Supplemental Material

Table of Contents

Missing Data and Descriptive Statistics on Self-Report Measures

Missing data on PT were imputed using the expectation maximization (EM) algorithm for 1 participant in the Biological Motion analysis, 1 participant in the ER-40 analysis, 0 in the RMET analysis, 0 in the EA analysis, and 0 in the affect sharing analysis. Missing data on ECp and ECn were imputed using EM for 4 in the Biological Motion analysis, 4 participants in the ER-40 analysis, 1 in the RMET analysis, 2 in the EA analysis, and 1 in the affect sharing analysis. Missing data on TAS-20 were imputed using EM for 14 in the Biological Motion analysis, 4 participants in the ER-40 analysis, 11 in the RMET analysis, 4 in the EA analysis, and 1 in the affect sharing analysis. For the measure of SES, data were missing for 10 participants in the Biological Motion analysis, 4 participants in the ER-40 analysis, 6 in the RMET analysis, 7 in the EA analysis, and 3 in the affect sharing analysis.

Means, standard deviations and internal consistency estimates for each sample can be found in Supplemental Table 1 below.

Supplemental Table 1. Cronbach's αs, Mean Values, and Standard Deviations of all Self-Report Measures for Each Outcome Variable

	Bio. Motion (N = 1,756)			ER-40 (N = 384)			RMET (N = 1,473)			EA (N = 849)			Affect Sharing (N = 541)		
	M	SD	α	M	SD	α	M	SD	α	M	SD	α	M	SD	α
PT	2.58	.66	.83	2.62	.61	.82	2.57	.67	.83	2.58	.68	.86	2.60	.67	.80
ECp	3.21	.58	.86	3.30	.54	.85	3.20	.59	.86	3.23	.55	.83	3.14	.62	.86
ECn	2.71	.54	.81	2.72	.53	.80	2.71	.54	.81	2.73	.53	.81	2.71	.59	.84
Alexithymia	2.42	.60	.87	2.40	.60	.88	2.42	.60	.87	2.38	.64	.88	2.38	.69	.90

Factor Structure of Emotional Contagion Scale (ECS)

Statistical Analyses

We used exploratory factor analysis (EFA) to examine the factor structure of the Emotional Contagion Scale (ECS). First, to identify the maximum number of factors to extract, we performed Velicer's minimum average partial (MAP) test, as well as parallel analysis (PA) using the 95th percentile criteria. Results were obtained from both parallel analysis using principal component analysis (PA-PCA) and parallel analysis using principal axis factoring (PA-PAF).

Results

Velicer’s MAP test suggested that no more than two factors should be extracted (the smallest average squared partial correlation was of .022). These results were consistent with those from the PA-PCA, which suggested the retention of two components: the first three observed eigenvalues from the real data were 5.34, 1.94, and 1.05, and the corresponding first three 95th percentile eigenvalues from the random data were 1.22, 1.17, and 1.13. The PA-PAF, on the other hand, resulted in real-data eigenvalues which were consistently greater than the random-data eigenvalues, suggesting that the maximum number of factors (i.e., 15) could be retained.

As discussed in Murphy et al. (2018), research using self-report measures of emotional contagion often conceptualizes the construct of emotional contagion as unidimensional, representing an overall tendency towards emotional contagion; however, a two-factor model of emotional contagion may better capture important differences between contagion for positive and negative emotions with regards to interpersonal outcomes. Thus, we investigated one- and two-factor models using principal axis factoring with promax rotation. Factor loadings for both models can be found in Table 2 and 3, respectively. Consistent with theory, results from the two-factor model revealed that all items representing emotional contagion for negative emotions showed moderately high loadings on the first factor (mean factor loading .59, range .47 to .73), and cross-loadings below .25 with the second factor. Similarly, all items representing emotional contagion for positive emotions showed high loadings on the second factor (mean factor loading .68, range .61 to .83), and low cross-loading. The correlation between these two factors was .54, consistent with theory suggesting that these are related but distinct subcomponents of an overarching construct.

Thus, results from the EFA align well with those from the MAP test and PA-PCA, suggesting that a two-factor model of the ECS, which indexes emotional contagion for positive emotions and for negative emotions, may be warranted.

