

Fast and Slow Empathic Perceptions in Couples' Daily Lives Use Different Cues

Haran Sened¹
Bar Ilan University

Eran Bar-Kalifa²
Ben Gurion University

Rony Pshedetzky-Shochat¹
Bar Ilan University

Marci Gleason³
University of Texas at Austin

Eshkol Rafaeli^{1,4,5}
Bar Ilan University, Columbia University

Authors Notes

¹Department of Psychology, Bar-Ilan University

²Department of Psychology, Ben Gurion University

³Department of Human Development and Family Sciences, University of Texas at Austin

⁴Gonda Multidisciplinary Brain Research Center, Bar-Ilan University

⁵Department of Psychology, Barnard College, Columbia University

This work was supported by a grant from the US-Israel Binational Science Fund (BSF) awarded to the fourth and fifth authors and by a grant from the John Templeton Foundation awarded to the fifth author. The first author is grateful to the Azrieli Foundation for the award of an Azrieli Fellowship supporting his work.

Correspondence concerning this article should be addressed to Haran Sened, Department of Psychology, Bar-Ilan University, Ramat-Gan 5290002, ISRAEL. E-mail: haranse@gmail.com.

Accepted to Affective Science. This version may not reflect exactly the final published version and is not the copy of record.

Abstract

Empathic accuracy, the ability to infer another person's emotions, thoughts, and other fleeting mental states, has been linked to *assumed similarity* (wherein the perceiver assumes that another person's mental states are similar to their own) and *direct accuracy* (wherein the perceiver uses various external cues to reach their judgment). Previous research has linked this component model, as well as dual process models, to neuroscientific models of empathy, but has not linked these components with dual process accounts directly. Thus, we examined whether assumed similarity involves rapid (type-1) processing while direct accuracy involves slower (type-2) inferences. In three dyadic daily diary samples (Total N=262 romantic couples), we examined associations between both components and response times. As expected, direct accuracy, but not assumed similarity, was associated with slower response times. Our findings suggest links between previously disparate lines of research and identify situations which may tip the balance between the empathic components.

Fast and Slow Empathic Perceptions in Couples' Daily Lives Use Different Cues

Almost all social interactions involve participants' attempts to identify and understand each others' thoughts and emotions. These efforts utilize an ability often termed *empathic accuracy* (Hodges et al., 2015). Successful empathic accuracy is associated with a variety of beneficial outcomes such as better interpersonal interactions (Schmid Mast & Hall, 2018) and more satisfying romantic relationships (Sened, Lavidor, et al., 2017). On the other hand, in specific circumstances, such as when a relationship is under threat, greater empathic accuracy may be detrimental (Simpson et al., 1995).

Empathic accuracy has been studied in various areas of research (sometimes under different terms). Three of the prominent ones, detailed below, are neuroscientific research, dual-process models, and close relationship research. While integrating these lines of research has great potential, few efforts have been made to do so to date (Zaki & Ochsner, 2011). As we will detail below, there is some evidence connecting dual process and neuroscientific models, and some evidence suggesting connections between neuroscientific models and ones drawn from relationship research. The current study aims to strengthen this triple link by linking close relationship and dual process research, showing that the assumed similarity vs. direct accuracy distinction, common in close relationship studies, fits a dual process model. Below, we briefly summarize main models from all three lines of research and existing studies connecting them.

Close relationship Research

Close relationship studies, usually looking at empathic accuracy between members of romantic couples, suggest that empathic accuracy inferences can be partitioned into two main components (Kenny & Acitelli, 2001; Wilhelm & Perrez, 2004). The first, referred to as *assumed*

similarity, taps the extent to which perceivers' inferences are associated with their own emotions. It may reflect perceivers' tendency to assume that targets' emotions are similar to their own; importantly, if the assumption is correct (i.e., target and perceiver emotions are indeed similar), assumed similarity will *indirectly* lead to accuracy. The second, referred to as *direct accuracy*, taps the extent to which perceivers' inferences are "correct" – i.e., associated with the targets' actual emotions after adjusting for any effects of assumed similarity. It reflects all sources of information, including both verbal and non-verbal cues, other than assumed similarity (Hall & Schmid Mast, 2007). Both components of empathic accuracy have been researched extensively (e.g., Kouros & Papp, 2019; Powers et al., 2011)

Dual Process Research

Dual process models (Chaiken & Trope, 1999) distinguish between type-1 mental processes, which are autonomous and do not require attention or working memory resources, and type-2 processes, which are slower and sequential, and do utilize working memory resources. Dual process models have been proposed for a wide variety of psychological phenomena; examples from areas pertinent to empathic accuracy include emotion regulation (Gyurak et al., 2011) and impression formation (Wyer & Srull, 2014). Similarly, within empathic accuracy research, Ma-Kellams and Lerner (2016) demonstrated that participants who were encouraged to think intuitively (i.e., use type-1 processes) had less accurate empathic inferences than those who were encouraged to think reflectively (i.e., use type-2 processes).

Identifying dual processes in psychological phenomena can help generate a variety of hypotheses, as the two types of processing have been associated with various individual differences (e.g., in working memory; Barrett et al., 2004) and situational variables (e.g.,

intuitive vs. rational modes of thought; Shenhav et al., 2012). Such situational variables can then be experimentally manipulated, allowing for tighter causal inferences.

