (obviously skewed) estimate of total relationship length created by adding 4.644 to the length of cohabiting relationships and 8.874 to the length of married couples' relationships. A model checking for a quadratic relation between this estimate and effect size was a non-significant trend ($Q_{\text{between}(2)} = 5.127, p = 0.077, R^2 = 0.127$), with a linear trend ($\beta = 0.93, p = 0.082$) and a significant quadratic term ($\beta = -1.117, p = 0.037$).

Gender. An ANOVA analog analysis was used to compare the association between EA and relationship satisfaction for men and for women. The associations were not significantly different ($Q_{\text{between}(1)} = 0.704, p = 0.401$).

Employed paradigm. An ANOVA analog analysis was used to compare the association between EA and relationship satisfaction between studies measuring EA using content coding, numeric correlations and numeric discrepancy scores. The associations were not significantly different ($Q_{\text{between}(2)} = 0.45, p = 0.798$).

Exploratory Moderators

We examined two other possible moderators in an exploratory manner:

Discussion type. We checked to see if the studies in which the participants were instructed to talk about a conflict or disagreement found different associations from ones in which participants were instructed to support each other. The associations between EA and satisfaction did not differ significantly ($Q_{\text{between}(1)} = 0.089, p = 0.766$).

Recording timing. We also compared the studies in which the participants' thoughts and feelings were recorded whenever they indicated they "had a thought or feeling" with those in which the participants' thoughts and feelings were sampled at fixed intervals (e.g., every 45 seconds). Again, the association between EA and satisfaction was not found to differ significantly ($Q_{\text{between}(1)} = 0.02, p = 0.889$).

Discussion

Our results demonstrate a consistent—if small—association between partners' empathic accuracy and the extent to which they are satisfied with their romantic relationship. Perceiving negative moods and feelings accurately was found to be more closely associated with relationship satisfaction than perceiving positive ones accurately. An interesting quadratic association with relationship length was found, in which the association between EA and relationship satisfaction increased with relationship length up to a point, and then started decreasing. Other possible moderators (e.g., the perceivers' gender) did not affect the association. Below, we discuss the results and their implication for future research on EA.

Moderators

Target vs perceiver. We had no specific expectation for a specific difference (if any) between target and perceiver effects. As noted above, accurate perceivers might be able to manage relationship risks better (Murray, Holmes, J. G. & Collins, 2006), improving their satisfaction. Alternatively, participants whose partners are accurate might feel that their partner is more responsive (Reis & Clark, 2013), responding better and caring more about

their wishes and needs. The findings from existing studies that have examined the distinction between target and perceiver effects are inconclusive (e.g., Cohen et al., 2012).

Our analysis revealed no significant difference between effects focused on targets and ones focused on perceivers. One possible explanation for this null finding is that partners' EA scores are modestly correlated, which could cause these effects to be similar. Future research should therefore examine the underlying causes for target and perceiver effects. Both direct and indirect effects should be examined. Regarding direct effects, the partners' current motivation to communicate more/less clearly might simultaneously increase/decrease both partners' empathic accuracy. Regarding indirect effects, if high accuracy actively elevates the target's satisfaction, it might result in their partner (the perceiver) being more satisfied through various mechanisms (e.g. by increasing support), thereby creating an indirect association between the perceiver's accuracy and their own satisfaction. This could be examined by looking at accuracy and satisfaction over time to see which effect comes first.

Emotion type. We expected EA for negative emotions to be more closely associated with relationship satisfaction than EA for positive emotions, and the results revealed significant moderation in the predicted direction. Studies assessing EA for overall emotions, which did not distinguish emotions along valence lines, fell in the middle. This finding accords with the often-reported pattern of "bad [being] stronger than good" across many areas within psychology (Baumeister et al., 2001), including close relationships (e.g., Rafaeli, Cranford, Green, Shrout, & Bolger, 2008). From an information processing perspective, Anderson's (1981) information integration theory argues that negative information is accorded greater weight than positive information. This difference is also consistent with neurological data showing stronger brain reactions to negative information (Ito, Larsen, Smith, & Cacioppo, 1998). Thus, it seems reasonable to assume that being accurate (or

inaccurate) about negative information such as one's partner's negative emotions might have more impactful consequences than being accurate about one's partner's positive emotions.