Supplemental Table 2. Factor Loadings for Exploratory Factor Analysis with One Factor

	Factor
	1
Item 1	.542
Item 2	.603
Item 3	.671
Item 4	.621
Item 5	.317
Item 6	.583
Item 7	.421
Item 8	.578
Item 9	.650
Item 10	.514
Item 11	.630
Item 12	.629
Item 13	.557
Item 14	.529
Item 15	.407

Supplemental Table 3. Factor Loadings for Exploratory Factor Analysis with Two Factors

	Factor	
	1	**2**
Item 10	.733	-.081
Item 15	.675	-.073
Item 13	.661	.010
Item 14	.618	.045
Item 5	.595	-.199
Item 1	.539	.156
Item 8	.515	.182
Item 7	.474	.049
Item 4	.474	.235
Item 12	-.162	.833
Item 6	-.055	.721
Item 9	.053	.671
Item 11	.024	.653
Item 2	.015	.622
Item 3	.076	.606

Demographic Information by Outcome Measure

Biological Motion Task Analyses

Overall, participants were predominantly female (67.90%), were on average 24.75 years old (SD = 10.94; age range = 18-77 years), and self-identified as White (71.90%), Black or African American (5.60%), Asian (16.50%), Native American or Alaska Native (0.10%) and Other (5.10%). On average, participants had completed 14.06 years of education (SD = 2.25 years).

ER-40 Task Analyses

Participants were predominantly female (67.70%), were on average 19.70 years old (SD = 1.79; age range = 18-31 years), and self-identified as White (75.00%), Black or African American (3.40%), Asian (16.70%), Native American or Alaska Native (1.00%) and Other (3.60%). On average, participants had completed 13.55 years of education (SD = 1.29 years).

RMET Analyses

Participants were predominantly female (69.20%), were on average 25.75 years old (SD = 11.66; age range = 18-77 years), and self-identified as White (71.20%), Black or African American (6.30%), Asian (16.20%), Native American or Alaska Native (0.50%) and Other (5.40%). On average, participants had completed 14.20 years of education (SD = 2.36 years).

EA Analyses

Participants were predominantly female (66.10%), were on average 25.32 years old (SD = 10.80; age range = 18-80 years), and self-identified as White (77.60%), Black or African American (7.50%), Asian (10.50%), Native American or Alaska Native (0.60%) and Other (3.70%). On average, participants had completed 14.14 years of education (SD = 2.23 years).

Affect Sharing Analyses

Participants were predominantly female (54.70%), were on average 30.12 years old (SD = 11.90; age range = 18-70 years), and self-identified as White (78.70%), Black or African American (9.00%), Asian (8.80%), Native American or Alaska Native (1.50%) and Other (1.90%). On average, participants had completed 15.07 years of education (SD = 2.88 years).

Biological Motion Task Exploratory Analyses

Gender-Moderation

Statistical Analysis. To examine whether participant gender moderated any of the associations between self-reported empathic tendencies and performance on the Biological Motion task, we conducted a hierarchical multiple regression. The model included our main variables of interest (i.e., PT, ECp, and ECn), two dummy variables coding for the three sites (i.e., University 1, University 2, and Mturk, with the latter as the reference group), demographic covariates (i.e., age, gender, and SES), alexithymia, and three gender-moderation interaction terms: PT x Female Gender, ECp x Female Gender, and ECn × Female Gender. FDR correction was applied to the four tests of social cognitive ability (i.e., Biological Motion, ER-40, RMET and EA).

Results. Our results revealed a significant interaction between ECp and Female Gender, as well as between ECn and Female Gender, however, these interactions were no longer significant after correcting for multiple tests, $\beta = -.126$, 95% CI = [-.236, -.017], $t(1743) = 2.267$, $p = .023$ (FDR corrected, $p = .062$), $sr^2 = .003$, and $\beta = .128$, 95% CI = [.011, .245], $t(1743) = 2.150$, $p = .032$ (FDR corrected, $p = .102$), $sr^2 = .003$, respectively. The interaction between PT and Female Gender was not significant, $\beta = -.087$, 95% CI = [-.188, .014], $t(1743) = 1.687$, $p = .092$ (FDR corrected, $p = .368$), $sr^2 = .002$.