Neuroscientific Research

Social-neuroscience models of empathy distinguish between two brain systems involved in empathic inferences (e.g., Shamay-Tsoory, 2011; Singer, 2006; Zaki & Ochsner, 2011)¹: Emotional empathy, which involves sharing another person's emotions and assuming correctly that the shared emotion reflects the other person's emotion, tends to be fast, intuitive, and automatic. Cognitive empathy, which involves explicitly processing all available information, tends to be slow and deliberate.

The distinction between these two systems helps clarify the ways in which various biological mechanisms (e.g., hormones; Bethlehem et al., 2013) might play a role in empathic accuracy and how it might operate in individuals characterized by various conditions known to have neural correlates (e.g., borderline personality disorder; Dziobek et al., 2011).

Integrating lines of research

Some links between these lines of research have already been demonstrated in previous studies. For example, Bohl and van den Bos (2012) and Spunt and Lieberman (2013) linked neuroscientific and dual process models by showing that emotional empathy involves type-1 processes and cognitive empathy involves type-2 processes. Other studies have linked neuroscientific models with the accuracy components model in close relationship research. For

¹ Detailing the differences between terminologies is beyond the scope of this article; for clarity, we use terms from Shamay-Tsoory's (2011) review. We are grateful for Simone Shamay-Tsoory's comments on an earlier draft of this manuscript.

example, Sened, Yovel, and colleagues (2017) provided evidence supporting the identification of emotional empathy with assumed similarity and of cognitive empathy with direct accuracy.

In an attempt to integrate these findings, we outline a conceptual framework for understanding empathic accuracy according to these three lines of research (see Figure 1). The framework suggests that direct accuracy results from type-2 processes carried out by the cognitive empathy system, whereas assumed similarity results from type-1 processes carried out by the emotional empathy system. It can help cross-pollinate these lines of research, leading to new hypotheses, research methods, and possible applications. However, the framework suggests that dual process models should also be linked directly to the accuracy components model, a link for which we could not find evidence in previous studies.

--- Figure 1 about here ----

The current study

Thus, the current study aims to provide more evidence to support the suggested conceptual framework by completing the “missing link” between these three lines of empathic accuracy research, the link between the accuracy components model from close relationship research and dual process models. In line with the proposed framework, we hypothesized that assumed similarity reflects type-1 processes, as both have been linked to emotional empathy; in contrast, direct accuracy may require slower and more involved type-2 processes, as both have been linked to cognitive empathy. Thus, we expected slower inferences to reflect greater direct accuracy but not greater assumed similarity, as in the time scale typical of our methods (tens of seconds) fast type-1 processes would have already had ample time to achieve their final results.

To test our hypothesis, we examined empathic inferences in three daily diary samples of romantic couples, and measured the time taken to reach each inference. The main reason we looked at this population is that recruiting couples is a simple way to recruit dyads of people who can be reasonably expected to be motivated and have opportunity to note each other's emotions on a daily basis. Importantly, we are looking at the way spending more time on empathic inferences changes the involvement of these two components in arriving at the final inference; our hypothesis is indifferent as to why participants spend more time on specific days (e.g., higher motivation to be accurate on a specific day, or having more free time to spend on the task).

Method

De-identified data used in the study, analysis code and a list of measures participants completed which were not used in this study are openly available at <https://osf.io/wsuta/>

All data collection was approved by institution IRBs.

Participants

Sample 1. Eighty-nine cohabiting mixed-gender couples were recruited to take part in a study on dyadic relationships in which they completed a background questionnaire and a 4-week daily diary. To be included, participants were required to be over 18 years old. Nine couples were excluded because at least one partner had completed less than 6 entries of the required 28 diary entries². Demographics are reported in Table 1. T-test analyses³ revealed that participants who remained in the study were significantly older (Mean age difference in years 4.087, 95% CI

² We are not aware of any research on the best number of entries to count as valid; We have used 6 as a cutoff in our earlier studies (e.g., Sened, Yovel, et al., 2017) , and it was pre-registered in Sample 3.

³ These analyses were performed on couple means for all samples to avoid dependence issues; performing them on each participant separately did not meaningfully change the results.

(1.723,6.452), $t(80.712) = 3.44$, $p < .001$) and with longer relationships (Mean relationship length difference in years 4.013, 95% CI (1.862,6.165), $t(77.314) = 3.714$, $p < .001$) than participants who were excluded. Wilcoxon tests revealed no difference in income or education level between participants who remained in the study and participants who were excluded ($p > .1$). Participants did not receive any compensation for this study.

Sample 2. After results from Study 1 confirmed our hypothesis (see the Results section, below), we ran similar analyses on data from a large sample of new parents which was already available at the time (e.g., Sened et al., 2019). This was done to replicate our results and to ensure that they remain consistent with couples undergoing a specific major life transition – in this case, into parenthood. One hundred and eight mixed-gender couples, expecting their first child, were recruited for a larger study on the transition to parenthood. To be included, participants were required to be over 18 years old, and expecting a single child (i.e. not twins). Five couples left the study before beginning the daily diary component, and 3 more were excluded because at least one partner had completed less than 6 entries of the required 21 diary entries. Demographics are reported in Table 1. T-test analyses revealed no difference in age or relationship length between participants who remained in the study and participants who dropped out ($p > .1$). Wilcoxon tests revealed no difference in income or education level between participants who remained in the study and participants who were excluded ($p > .1$). Participants received the equivalent of 150 USD for completing the daily diary alongside other procedures, including a lab meeting.