With this finding in mind, we believe that future EA studies that examine the antecedents and outcomes of EA should routinely assess the valence of the emotions, and test for differences in EA for both positive and negative emotions, because they seem to be associated with different patterns of results. For example, a variation on Ickes et al.'s (1990) original coding procedure could be used to focus on negative emotions, by coding positive and negative thoughts and feelings separately. This separate coding was in fact a feature of the coding procedure in some of the earliest EA studies by Ickes and his colleagues (see, for example, Graham & Ickes, 1997, and Buysse & Ickes, 1999, and it may be time to reincorporate it into the dyadic interaction paradigm for studying EA.

Gender. We had no specific expectation to find, and in fact did not find, a significant gender difference in the association between EA and relationship satisfaction. As for EA itself, previous reviews have suggested that the occasional gender differences in EA might be due to differences in motivation rather than to differences in actual ability (Ickes, Gesn & Graham, 2000). For example, Klein and Hodges (2001) demonstrated that when men and women are paid to be empathically accurate, gender differences disappear. In their review of the subject, Hodges, Laurent and Lewis (2011) further note that even if gender differences in accuracy are not entirely motivational, they are generally quite small.

Less research exists on gender differences in the association between EA and satisfaction. The few existing studies that examined the question directly are inconsistent in their findings (e.g., Kilpatrick et al., 2002, Cohen et al, 2012). And, indeed, we did not find a gender difference in the association between EA and relationship satisfaction. It should be noted that the absence of a gender moderation does not, of course, preclude the possibility that EA functions differently for women and men (see, e.g., Sened, Yovel, Bar-Kalifa,

Gadassi, & Rafaeli, 2016) or that it has gender-specific effects on outcomes other than satisfaction (e.g., on depression; Gadassi, Mor & Rafaeli, 2011).

Relationship length. Based on the findings of Thomas and Fletcher (1997) and Kilpatrick et al. (2002), Ickes and Simpson (2001) posited that EA plays a progressively weaker role as relationships become lengthier. For this reason, we expected weaker associations between EA and relationship satisfaction in studies in which the relationships were of longer duration. This was indeed the case in studies of married couples. In contrast, in studies of dating couples, the associations became *stronger* as average relationship length increased. Our two follow-up analyses, using estimates of total relationship length, suggest that the different pattern of results in dating vs. married samples is not due to a qualitative difference between the samples; instead, they suggest that the association between EA and relationship satisfaction rises in the first years of relationships and then begins to decline.

Various relational theories (e.g., Reese-Weber, 2015; Solomon & Knobloch, 2004) suggest that most lasting relationships begin in a largely positive, “honeymoon” phase, followed by a middle phase characterized by uncertainty, communication problems, or even aggression, and culminating in a final stable and established phase. In light of these theories, EA may have its largest effect on satisfaction in the middle phase, where it might help partners deal with problems that did not arise in the early honeymoon stage, but which have been resolved (or at least contained) by the final stable and established stage. Future studies could examine this explanation by testing whether relational uncertainty (or other constructs typical in this middle phase) moderate the association between EA and satisfaction.

Employed paradigm and other procedural moderators. All of the moderators we examined that concerned variations on the DIP procedure (Ickes et al., 1990) did not affect the association between EA and satisfaction. In our view, this indicates a general robustness of the DIP procedure, and supports the validity of elaborating or expanding on the classical

paradigm so as to explore empathic accuracy in different types of conversations or using various methods of mood recording (e.g., verbal or numeric coding), which fit researchers' specific research questions, goals, and/or resources.

Overall Effect Size

Our analysis found a relatively small, if positive, overall effect size. There are several possible explanations for this finding. First, our examination of moderators found that EA for negative emotions was more strongly correlated with satisfaction than EA for positive emotions. The small-but-significant moderation effect might have been stronger if EA for both types of emotions was assessed separately in all studies. However, our review included a predominance of studies in which global EA (not broken down by emotion type) was assessed, and the inclusion of those studies may have masked what might have otherwise been an even stronger effect size for emotion type.