Valence-Moderation

Statistical Analysis. To examine whether stimuli valence (i.e., positive, negative or neutral) moderated any of the associations between self-reported empathic tendencies and performance on the Biological Motion task, we used multilevel modeling. Because the Biological Motion data were negatively skewed (-1.21), we conducted a generalized linear mixed model (GLMM); we first reversed the data (by multiplying the participants' accuracy score for each stimuli by -1), and then added +2 to these scores. We used a GLMM with γ distribution and a log linking function, with participants' accuracy for each stimuli as the outcome variable. Level 1 (within-person) predictors included: two dummy variables coding for stimuli valence (i.e., with negative valence as the reference group). Level 2 (between-person) predictors included: our main variables of interest (i.e., PT, ECp, and ECn), two dummy variables coding for the three sites (i.e., University 1, University 2, and Mturk, with the latter as the reference group), demographic covariates (i.e., age, gender, and SES), alexithymia, and six two-way interactions coding for valence-specific effects: PT × Positive Valence, PT × Neutral Valence, ECp × Positive Valence, ECp × Neutral Valence, ECn × Positive Valence, ECn × Neutral Valence. FDR correction was applied to the four tests of social cognitive ability (i.e., Biological Motion, ER-40, RMET and EA); specifically, we corrected the interaction between our main effects of interest (i.e., PT, ECp, and ECn) and the dummy variable coding for positive vs. negative valence, across the four tests, given that this interaction was common to all tests (whereas stimuli of neutral valence were not present in the EA task). For the interaction between our main effects of interest and the dummy variable coding for neutral vs. negative valence, FDR correction was applied to the three tests which included neutral stimuli (i.e., Biological Motion, ER-40, and RMET).

Results. Of note, all *b* values presented hereon for these analyses have been reversed for presentation purposes, following the initial reverse coding of the Biological Motion data for GLMM analysis. There was a significant interaction between PT and Positive Valence, however this interaction was no longer significant after correcting for multiple tests, $b = -.009$, 95% CI = [-.017, -.002], $t(41406) = 2.497$, $p = .013$ (FDR corrected, $p = .052$). Similarly, there was no significant interaction between PT and Neutral Valence, $b = -.004$, 95% CI = [-.012, .003], $t(41406) = 1.189$, $p = .234$ (FDR corrected, $p = .234$).

PT × EC interaction

Statistical Analysis. To examine whether PT interacted with EC (for both positive and negative emotions) in predicting performance on the Biological Motion task, we conducted a hierarchical multiple regression. The model included our main variables of interest (i.e., PT, ECp, and ECn), two dummy variables coding for the three sites (i.e., University 1, University 2, and Mturk, with the latter as the reference group), demographic covariates (i.e., age, gender, and SES), alexithymia, and two interaction terms: PT × ECp, and PT × ECn. FDR correction was applied to the four tests of social cognitive ability (i.e., Biological Motion, ER-40, RMET and EA).

Results. Our results revealed no significant interaction between PT and ECp, nor between PT and ECn, $\beta = -.042$, 95% CI = [-.089, .006], $t(1744) = 1.731$, $p = .084$ (FDR corrected, $p = .168$), $sr^2 = .002$, and $\beta = .015$, 95% CI = [-.032, .061], $t(1744) = 0.620$, $p = .535$ (FDR corrected, $p = .713$), $sr^2 = .000$, respectively.

Emotion Recognition Task (ER-40 Task) Exploratory Analyses

Gender-Moderation

Statistical Analysis. To examine whether participant gender moderated any of the associations between self-reported empathic tendencies and performance on the ER-40 task, we conducted a hierarchical multiple regression. The model included our main variables of interest (i.e., PT, ECp, and ECn), demographic covariates (i.e., age, gender, and SES), alexithymia, and three gender-moderation interaction terms: PT × Female Gender, ECp × Female Gender, and ECn × Female Gender. FDR correction was applied to the four tests of social cognitive ability (i.e., Biological Motion, ER-40, RMET and EA).

Results. Our results revealed no significant interaction between PT and Female Gender, $\beta = .079$, 95% CI = [-.144, .301], $t(373) = 0.694$, $p = .488$ (FDR corrected, $p = .488$), $sr^2 = .001$, between ECp and Female Gender, $\beta = .083$, 95% CI = [-.156, .321], $t(373) = 0.682$, $p = .496$ (FDR corrected, $p = .496$), $sr^2 = .001$, nor between ECn and Female Gender, $\beta = .016$, 95% CI = [-.251, .283], $t(373) = 0.117$, $p = .907$ (FDR corrected, $p = .907$), $sr^2 = .000$.