Sample 3. To ensure that our findings do not rely (unintendedly) on “researcher degrees of freedom” (Wicherts et al., 2016), we pre-registered our hypotheses and analytic code, and recruited a new sample. Ninety-five cohabiting couples were recruited in a manner similar to

Sample 1, but couples completed diaries for 3 weeks only. Eight couples were excluded because at least one partner had completed less than 6 entries of the required 21 diary entries.

Demographics are reported in Table 1. T-test analyses revealed no difference in age or relationship length between participants who remained in the study and participants who dropped out ($p > .1$). Wilcox tests revealed no difference in income or education level between participants who remained in the study and participants who were excluded ($p > .1$). Participants did not receive any compensation for this study. Three couples (3.1%) were same-gender women couples, all other couples were mixed-gender.

----Table 1 about here-----

Procedure

Sample 1. Participants completed a daily diary questionnaire every day for 28 days using the Qualtrics platform. They were instructed to complete the questionnaire an hour before they go to sleep and were asked not to share their answers with their partners. Participants who missed a questionnaire could complete it during the following day; they were asked to reply as if they were completing it at the appropriate time (i.e., the previous evening). Over eighty-three percent of the entries (3760/4480) were completed. The daily diary included a self-reported mood questionnaire, several items unrelated to the current study, and then the partner-report (i.e., empathic inference) mood questionnaire.

Data cleaning included removing entries in which the whole diary (including various questionnaires not reported in this study) was completed in less than two minutes, which was assumed to indicate careless completion, leaving 3678 entries. We then removed the 5% of entries with the longest response times for self-reported moods and for empathic inferences,

leaving 3354 entries.

Sample 2. Three months after their first child was born, participants completed a daily diary questionnaire every day for 21 days in the same manner used in Sample 1. Close to ninety-seven percent of the entries (4067/4200) were completed.

Data cleaning included removing entries in which the whole diary (including various questionnaires not reported in this study) was completed in less than two minutes, which was assumed to indicate careless completion. No entries were removed. We then removed the 5% of entries with the longest response times for self-reported moods and for empathic inferences, leaving 3686 entries.

Sample 3. Procedure for Sample 3 was identical to Sample 1, on the same site. Items not analyzed in this study were different, and the diary was only 21 days. Data collection and full analysis code was preregistered at <https://osf.io/cfhsn>⁴. Over eighty-five percent of the entries (3128/3654) were completed.

Data cleaning included removing entries in which the whole diary (including various questionnaires not reported in this study) was completed in less than two minutes, which was assumed to indicate careless completion, leaving 3123 entries. We then removed the 5% of entries with the longest response times for self-reported moods and for empathic inferences, leaving 2835 entries. Unfortunately, though these cleaning steps were pre-planned and pre-coded (and thus performed in a manner identical to the previous two samples), we failed to mention

⁴ There are two changes between the code used to produce the final results presented here and the pre-registered analytic code. The first is the data cleaning procedure reported below. The second is a small change to the encoding of the Study 3 gender variable, done to correctly account for same-gender couples. The results with and without this second edit were similar; no non-significant effects became significant or vice versa.

them in the pre-registration document.

Measures

Negative affect. To be able to calculate empathic accuracy, we first measured participants' self-reported moods and their reports of their partners' moods using a version of Lorr and McNair's (1971) Profile of Mood States (POMS) questionnaire, adapted for daily diary use by Cranford and colleagues (2006). For self-reported moods, the questionnaire prompt was "Please indicate the extent to which **you** are experiencing these feelings **at the moment, in the evening**". For reports on partner moods the prompt was "Please indicate the extent to which you think that **your partner** is experiencing these feelings **at the moment, in the evening**". To avoid confusion, from now on we will refer to reports on partner moods as *empathic inferences*.

In Sample 1 and Sample 3, participants rated six positive and negative moods (e.g., anger, calmness) on a 0-100 sliding scale, with 0 marked "not at all" and 100 marked "to a very large extent". In Sample 2 participants rated the same moods using 3 items each, on a 5-point Likert-type scale, with 0 marked "not at all" and 5 marked "extremely". The current study focused only on negative moods, as existing evidence from psychology research in general (cf. Baumeister et al., 2001) and from empathic accuracy studies in couples specifically (e.g., Sened, Lavidor, et al., 2017) shows that across a wide range of phenomena, effects for negative emotions are more pronounced than for positive ones⁵. Thus, in each sample scores for the three negative mood items (anxiety, anger, and sadness) were averaged to create a general negative mood measure.

⁵ Following a question raised about an earlier draft of the manuscript, we ran all analyses on accuracy regarding positive emotions. As expected, results for positive emotions were generally similar to results for negative emotions, but weaker. Some effects failed to reach significance for some samples. Full results for positive emotions can be found at <https://osf.io/wsuta/>

We used the techniques suggested by Cranford and colleagues (2006) to calculate reliability for detecting within-person changes, which are roughly equivalent to calculating Cronbach's alpha after subtracting each person's mean over the diary period from the ratings. Shrout (1998) suggests that a reliability of 0.0 to .1 can be seen as "virtually no reliability", .11 to .4 as "slight", 0.41-0.6 as "fair", .61 to .80 as "moderate" and .81 to 1.0 as "substantial". However, Nezlek (2017) notes that in within-person studies standards should be somewhat relaxed, for two reasons: First, we expect participants to complete these measures daily, which means they must be kept shorter. Second, multi-level analysis methods deal better with low reliability than traditional analyses (e.g., multiple regression).