A second reason for the relatively low effect sizes in this moderation analysis is that most of the sampled studies focused either on short-term relationships (couples dating 1.8 years on average) or on long-term relationships (couples cohabiting for 7.7 years on average or married for 9.1 years on average). Our findings of curvilinear moderation by relationship length suggest that the association between EA and satisfaction would be low for very short and very long relationships, and higher for those of medium length. Indeed, in one of the only studies that examined medium-length relationships (Kilpatrick et al., 2002), associations between EA and satisfaction were very high (over 0.35 for both men and women). Additional studies of such relationships are needed to see if Kilpatrick et al.'s finding replicates.

A third reason for the relatively low overall effect size may have to do with qualities of the EA index. Specifically, the DIP procedure that was used in all of the studies in this meta-analysis may be particularly useful for capturing EA at a specific moment in time. Indeed, Kilpatrick et al. (2002) found very low test-retest reliability for the DIP-based EA

index assessed across one-year intervals (.11 and .07). Thus, the weak effect size for stage-of-relationship might be attributable, at least in part, to studies that tried to link a local index (i.e., EA assessed during a specific interaction) with a more global outcome (i.e., long-term relationship satisfaction). Future studies could examine this idea by assessing EA using a more global index (e.g., using diary methods) or by assessing the outcome measure (relationship satisfaction) using a more momentary index.

A fourth explanation for the small effect size found might be the presence of motivated inaccuracy, a phenomenon discussed in the introduction in which perceivers with higher EA for the partner's threatening thoughts and feelings report feeling *less* close to their partners than perceivers with lower EA do (Simpson et al., 1995; Simpson et al., 2003). Ickes and Simpson (2001) propose that motivated inaccuracy is an *exception* to the "rule" that greater empathic accuracy is generally good (or at least not bad) for relationships, a predicted generalization that was validated in the current study. However, because the conditions for motivated accuracy—mainly, relationship threat—could have been met in some of the reviewed samples, it is possible that the inclusion of these samples reduced the overall effect size. Additional research on motivated inaccuracy could help clarify in which situations it is the dominant effect. We strongly encourage such research.

An important corollary of the small overall effect size is that most studies in the area, which tend to have samples of several dozen (but rarely more than 100) couples, can be underpowered to detect any effect. Our power analysis, using the calculator developed by Ackerman, Ledermann, and Kenny (2016) for dyadic studies, shows that to detect the association found in the current study between relationship satisfaction and EA ($r = .134$) with a power ($1 - \beta$) of 0.8, $N = 211$ couples would be required. Even detecting the (stronger) association found between satisfaction and *negative* EA would require $N = 127$ couples to achieve a similar power of 0.8. Accordingly, future studies should either be designed to

include larger samples or use the recommendations above to enlarge the effect size. Until that happens, our meta-analytic results suggest that many of the null findings should be taken with a grain of salt, as an underlying effect (albeit a small one) might still exist.

Limitations

The present meta-analysis has a few limitations that should be noted. The first limitation is its reliance on correlational data. Our results cannot help us determine whether high satisfaction causes partners to be accurate, high accuracy causes partners to be satisfied, or, alternatively, some third variable causes both. Future studies should consider adopting experimental designs in which causality and causal direction can be established.

Second, the number of studies in the area is not very large. DIP studies are difficult to conduct because they require having both dyadic partners come to the lab simultaneously to engage in a relatively lengthy and somewhat intrusive procedure. Unfortunately, some DIP studies have not obtained a measure of relationship satisfaction as an outcome variable. Other studies may have obtained it, but then failed to report its association with EA (possibly because of null findings that can result from low power, as we discussed earlier).

Third, our report of a quadratic association between relationship length and effect sizes was based, in part, on proxy indices created to estimate total relationship length in some of the studies. Future studies should employ a longitudinal design that enables them to test this quadratic association directly, or a cross-sectional design that studies couples of varying relationship lengths and attempts to tease apart the effects of age and relationship duration.