Valence-Moderation

Statistical Analysis. To examine whether stimuli valence (i.e., positive, negative or neutral) moderated any of the associations between self-reported empathic tendencies and performance on the ER-40 task, we used multilevel modeling. For each stimuli in the ER-40 task, participants received either a score of 0 (i.e., for incorrect responses) or a score of 1 (i.e., for correct responses); therefore, we conducted a binary logistic regression using GLMM (i.e., a GLMM with binomial distribution and a logit linking function), with participants' binary accuracy score as the outcome variable. Level 1 (within-person) predictors included: two dummy variables coding for stimuli valence (i.e., with negative valence as the reference group). Level 2 (between-person) predictors included: our main variables of interest (i.e., PT, ECp, and ECn), demographic covariates (i.e., age, gender, and SES), alexithymia, and six two-way interactions coding for valence-specific effects: PT × Positive Valence, PT × Neutral Valence, ECp × Positive Valence, ECp × Neutral Valence, ECn × Positive Valence, ECn × Neutral Valence. FDR correction was applied to the four tests of social cognitive ability (i.e., Biological Motion, ER-40, RMET and EA); specifically, we corrected the interaction between our main effects of interest (i.e., PT, ECp, and ECn) and the dummy variable coding for positive vs. negative valence, across the four tests, given that this interaction was common to all tests (whereas stimuli of neutral valence were not present in the EA task). For the interaction between our main effects of interest and the dummy variable coding for neutral vs. negative valence, FDR correction was applied to the three tests which included neutral stimuli (i.e., Biological Motion, ER-40, and RMET).

Results. There was a significant interaction between PT and Neutral Valence, and this interaction remained significant after correcting for multiple tests, $b = -.185$, $t(15244) = 3.522$, OR = 0.831, 95% CI of OR = [0.750, 0.921], $p < .001$ (FDR corrected, $p < .001$). For stimuli of negative valence, PT was not significantly associated with performance on the ER-40 task, $b = .046$, $t(15244) = 1.333$, OR = 1.047, 95% CI of OR = [0.979, 1.121], $p = .182$. For neutral stimuli, however, PT was associated with lower odds of a correct answer, $b = -.139$, $t(15244) = 2.863$, OR = 0.870, 95% CI of OR = [0.792, 0.957], $p = .004$.

PT × EC interaction

Statistical Analysis. To examine whether PT interacted with EC (for both positive and negative emotions) in predicting performance on the ER-40 task, we conducted a hierarchical multiple regression. The model included our main variables of interest (i.e., PT, ECp, and ECn), demographic covariates (i.e., age, gender, and SES), alexithymia, and two interaction terms: PT × ECp, and PT × ECn. FDR correction was applied to the four tests of social cognitive ability (i.e., Biological Motion, ER-40, RMET and EA).

Results. Our results revealed no significant interaction between PT and ECp in predicting performance on the ER-40 task, $\beta = .030$, 95% CI = [-.089, .150], $t(374) = 0.498$, $p = .619$ (FDR corrected, $p = .619$), $sr^2 = .001$, nor between PT and ECn, $\beta = -.020$, 95% CI = [-.131, .091], $t(374) = 0.347$, $p = .729$ (FDR corrected, $p = .729$), $sr^2 = .000$.

Reading the Mind in the Eyes Test (RMET) Exploratory Analyses

Gender-Moderation

Statistical Analysis. To examine whether participant gender moderated any of the associations between self-reported empathic tendencies and performance on the RMET, we conducted a hierarchical multiple regression. The model included our main variables of interest (i.e., PT, ECp, and ECn), two dummy variables coding for the three sites (i.e., University 1, University 2, and Mturk, with the latter as the reference group), demographic covariates (i.e., age, gender, and SES), alexithymia, and three gender-moderation interaction terms: PT × Female Gender, ECp × Female Gender, and ECn × Female Gender. FDR correction was applied to the four tests of social cognitive ability (i.e., Biological Motion, ER-40, RMET and EA).

Results. Our results revealed a significant interaction between ECp and Female Gender, however this interaction was no longer significant after correcting for multiple tests, $\beta = -.131$, 95% CI = [-.249, -.012], $t(1460) = 2.163$, $p = .031$ (FDR corrected, $p = .062$), $sr^2 = .003$.