Reliability for self-reported negative moods was .55 for men and .53 for women in Sample 1, .83 for men and .86 for women in Sample 2, and .58 for men and .60 for women in Sample 3. Reliability for empathic inferences was .59 for men and .41 for women in Sample 1, .87 for men and .85 for women in Sample 2, and .59 for men and .63 for women in Sample 3.

Assumed similarity and direct accuracy. Both indices were obtained using multilevel models. Assumed similarity was calculated as the slope of *perceivers'* self reported moods when predicting their empathic inferences. Direct accuracy was calculated as the slope of *targets'* self reported moods when predicting the perceivers' empathic inferences, in the same model (see statistical analysis below for more details).

Response time. In all three samples, the response time for each report was assessed using the Qualtrics platform. Response times were measured to a one millisecond precision and are reported in seconds.

The main response time variable, which we refer to as *empathic inference response time* is the time it took each participant to make empathic inferences, measured as time between entering the questionnaire page inquiring about the partner's moods (which included no other questions) and clicking the button to exit this page and progress to the next one. Values that were 3 standard deviations above or below the mean were considered outliers and removed from analysis, leaving 3205 entries in Sample 1, 3556 entries in Sample 2, and 2769 entries in Sample 3.

Importantly, response time included both positive and negative mood questions, even though we focused on accuracy regarding negative emotions. With our procedures, we cannot be sure whether (faced with mood inference items) participants infer their partners' overall affective states and then answer all items accordingly, or instead make separate inferences for each item. If the former is the case (which we strongly suspect), the effects of inference processes on response times would be reflected only in the first few items; measuring separate response times by item type (positive vs. negative) would render the results meaningless. More importantly, even if the latter is the case, the only adverse consequence of measuring total response time to all items would be adding some random noise.

To control for general processing speed as well as for individual differences in motor operation (e.g., mouse use) speed, we regressed the response time for empathic inferences on the *self-reported mood response time*, and then used the residual scores⁶. The self-reported mood response time is the time between entering the questionnaire page for reporting one's own moods (again, a page which included no other questions) and clicking on the button to exit the page and

⁶ We also conducted our analysis without this adjustment, reaching similar results to those reported below.

progress to the next one. The self-reported mood questionnaire was presented and rated earlier than the empathic inference questionnaire (inquiring about the partner's moods), with other questionnaires unrelated to the current study in between the two.

Analyses

Data were analyzed in a mixed-model analysis, using SAS PROC MIXED, according to recommendations by Bolger and Laurenceau (2013) for analyzing longitudinal dyadic data. According to these recommendations, we treated our data as containing two levels: a daily level (level 1: a specific day for a particular participant) and a couple level (level 2: the couple within which this participant was a partner). Such mixed models take into account the statistical non-independence of partners in each couple. This permits us to simultaneously test the prediction with regards to both partners. Our dependent variable was the perceiver's empathic inference, which was unique for each partner and day. The independent variables were the empathic inference response time (similarly unique for each partner and day), alongside both the perceiver's and the target's self-reported mood (which were unique for each day, and served dual roles - a specific partner's self-reported mood served as a perceiver variable when predicting their own empathic inference and as a target variable when predicting their partner's empathic inference).

Our statistical model was designed as a Truth and Bias model (T&B; West & Kenny, 2011). West and Kenny provide useful terminology and some guidelines for situations in which some kind of judgement or assessment is performed – in our case, the perceiver's attempt to assess the target's mood. The authors suggest predicting the judgement (the empathic inference) using two independent variables. The first is the actual target of inference – in our case, the target's actual mood. This is termed the truth variable. The effect of the truth variable on the

judgement, termed the truth force, is our operationalization of direct accuracy. The second is some alternative source of influence on the judgement – in our case, the perceiver’s own mood. This is termed the bias variable. The effect of the bias variable on the judgement, termed the bias force, is our operationalization of assumed similarity. These associations are moderated by the daily empathic inference response time (M_{ik}) and by a gender difference dummy variable for each participant (G_i) – coded -0.5 for men and 0.5 for women. The full model is as follows:

$$J_{ik} = (b_0 + u_{0i}) + (b_1 + u_{1i})T_{ik} + (b_2 + u_{2i})B_{ik} + (b_3)T_{ik}M_{ik} + (b_4)B_{ik}M_{ik} + (b_5 + u_{5i}) G_i + (b_6 + u_{6i})G_iT_{ik} + (b_7 + u_{7i}) G_iB_{ik} + (b_8) G_iT_{ik}M_{ik} + (b_9) G_iB_{ik}M_{ik} + e_{ik}$$

with J_{ik} , T_{ik} , B_{ik} , M_{ik} being the judgement, truth, bias, and moderator variables respectively for participant i on day k . The model estimates a coefficient for each variable and for interactions between the truth and bias variables on the one hand and the moderator on the other. Each coefficient includes a fixed component (e.g., b_0). All main effect coefficients included a random component varying by participant (e.g., u_{0i}); we could not estimate random components for interaction effects as the models did not converge when such components were included.