Finally, it should be noted that the studies included in the analysis, while ranging over multiple countries across four continents, come in general from western, educated, industrialized, rich and democratic (WEIRD; Henrich, Heine & Norenzayan) societies. Although not enough studies listed ethnicity and socio-economic status of participants to enable a quantified analysis, it is probable that couples come from relatively high SES

backgrounds even within those societies. There are reasons to believe that results would somewhat generalize (e.g., Dalton's (2005) EA results in couples' therapy for a general Canadian population replicated by Ahmad (2012) with couples of a South-Asian Canadian minority), but some moderators could prove to be different. For example, motivated inaccuracy could prove more significant in populations with lower general satisfaction, and relationship length effects could differ in cultures with different culture expectations for relationship development (Hatfield & Rapson, 1996). EA studies in other countries and among more diverse populations would certainly be welcome in this regard.

Conclusion and Implications

The current meta-analysis reviewed 21 studies reporting the association between EA and relationship satisfaction. Our overall finding shows a stable, though small, positive association between these two constructs. Even more important, our moderation findings help to specify the conditions in which increased EA is most likely to affect couples' lives. Specifically, the positive association between EA and relationship satisfaction was significantly stronger when EA was assessed in regards to the partner's negative emotions. This association was also stronger in studies of couples with moderate-length relationships, as compared to couples with shorter- and longer-length relationships.

From a practical perspective, the findings of the current study suggest that clinicians should look into partners' understanding of each other, especially regarding negative emotions, as a possible cause of relationship problems, as well as a potential pathway for improving satisfaction. This would be more strongly indicated for couples in medium-length relationships (about a few years long). Of course, further studies would still be required to test specific assessment methods and actual interventions.

We hope this review, and the recommendations it provides for researchers interested in EA, helps to stimulate further research on this important relationship process.

References

- Ackerman, R. A., Ledermann, T., & Kenny, D. A. (2016). Power analysis for the Actor-Partner Interdependence Model. *Manuscript under preparation*.
- Ahmad, S. S. (2012). *Enhancing the Quality of South Asian Marriages Through Systemic-Constructivist Couple Therapy (SCCT): Investigating the Role of Couple Identity in Marital Satisfaction* (unpublished doctoral dissertation). York University, Toronto, Ontario, Canada.
- Anderson, N. H. (1981). *Foundations of information integration theory*. New York, NY: Academic Press.
- Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of general psychology, 5(4)*, 323.
- Bentley, C. G. (2010). Does Empathic Accuracy Mediate the Relationships Between Individual Psychological Characteristics and Adolescent Romantic Relationship Functioning?. *Graduate Theses and Dissertations, 542*.
- Blackmon, A. D. (1999). *Empathy in Marriage: Implications for Marital Satisfaction and Depression* (unpublished doctoral dissertation). Yale University, New Haven, Connecticut.
- Buysse, A., & Ickes, W. (1999). Topic-relevant cognition and empathic accuracy in laboratory discussions of safer sex. *Psychology and Health, 14*, 351-366.
- Clements, K., Holtzworth-Munroe, A., Schweinle, W., & Ickes, W. (2007). Empathic accuracy of intimate partners in violent versus nonviolent relationships. *Personal Relationships, 14(3)*, 369-388.
- Cohen, S., Schulz, M. S., Weiss, E., & Waldinger, R. J. (2012). Eye of the beholder: The individual and dyadic contributions of empathic accuracy and perceived empathic effort to relationship satisfaction. *Journal of Family Psychology, 26(2)*, 236.