Valence-Moderation

Statistical Analysis. To examine whether stimuli valence (i.e., positive, negative or neutral) moderated any of the associations between self-reported empathic tendencies and performance on the RMET, we used multilevel modeling. For each RMET stimuli, participants received either a score of 0 (i.e., for incorrect responses) or a score of 1 (i.e., for correct responses); therefore, we conducted a binary logistic regression using GLMM (i.e., a GLMM with binomial distribution and a logit linking function), with participants' binary accuracy score as the outcome variable. Level 1 (within-person) predictors included: two dummy variables coding for stimuli valence (i.e., with negative valence as the reference group). Level 2 (between-person) predictors included: our main variables of interest (i.e., PT, ECp, and ECn), two dummy variables coding for the three sites (i.e., University 1, University 2, and Mturk, with the latter as the reference group), demographic covariates (i.e., age, gender, and SES), alexithymia, and six two-way interactions coding for valence-specific effects: PT × Positive Valence, PT × Neutral Valence, ECp × Positive Valence, ECp × Neutral Valence, ECn × Positive Valence, ECn × Neutral Valence. FDR correction was applied to the four tests of social cognitive ability (i.e., Biological Motion, ER-40, RMET and EA); specifically, we corrected the interaction between our main effects of interest (i.e., PT, ECp, and ECn) and the dummy variable coding for positive vs. negative valence, across the four tests, given that this interaction was common to all tests (whereas stimuli of neutral valence were not present in the EA task). For the interaction between our main effects of interest and the dummy variable coding for neutral vs. negative valence, FDR correction was applied to the three tests which included neutral stimuli (i.e., Biological Motion, ER-40, and RMET).

Results. There was a significant interaction between ECn and Positive Valence, $b = -.076$, $t(54430) = 2.409$, OR = 0.926, 95% CI of OR = [0.871, 0.986], $p = .016$, however this interaction was no longer significant after correcting for multiple tests (FDR corrected, $p = .064$).

PT × EC interaction

Statistical Analysis. To examine whether PT interacted with EC (for both positive and negative emotions) in predicting performance on the RMET, we conducted a hierarchical multiple regression. The model included our main variables of interest (i.e., PT, ECp, and ECn), two dummy variables coding for the three sites (i.e., University 1, University 2, and Mturk, with the latter as the reference group), demographic covariates (i.e., age, gender, and SES), alexithymia, and two interaction terms: PT × ECp, and PT × ECn. FDR correction was applied to the four tests of social cognitive ability (i.e., Biological Motion, ER-40, RMET and EA).

Results. Our results revealed a significant interaction between PT and ECp, $\beta = -.105$, 95% CI = [-.154, -.056], $t(1461) = 4.168$, $p < .001$, $sr^2 = .011$, and this interaction remained significant following multiple test correction (FDR corrected, $p < .001$). For individuals with low levels of emotional contagion for positive emotions (i.e., with ECp scores that were one standard deviation below the mean), PT was positively associated with RMET performance, $\beta = .173$, 95% CI = [.101, .245], $t(1461) = 4.717$, $p < .001$, $sr^2 = .013$. Similarly, for individuals with average levels of ECp, PT was positively associated with RMET performance, $\beta = .068$, 95% CI = [.016, .121], $t(1461) = 2.558$, $p = .011$, $sr^2 = .004$. However, for individuals with high levels of ECp (i.e., with ECp scores that were one standard deviation above the mean), PT was not significantly associated with RMET performance, $\beta = -.037$, 95% CI = [-.109, .035], $t(1461) = 0.997$, $p = .319$, $sr^2 = .001$. There was no significant interaction between PT and ECn, $\beta = .034$, 95% CI = [-.016, .083], $t(1461) = 1.341$, $p = .180$ (FDR corrected, $p = .676$), $sr^2 = .001$.