According to West and Kenny’s (2011) suggestion, the judgement, bias, and truth variables were all centered on the mean truth value for the target. Response times were centered on each participant’s mean. To calculate effect sizes for specific predictors, we estimated Cohen’s f^2 using the procedures outlined by Selya, Rose, Dierker, Hedeker, and Mermelstein (2012). We evaluated moderation effect sizes by reference to a widely-cited review of moderation effects (Aguinis et al., 2005), which notes that moderation effect sizes tend to be small, and reports a median f^2 across studies of .002, with a 4th quartile at .0053. Thus, we considered moderation effects above .002 as meaningful.

Our operational hypotheses were as follows:

1. The perceiver's empathic inference response time will positively moderate the direct accuracy effect. This moderation will be reflected by a positive interaction between the target's self-report of negative mood and the perceiver's empathic inference response time; $b_3 > 0$.
2. The perceiver's empathic inference response time will not positively moderate the assumed similarity effect. This lack of moderation will be reflected by a non-positive interaction between the perceiver's self-report of negative mood and the perceiver's empathic inference response time; $b_4 \leq 0$.
3. The moderation by the perceiver's empathic inference response time of direct accuracy will differ significantly from (and be more positive than) its moderation of assumed similarity. This difference will be reflected by a positive contrast between the two effects; $b_3 - b_4 > 0$.

Power Analysis

For all samples, data was collected as part of larger studies and sample size was not determined by the considerations of the current study. To ensure the adequacy of these samples for testing our hypothesis, we ran a post-hoc power analysis on the results of Sample 1, using the R simulation package *simr* (Green & MacLeod, 2016). Since our hypothesis was directional, we calculated power for a one-tailed test. Post-hoc power for Sample 1 was .706 (95% CI .677, .734). Using the data from Sample 1 to extrapolate the simulation to higher sample sizes, we determined that 100 couples were needed to achieve adequate power ($\beta = .804$, 95% CI .778, .828), a sample size which was achieved in Sample 2. Unfortunately, due to budget

considerations, the sample size for Sample 3 was slightly smaller; extrapolating the data from Sample 1 to the achieved sample size of 87, power for Sample 3 was .769 (95% CI .742,.795).

Post-hoc power analyses revealed a power of .797 (95% CI .771,.862) for Sample 2 and .764 (95% CI .736,.79) for Sample 3. Aggregating the post-hoc results, power for finding the effect in at least two out of three studies was .851.

Results

Descriptives. Means, standard deviations, ranges and quartile values are provided for self-reported moods, self-reported mood response times, empathic inferences, and empathic inference response times in Table 2. We performed mixed model analyses to compare self-reported mood response time and empathic inference response time; self-reported mood response time was higher in Sample 1 ($b(SD) = 4.053(.358), t(6984) = 11.33, p < .0001, f^2$ effect size = .022), Sample 2 ($b(SD) = 41.155(.354), t(7724) = 5.56, p < .0001, f^2$ effect size = .004), and Sample 3 ($b(SD) = 4.478(.336), t(5925) = 13.34, p < .0001, f^2$ effect size = .032).

Correlations between the three negative mood measures (self, partner and empathic inference), accounting for repeated measures (Bakdash & Marusich, 2017), are provided in Table 3.

--- Tables 2 & 3 about here -----

Response time interactions. To test whether response time moderated assumed similarity and/or direct accuracy, we conducted a mixed-model analysis. Full results are presented in Table 4. The association between perceiver self-reported mood and perceiver empathic inference (i.e., assumed similarity) was not moderated by response time in any sample. The association between target self-reported mood and perceiver empathic inference (i.e., direct

accuracy) was moderated by response time in all samples (Sample 1 f^2 effect size = .006; Sample 2 f^2 effect size = .005; Sample 3 f^2 effect size = .004). In all samples, slower inferences involved more direct accuracy. To test whether moderation of assumed similarity and moderation of direct accuracy differed significantly, we contrasted these effects; moderation of direct accuracy was significantly larger in Sample 1 ($b(SD) = .005 (.003)$, $t(694) = 1.99$, $p = .047$); Sample 2 ($b(SD) = .005(.002)$, $t(810) = 2.78$, $p = .006$); and Sample 3 ($b(SD) = .007(.004)$, $t(596) = 2.04$, $p = .042$). To illustrate the moderation effect, Figure 2 shows standardized r coefficients for assumed similarity (i.e., the association between perceiver self-reported mood and perceiver empathic inference of target mood) and direct accuracy (i.e., the association between target self-reported negative moods and perceiver empathic inference of target mood). Slopes are presented for fast (i.e. <-1 SD residual response time) and slow (i.e. $>+1$ SD residual response time) inferences in each sample.

---- Table 4 and Figure 2 about here -----

Discussion

We set out to test the associations between direct accuracy, assumed similarity, and response times. As expected, direct accuracy was associated (in all three samples) with longer response times, suggesting that it might indeed reflect a slow, deliberate thought process. In contrast, assumed similarity showed no such association. These findings are in line with a dual process account of empathic inferences. Specifically, they support a default interventionist dual process model (Evans & Stanovich, 2013), or an anchoring and adjustment sequence (Tversky & Kahneman, 1974). Accordingly, type-1 processes (in this case, *assumed similarity*) are likely to provide an initial, quick judgement, and type-2 processes (in this case, *direct accuracy*), when implemented, are likely to adjust the initial judgement, but not replace it.