- Dalton, J. (2005). *Increasing marital satisfaction in clinically distressed couples: The role of empathic accuracy and "we-ness"* (Unpublished doctoral dissertation). York University, Toronto, Ontario, Canada.
- Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of personality and social psychology*, *44*(1), 113.
- Davis, M. H., & Kraus, L. A. (1997). Personality and empathic accuracy. In W. Ickes (Ed.), *Empathic accuracy* (pp. 117-143). New York: Guilford Press.
- Drwal, J. M. (2003). *The relationship among attachment styles, empathic accuracy, and relationship satisfaction*. (unpublished doctoral dissertation). University of Connecticut, Storrs, Connecticut.
- Fletcher, G. J. O., & Kerr, P. S. G. (2010). Through the Eyes of Love: Reality and Illusion in Intimate Relationships. *Psychological Bulletin*, *136*, 627-658.
- Gadassi, R., Mor, N., & Rafaeli, E. (2011). Depression and Empathic Accuracy in Couples: An Interpersonal Model of Gender Differences in Depression. *Psychological Science*, *22*(8), 1033-1041.
- Gleason, K.A., Jensen-Campbell, L.A., & Ickes, W. (2009). The role of empathic accuracy in adolescents' peer relations and adjustment. *Personality and Social Psychology Bulletin*, *35*, 997-1011.
- Graham, T., & Ickes, W. (1997). When women's intuition isn't greater than men's. In W. Ickes (Ed.), *Empathic accuracy* (pp. 117-143). New York: Guilford Press.
- Harvey, J. H., Hendrick, S. S., & Tucker, K. (1988). Self-report methods in studying personal relationships. In S. W. Duck and W. Ickes (Eds.), *Handbook of research on personal relationships* (pp. 99-113). New York: John Wiley.
- Hatfield, E., & Rapson, R. L. (1996). *Love and sex: Cross-cultural perspectives*. Needham Heights, MA, US: Allyn & Bacon.

- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and brain sciences*, *33*(2-3), 61-83.
- Howland, M., & Rafaeli, E. (2010). Bringing everyday mind reading into everyday life: Assessing empathic accuracy with daily diary data. *Journal of personality*, *78*(5), 1437-1468.
- Haugen, P. T., Welsh, D. P., & McNulty, J. K. (2008). Empathic accuracy and adolescent romantic relationships. *Journal of adolescence*, *31*(6), 709-727.
- Hodges, S. D., Laurent, S. M., & Lewis, K. L. (2011). Specially motivated, feminine, or just female: Do women have an empathic accuracy advantage. In Smith, J. L., Ickes, W., Hall, J. A. & Hodges, S.D. (Eds), *Managing interpersonal sensitivity: Knowing when and when not to understand others*, 59-74. NY: Nova Science Publishers.
- Hedges, L. V., & Olkin, I. (1985). *Statistical models for meta-analysis*. New York: Academic Press.
- Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *bmj*, *327*(7414), 557-560.
- Howland, M., & Simpson, J. A. (2014). Attachment orientations and reactivity to humor in a social support context. *Journal of Social and Personal Relationships*, *31*(1), 114-137.
- Ickes, W. (1993). Empathic accuracy. *Journal of Personality*, *61*(4), 587-610.
- Ickes, W. (2003). *Everyday mind reading: Understanding what other people think and feel*. Amherst, NY: Prometheus Books.
- Ickes, W., Gesn, P.R., & Graham, T. (2000). Gender differences in empathic accuracy: Differential ability or differential motivation? *Personal Relationships*, *7*, 95-109. doi: 10.1111/j.1475-6811.2000.tb00006.x

- Ickes, W., & Hodges, S. D. (2013). Empathic accuracy in close relationships. In J. A. Simpson & L. Campbell (Eds.), *The Oxford handbook of close relationships* (pp. 116–136). New York, NY: Oxford University Press.
- Ickes, W., & Simpson, J. (1997). Managing empathic accuracy in close relationships. In W. Ickes (Ed.), *Empathic accuracy* (pp. 218-250). New York: Guilford Press.
- Ickes, W., & Simpson, J. (2001). Motivational aspects of empathic accuracy. In G.J.O. Fletcher & M.S. Clark (Eds.), *Interpersonal Processes: Blackwell Handbook in Social Psychology* (pp. 229-249). Oxford, UK: Blackwell.
- Ickes, W., Stinson, L., Bissonnette, V., & Garcia, S. (1990). Naturalistic social cognition: Empathic accuracy in mixed-sex dyads. *Journal of Personality and Social Psychology, 59*(4), 730.
- Ito, T. A., Larsen, J. T., Smith, N. K., & Cacioppo, J. T. (1998). Negative information weighs more heavily on the brain: the negativity bias in evaluative categorizations. *Journal of personality and social psychology, 75*(4), 887.
- Kilpatrick, S. D., Bissonnette, V. L., & Rusbult, C. E. (2002). Empathic accuracy and accommodative behavior among newly married couples. *Personal Relationships, 9*(4), 369-393.
- Klein, K. J., & Hodges, S. D. (2001). Gender differences, motivation, and empathic accuracy: When it pays to understand. *Personality and Social Psychology Bulletin, 27*(6), 720-730.
- Knapp, M. L. (1978). *Social intercourse: From greeting to goodbye*. Boston, MA: Allyn & Bacon.
- Leonard, M. T. (2008). *Couples' Understanding of Pain and Interaction Dynamics: The CUPID Study*. (unpublished dissertation). Wayne State University, Detroit, Michigan.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