Our findings provide support for the conceptual framework stated above (see Figure 1), tying findings from close relationship, dual process, and neuroscientific research. However, while we focused on one link in the framework for which we could find no prior research, more studies are required to provide evidence for all three links. In our opinion, the ties between the accuracy component model and the other two lines of research warrant specific attention, whereas the tie between dual process and neuroscientific accounts of empathy is already relatively well established. Studies linking assumed similarity and direct accuracy to other characteristics of type-1 or type-2 processes (e.g., the use of working memory), and studies using neuro-imaging or neuro-stimulation methods while recording participants' empathic inferences alongside their own mental states could provide crucial evidence supporting this framework.

Beyond supporting this conceptual framework, our findings have direct implications for the understanding of empathic processes. Specifically, the fact that quicker thinking involves more assumed similarity and less direct accuracy can help predict the effects of different time investments in accuracy within different situations. For example, when people share experiences and spend time together their emotions may grow more similar (Golland et al., 2015; Hatfield et al., 1993). In such situations, quick and immediate empathic inferences which rely mainly on assumed similarity may be very accurate and useful. Consequently, any investment of deliberate and effortful social perception may be a waste of time and energy better spent on simply experiencing the moment; indeed, they may even lead to less accurate inferences. Conversely, when trying to assess the states of mind of strangers or even of close others who have spent some time apart, more effort might be helpful or even required to achieve accurate inferences through direct accuracy.

The current study has some limitations, which can be addressed in future studies. First, reliability of the negative affect measure in Sample 1, and to a slightly lesser extent, Sample 3, was quite low – in all likelihood, because of the brevity of the scales used therein. Given the complete replication with longer (and thus, higher reliability) scales in Sample 2, the pre-registration of Sample 3 hypotheses and analytic code, and the fact that the negative affect construct itself (unlike the construct of empathic accuracy) is not central to our findings, we are not overly concerned about this issue; nevertheless, future studies would benefit from measures with higher reliability.

Second, our study is correlational, raising the possibility that some other variable would lead participants to both be more accurate and to spend more time on their inferences with no real association between the two. We believe such an explanation to be unlikely, as it would mean that participants are consistently spending time working on the inference task without that time investment influencing their performance. Nevertheless, ruling it out completely would require an experimental study⁷.

Third, our measurement of response times was not as precise as it could be; we do not have a record of what device the participant used to complete each questionnaire (e.g., cell phone or computer), a factor which might have influenced completion time. Additionally, the measured response times reflect the time it took to complete all mood items (positive and negative). Future studies could retain information on the device used and separate timing data for the rating of each individual item, although this method has its own caveats (see our method section for a detailed explanation).

⁷ We wish to thank an anonymous reviewer of a previous version of the manuscript for pointing out this limitation.

Finally, all samples involved romantic couples. While we have no reason to suspect the results are not generalizable to other dyads, future studies should test this question with other types of dyads (e.g. parents and children, friends, patients and therapists). Additionally, though we had no hypotheses as to the effects of relational variables such as relationship length or quality on these findings in this initial examination of response time effects on empathic accuracy, such variables could be examined in future studies.

Open practices statement

De-identified data used in the study and analysis code are openly available at <https://osf.io/wsuta/>. See the methods sections for links to full procedure for each study, including information gathered but not used in the current study. Hypotheses, analyses and code for Sample 3 were pre-registered before data collection began at <https://osf.io/cfhsn>.

Compliance with Ethical Standards

All study procedures were approved by the appropriate institutional IRBs. Participants gave their informed consent to enlist in the study. The authors declare no conflicts of interest with respect to the authorship or the publication of this article.

Bibliography

Aguinis, H., Beaty, J. C., Boik, R. J., & Pierce, C. A. (2005). Effect size and power in assessing moderating effects of categorical variables using multiple regression: A 30-year review.

Journal of Applied Psychology, 90(1), 94.

Bakdash, J. Z., & Marusich, L. R. (2017). Repeated measures correlation. *Frontiers in*

Psychology, 8, 456.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology, 5*(4), 323–370.

Bethlehem, R. A., van Honk, J., Auyeung, B., & Baron-Cohen, S. (2013). Oxytocin, brain physiology, and functional connectivity: A review of intranasal oxytocin fMRI studies.

Psychoneuroendocrinology, 38(7), 962–974.

Bohl, V., & van den Bos, W. (2012). Toward an integrative account of social cognition:

Marrying theory of mind and interactionism to study the interplay of Type 1 and Type 2 processes. *Frontiers in Human Neuroscience, 6*.

<https://doi.org/10.3389/fnhum.2012.00274>

Cranford, J. A., Shrout, P. E., Iida, M., Rafaeli, E., Yip, T., & Bolger, N. (2006). A procedure for evaluating sensitivity to within-person change: Can mood measures in diary studies

detect change reliably? *Personality and Social Psychology Bulletin, 32*(7), 917–929.

Evans, J. St. B. T., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition:

Advancing the Debate. *Perspectives on Psychological Science, 8*(3), 223–241.