- Long, E. C., & Andrews, D. W. (1990). Perspective taking as a predictor of marital adjustment. *Journal of Personality and Social Psychology*, *59*(1), 126.
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Research methods & reporting-preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Bmj* *338*(7716).
- Overall, N. C., Fletcher, G. J., & Kenny, D. A. (2012). When bias and insecurity promote accuracy mean-level bias and tracking accuracy in couples' conflict discussions. *Personality and Social Psychology Bulletin*, *38*(5), 642-655.
- Overall, N. C., Fletcher, G. J., Simpson, J. A., & Fillo, J. (2015). Attachment insecurity, biased perceptions of romantic partners' negative emotions, and hostile relationship behavior. *Journal of personality and social psychology*, *108*(5), 730.
- Rafaeli, R., Gadassi, R., Howland, M., Boussi, A., & Lazarus, G. (2016) *Seeing bad does good: Accuracy regarding negative moods predicts positive relationship outcomes*. Manuscript under review.
- Rafaeli, E., Cranford, J. A., Green, A. S., Shrouf, P. E., and Bolger, N. (2008) The good and bad of relationships: How social hindrance and social support affect relationship feelings in daily life. *Personality and Social Psychology Bulletin*, *34*(12), 1703-1718.
- Reese-Weber, M. (2015). Intimacy, communication, and aggressive behaviors: Variations by phases of romantic relationship development. *Personal Relationships*, *22*(2), 204-215.
- Reis, H. T., & Clark, M. S. (2013). Responsiveness. In L. Campbell & J. Simpson (Eds.), *The Oxford handbook of close relationships*. (pp. 400-423) New York: Oxford University Press.
- Sened, H., Yovel, I., Bar-Kalifa, E., Gadassi, R., & Rafaeli, E. (2016). Now You Have My Attention: Empathic Accuracy Pathways in Couples and the Role of Conflict. *Emotion*, In Press.

- Sillars, A. L., Pike, G. R., Jones, T. S., & Redmon, K. (1983). Communication and conflict in marriage. *Communication yearbook*, 7, 414-429.
- Sillars, A. L., & Scott, M. D. (1983). Interpersonal perception between intimates an integrative review. *Human Communication Research*, 10(1), 153-176.
- Simpson, J. A., Ickes, W., & Blackstone, T. (1995). When the head protects the heart: Empathic accuracy in dating relationships. *Journal of Personality and Social Psychology*, 69(4), 629. doi: 10.1037/0022-3514.69.4.629
- Simpson, J.A., Ickes, W., & Oriña, M. (2001). Empathic accuracy and pre-emptive relationship maintenance. In J. Harvey & A. Wenzel (Eds.), *Close romantic relationships: Maintenance and enhancement* (pp. 27-46). Mahwah, N.J.: Erlbaum.
- Simpson, J. A., Oriña, M. M., & Ickes, W. (2003). When accuracy hurts, and when it helps: A test of the empathic accuracy model in marital interactions. *Journal of Personality and Social Psychology*, 85, 881-893. doi: 10.1037/0022-3514.85.5.881
- Simpson, J. A., Kim, J. S., Fillo, J., Ickes, W., Rholes, W. S., Oriña, M. M., & Winterheld, H. A. (2011). Attachment and the management of empathic accuracy in relationship-threatening situations. *Personality and Social Psychology Bulletin*, 37(2), 242-254.
- Solomon, D. H., & Knobloch, L. K. (2004). A model of relational turbulence: The role of intimacy, relational uncertainty, and interference from partners in appraisals of irritations. *Journal of Social and Personal Relationships*, 21(6), 795-816.
- Stinson, L., & Ickes, W. (1992). Empathic accuracy in the interactions of male friends versus male strangers. *Journal of Personality and Social Psychology*, 62(5), 787.
- Suppes, A., Zaki, J., Stadler, G., Bolger, N. (2016) *Stress and Social Support Study*. manuscript under preparation.

- Swann, W. B., Jr., Chang-Schneider, C., & Angulo, S. (2007). Self-verification in relationships as an adaptive process. In J. Wood, A. Tesser, & J. Holmes (Eds.), *Self and relationships* (pp. 49-72). New York, NY: Psychology Press.
- Thomas, G., & Fletcher, G. J. (2003). Mind-reading accuracy in intimate relationships: assessing the roles of the relationship, the target, and the judge. *Journal of personality and social psychology*, *85*(6), 1079.
- Thomas, G., Fletcher, G. J., & Lange, C. (1997). On-line empathic accuracy in marital interaction. *Journal of Personality and Social Psychology*, *72*(4), 839.
- Verhofstadt, L. L., Buysse, A., Ickes, W., Davis, M., & Devoldre, I. (2008). Support provision in marriage: the role of emotional similarity and empathic accuracy. *Emotion*, *8*(6), 792.
- Verhofstadt, L., Devoldre, I., Buysse, A., Stevens, M., Hinnekens, C., Ickes, W., & Davis, M. (2016). The Role of Cognitive and Affective Empathy in Spouses' Support Interactions: An Observational Study. *PloS one*, *11*(2).
- Wilhelm, P., & Perrez, M. (2004). How is my partner feeling in different daily-life settings? Accuracy of spouses' judgements about their partner's feelings at work and at home. *Social Indicators Research*, *67*, 183-246. doi: 10.1023/B:SOCI.00000007339.48649.20
- Wilson, D. B. (2011). *Meta-analysis macros for SAS, SPSS, and Stata*. Retrieved, June 16th, 2016, from <http://mason.gmu.edu/~dwilsonb/ma.html>.
- Winczewski, L. A., Bowen, J. D., & Collins, N. L. (2016). Is Empathic Accuracy Enough to Facilitate Responsive Behavior in Dyadic Interaction? Distinguishing Ability from Motivation. *Psychological science*, *27*(3), 394.
- Zell, E., Krizan, Z., & Teeter, S. R. (2015). Evaluating gender similarities and differences using metasynthesis. *American Psychologist*, *70*(1), 10.