<https://doi.org/10.1177/1745691612460685>

- Golland, Y., Arzouan, Y., & Levit-Binnun, N. (2015). The mere co-presence: Synchronization of autonomic signals and emotional responses across co-present individuals not engaged in direct interaction. *PloS One*, *10*(5), e0125804.
- Green, P., & MacLeod, C. J. (2016). simr: An R package for power analysis of generalised linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498.
<https://doi.org/10.1111/2041-210X.12504>
- Hall, J. A., & Schmid Mast, M. (2007). Sources of accuracy in the empathic accuracy paradigm. *Emotion*, *7*(2), 438–446. <https://doi.org/10.1037/1528-3542.7.2.438>
- Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1993). Emotional contagion. *Current Directions in Psychological Science*, *2*(3), 96–100.
- Hodges, S. D., Lewis, K. L., & Ickes, W. (2015). The matter of other minds: Empathic accuracy and the factors that influence it. In *APA handbook of personality and social psychology, Volume 3: Interpersonal relations* (pp. 319–348). American Psychological Association.
<https://doi.org/10.1037/14344-012>
- Kenny, D. A., & Acitelli, L. K. (2001). Accuracy and bias in the perception of the partner in a close relationship. *Journal of Personality and Social Psychology*, *80*(3), 439–448.
- Kouros, C. D., & Papp, L. M. (2019). Couples' Perceptions of Each Other's Daily Affect: Empathic Accuracy, Assumed Similarity, and Indirect Accuracy. *Family Process*, *58*(1), 179–196.
- Lorr, M., & McNair, D. M. (1971). The profile of mood states manual. *San Diego, CA: Educational and Industrial Testing Service.*
- Nezlek, J. B. (2017). A practical guide to understanding reliability in studies of within-person variability. *Journal of Research in Personality*, *69*, 149–155.

- Powers, S. R., Rauh, C., Henning, R. A., Buck, R. W., & West, T. V. (2011). The effect of video feedback delay on frustration and emotion communication accuracy. *Computers in Human Behavior, 27*(5), 1651–1657.
- Schmid Mast, M., & Hall, J. A. (2018). The Impact of Interpersonal Accuracy on Behavioral Outcomes. *Current Directions in Psychological Science, 27*(5), 309–314.
<https://doi.org/10.1177/0963721418758437>
- Selya, A. S., Rose, J. S., Dierker, L. C., Hedeker, D., & Mermelstein, R. J. (2012). A practical guide to calculating Cohen's f^2 , a measure of local effect size, from PROC MIXED. *Frontiers in Psychology, 3*, 111.
- Sened, H., Bar-Kalifa, E., Pshedetzky-Shochat, R., Gleason, M., & Rafaeli, E. (2019). Mom-and-pop narcissism: The impact of attention seeking and grandiosity on couples' experience of the transition to parenthood. *Journal of Personality Disorders, 67*–86.
- Sened, H., Lavidor, M., Lazarus, G., Bar-Kalifa, E., Rafaeli, E., & Ickes, W. (2017). Empathic accuracy and relationship satisfaction: A meta-analytic review. *Journal of Family Psychology, 31*(6), 742–752. <https://doi.org/10.1037/fam0000320>
- Sened, H., Yovel, I., Bar-Kalifa, E., Gadassi, R., & Rafaeli, E. (2017). Now you have my attention: Empathic accuracy pathways in couples and the role of conflict. *Emotion, 17*(1), 155–168.
- Shrout, P. E. (1998). Measurement reliability and agreement in psychiatry. *Statistical Methods in Medical Research, 7*(3), 301–317.
- 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), 629–641. <https://doi.org/10.1037/0022-3514.69.4.629>

- Spunt, R. P., & Lieberman, M. D. (2013). The Busy Social Brain: Evidence for Automaticity and Control in the Neural Systems Supporting Social Cognition and Action Understanding. *Psychological Science, 24*(1), 80–86. <https://doi.org/10.1177/0956797612450884>
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124–1131.
- West, T. V., & Kenny, D. A. (2011). The truth and bias model of judgment. *Psychological Review, 118*(2), 357.
- Wicherts, J. M., Veldkamp, C. L., Augusteijn, H. E., Bakker, M., Van Aert, R., & Van Assen, M. A. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid p-hacking. *Frontiers in Psychology, 7*, 1832.
- 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*(1–2), 183–246.
- Zaki, J., & Ochsner, K. (2011). Reintegrating the study of accuracy into social cognition research. *Psychological Inquiry, 22*(3), 159–182.

Table 1 – Demographics

	Sample 1	Sample 2	Sample 3
Age M(SD)	30.3(9.9)	29.4(4.3)	30(9.3)
Relationship Length M(SD)	7(8.9)	4.9(2.9)	7.3(8.4)
Joint Monthly Income			
0-1250 USD equivalent	37.5%	9.5%	28.2%
1250-2500 USD equivalent	20%	23%	26.4%
2500-3750 USD equivalent	19.4%	29.5%	16.1%
3750-5000 USD equivalent	7.5%	20%	13.8%
5000+ USD equivalent	15.6%	18%	15.5%
Education			
Did not complete high school	0%	0%	0%
High school Graduate	35.6%	16%	31.6%
Bachelor's Degree	54.4%	52.5%	56.9%
Master's Degree or Above	10%	31.5%	11.5%

Table 2 – Descriptive statistics

	Mean(SD)	Minimum	1 st quartile	median	3 rd quartile	Maximum
Sample 1						
Self Reported Negative Mood	10.671(13.309)	0	.333	6	16	100
Empathic Inference of Negative Mood	11.988(13.531)	0	1.333	7.667	18	100
Self Reported Mood Response Time	28.351(14.356)	4.484	17.818	24.584	34.847	80.761
Empathic Inference Response Time	24.47(11.581)	4.568	15.848	21.703	30.189	64.421
Sample 2						
Self Reported Negative Mood	.344(.514)	0	0	.111	.444	3.889
Empathic Inference of Negative Mood	.424(.594)	0	0	.222	.611	3.778
Self Reported Mood Response Time	41.927(20.659)	0	27.889	36.712	50.099	133.672
Empathic Inference Response Time	39.379(17.617)	3.284	26.948	35.167	47.602	104.549
Sample 3						
Self Reported Negative Mood	14.354(14.814)	0	2.333	10	22	98.667
Empathic Inference of Negative Mood	15.667(15.754)	0	3	11.333	23.333	92.667
Self Reported Mood Response Time	26.676(13.401)	7.922	16.963	22.919	32.844	76.747
Empathic Inference Response Time	22.495(9.585)	6.547	15.504	20.106	27.319	54.058

Table 3 – Negative mood correlations accounting for repeated measures

	Sample 1 r	Sample 2 r	Sample 3 r
Perceiver self-reported mood <-> Target self-reported mood	.136	.335	.293
Perceiver empathic inference <-> Target self-reported mood	.281	.444	.434
Perceiver empathic inference <-> Perceiver self-reported mood	.309	.59	.461

Table 4. Effects of Perceiver and Target Negative Affect and Residual Response Time on Perceiver's Inference of Target Negative Affect

	Mean(SE)	t(df)	p	95% CI	Gender Differences	
					t(df)	p
Sample 1						
Intercept	1.7(.374)	4.55(74.4)	< .001***	(.955,2.446)	-3.61(75)	.001***
Perceiver NA	.274(.025)	11.13(55.2)	< .001***	(.225,.324)	-2.07(79.3)	.042*
Target NA	.268(.029)	9.35(67.4)	< .001***	(.211,.326)	-2.22(72.4)	.029*
Residual RT	.119(.027)	4.42(65.5)	< .001***	(.065,.172)	.07(64.1)	.947
Perceiver NA*Residual RT	0(.002)	.09(176)	.930	(-.003,.003)	.04(270)	.968
Target NA*Residual RT	.006(.002)	2.52(2360)	.012*	(.001,.01)	-1.84(2361)	.066
Sample 2						
Intercept	.092(.013)	6.98(84.1)	< .001***	(.066,.118)	-3.41(87.6)	.001**
Perceiver NA	.52(.028)	18.71(83.6)	< .001***	(.465,.576)	-3.43(97.7)	.001***
Target NA	.324(.024)	13.45(77.5)	< .001***	(.276,.371)	-.24(95.2)	.814
Residual RT	.003(0)	7.26(87.6)	< .001***	(.002,.004)	-.49(95.4)	.628
Perceiver NA*Residual RT	-.001(.001)	-.65(359)	.517	(-.003,.001)	-1.23(326)	.220
Target NA*Residual RT	.004(.001)	3.34(2084)	.001***	(.002,.007)	1.28(2063)	.200
Sample 3						
Intercept	1.455(.384)	3.79(81.4)	< .001***	(.692,2.218)	-3.91(77.1)	< .001***
Perceiver NA	.417(.029)	14.32(86.5)	< .001***	(.359,.475)	-2.48(69.5)	.016*
Target NA	.309(.029)	1.63(62.8)	< .001***	(.251,.367)	-.9(58.1)	.373
Residual RT	.104(.032)	3.22(58.6)	.002**	(.039,.169)	.89(62)	.375
Perceiver NA*Residual RT	0(.002)	.19(212)	.847	(-.004,.004)	-.71(199)	.480
Target NA*Residual RT	.008(.003)	2.92(1723)	.004**	(.003,.013)	-1.37(1716)	.172

*p < .05 **p < .01 ***p < .001 NA – Negative Affect RT – Response time

Positive gender difference values indicate more positive and less negative associations for women.

Figure 1. Conceptual framework of empathic accuracy

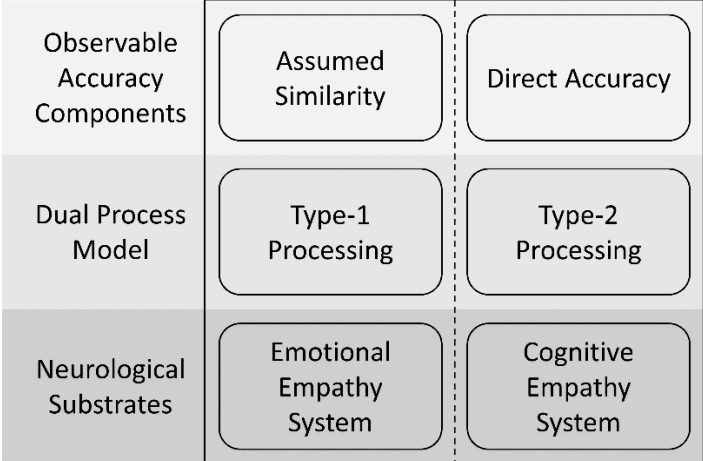


Figure 2. Direct Accuracy and Assumed Similarity Standardized coefficients in Fast and Slow responses