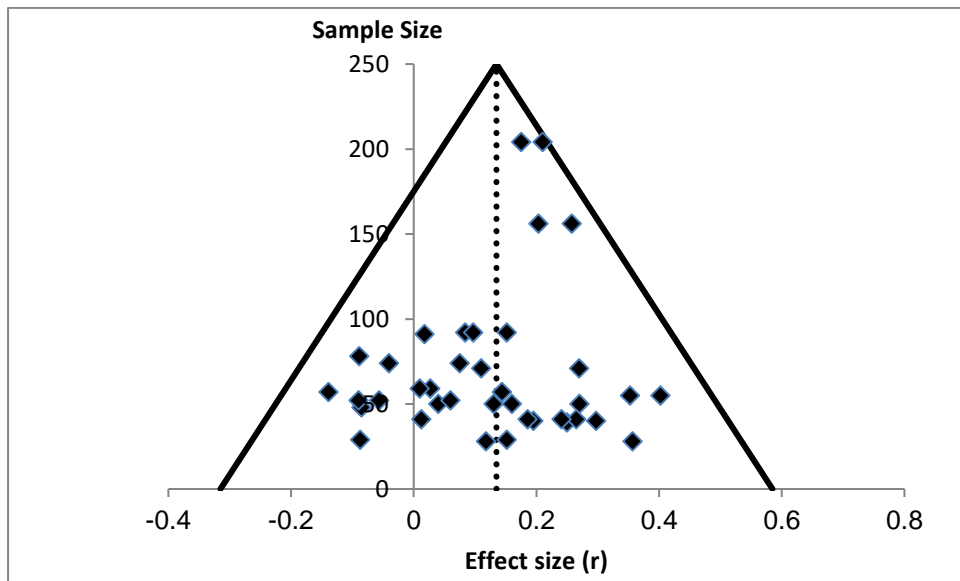


Figure 1. Funnel plot of studies by sample and effect size.

Table 1. Studies included in meta-analysis and sample characteristics.

Study	N (individuals)	Effect Size (r) ¹
Ahmad(2012) ³	28	0.357
Bentley(2010) ³	184	0.084,0.097
Blackmon(1999) ³	100	0.16,0.13
Cohen, Schulz, Weiss, & Waldinger (2012) ⁵	312	0.258,0.203
Clements, Holtzworth-Munroe, Schweinle, & Ickes (2007) ⁵	142	0.11,0.27
Dalton(2005) ²	28	0.118
Drwal(2003) ²	118	0.027,0.01
Haugen, Welsh, & McNulty (2008) ⁵	408	0.175,0.21
Howland & Simpson (2014) ⁴	82	0.265,0.241
Kilpatrick, Bissonette, & Rusbult (2002) ⁵		
Time 1	110	0.353,0.402
Time 2	104	-0.056,0.06
Time 3	80	0.195,0.297
Leonard (2008) ³	48	-0.085
Overall, Fletcher, & Kenny (2012); Overall, Simpson, Fletcher, & Fillo (2015) ⁴	114	-0.139,0.144
Thomas & Fletcher (2003) ⁵	100	0.27,0.04
Thomas, Fletcher, & Lange (1997) ⁵	148	0.075,-0.04
Rafaeli, Gadassi, Howland, Boussi & Lazarus (2016) ²	141	0.121,0.086
Simpson et al. (2011) ⁴	184	0.018,0.152
Suppes, Zaki, Stadler, & Bolger (2016) ²	156	-0.089
Verhofstadt, Buysse, Ickes, Davis, & Devoldre (2008) ⁴	58	0.152, -0.088
Verhofstadt et al. (2016) ⁴	82	0.012,0.186
Winczewski, Bowen, & Collins (2016) ⁴	91	-0.08,0.25

¹ Effect sizes averaged for the whole study across moderators (e.g. EA for positive and negative moods both recorded) by converting to fisher's z, averaging and converting back; When women and men were separable, first effect size is for women.

²Manuscript under preparation ³Unpublished Dissertation ⁴Published manuscript, additional data requested from authors ⁵Published data

Table 2. Number of effects, effect sizes, and homogeneity for categorical moderators

Study	k	n	r	95% CI	Q	I ²
Total	40	2739	0.134(0.053)*	0.031,0.235	35.912	0%
Target vs Perceiver						
Target Satisfaction	30	2190	0.153(0.055)*	0.046,0.258	35.755	18.9%
Perceiver Satisfaction	37	2605	0.13(0.055)*	0.022,0.235	36.592	1.6%
Emotion Type						
Positive	10	1136	0.068(0.115)	-0.158,0.294	15.994 [†]	43.7%
Negative	10	1136	0.171(0.073)*	0.029,0.314	11.787	23.6%
Overall	32	1721	0.121(0.071) [†]	-0.018,0.255	25.318	0%
Gender						
Female	18	1284	0.126(0.079)	-0.03,0.283	15.583	0%
Male	18	1273	0.159(0.075)*	0.014,0.297	12.171	0%
Employed paradigm						
Content Coding	29	1543	0.13(0.072) [†]	-0.011,0.266	21.647	0%
Correlations	13	1314	0.133(0.076) [†]	-0.016,0.276	16.23	26.1%
Discrepancy Scores	6	398	0.095(0.164)	-0.226,0.398	0.784	0%

[†]p<.1 *p<.05