Empathic Accuracy (EA) Exploratory Analyses

Gender-Moderation

Statistical Analysis. To examine whether participant gender moderated any of the associations between self-reported empathic tendencies and EA performance, we conducted a similar GLMM to our primary analyses with EA. As with our primary analyses, EA scores were first reversed (by multiplying the participants' accuracy score for each stimuli by -1), and +2 was added to these scores in order for all scores to be greater than 1. We used a GLMM with γ distribution and a log linking function, with participants' accuracy for each stimuli as the outcome variable. Video valence (negative=0, positive=1) and target expressivity were entered as Level 1 (within-person) predictors. The Level 2 (between-person) predictors included two dummy variables indicating which study participants completed (Study 3 was set as the reference group), PT, EC for positive emotions, EC for negative emotions, demographic covariates (age, gender, and SES), and alexithymia. Due to the study-specific effects observed in our primary analyses, we also tested for study-specific effects in these exploratory analyses. Thus, we included the six three-way interactions of the three variables of interest, Female Gender, and the two study variables (PT × Study 1 × Female Gender, PT × Study 2 × Female Gender, ECp × Study 1 × Female Gender, ECp × Study 2 × Female Gender, ECn × Study 1 × Female Gender, ECn × Study 2 × Female Gender), along with all lower-level interaction terms. A random intercept was included for each participant. FDR correction was applied to the four tests of social cognitive ability (i.e., Biological Motion, ER-40, RMET and EA); for this analysis, the correction was applied to the results at the "average study." A separate correction was applied to the three tests of study-specific effects.

Results. Of note, all *b* values presented hereon for these analyses have been reversed for presentation purposes, following the initial reverse coding of the EA data for GLMM analysis. At the average study, the gender interactions with PT, ECp, and ECn were not significant, $b = -.007$, 95% CI = [-.023, .008], $t(5852) = 0.946$, $p = .344$ (FDR corrected, $p = .459$), $b = -.012$, 95% CI = [-.027, .004], $t(5852) = 1.467$, $p = .142$ (FDR corrected, $p = .189$), and $b = .016$, 95% CI = [.000, .032], $t(5852) = 1.954$, $p = .051$ (FDR corrected, $p = .102$), respectively.

There was a significant three-way interaction between PT, Study 1-vs-3 and Female Gender, $b = .035$, 95% CI = [.001, .069], t(5852) = 2.037, p = .042. In Study 1, the two-way interaction between PT and Female Gender was not significant, suggesting that the association between PT and EA did not differ by gender in Study 1, $b = .009$, 95% CI = [.017, .034], $t(5852) = 0.656$, $p = .512$ (FDR corrected, $p = .512$). In Study 3, there was a significant interaction between PT and Female Gender, but this interaction was no longer significant after multiple test correction, $b = -.027$, 95% CI = [-.049, -.005], $t(5852) = 2.359$, $p = .018$ (FDR corrected, $p = .054$).

For ECn, two significant three-way interactions were observed: between ECn, Study 1-vs-3 and Female Gender, $b = -.068$, 95% CI = [-.102, -.035], t(5852) = 3.981, $p < .001$, and between ECn, Study 2-vs-3 and Female Gender, $b = -.089$, 95% CI = [-.131, -.047], $t(5852) = 4.146$, $p < .001$, which suggests that the gender-moderated effect of ECn differed significantly between Study 1 and Study 3, and between Study 2 and Study 3. In both Study 1 and Study 2, the interaction between ECn and Female Gender was not significant, $b = -.001$, 95% CI = [-.028, .025], $t(5852) = 0.112$, $p = .910$ (FDR corrected, $p = .910$), and $b = -.022$, 95% CI = [-.059, .014], $t(5852) = 1.205$, $p = .228$ (FDR corrected, $p = .342$), respectively. This suggests that gender did not moderate the association between ECn and EA in both Study 1 and Study 2. By contrast, in Study 3, the interaction between ECn and Female Gender was significant, $b = .067$, 95% CI = [.045, .088], $t(5852) = 6.175$, $p < .001$, and this interaction was maintained after multiple test correction (FDR corrected, $p < .001$). For men, ECn was associated with worse EA performance, $b = -.044$, 95% CI = [-.059, -.030], $t(5852) = 5.872$, $p < .001$, whereas for women, ECn was related to better EA performance, $b = .022$, 95% CI = [.007, .038], $t(5852) = 2.833$, $p = .005$.

Valence-Moderation

Statistical Analysis. To examine whether stimuli valence (i.e., positive and negative) moderated any of the associations between self-reported empathic tendencies and EA performance, we used a similar GLMM to the model used in the gender moderation analysis (see previous analysis plan). Video valence (negative=0, positive=1) and target expressivity were entered as Level 1 (within-person) predictors. The Level 2 (between-person) predictors included two dummy variables indicating which study participants completed (Study 3 was set as the reference group), PT, ECp, ECn, demographic covariates (age, gender, and SES), and alexithymia. Similarly, we also tested for study-specific effects in these exploratory analyses. Therefore, we included the six three-way interactions of the three variables of interest, stimuli valence, and the two study variables (PT × Study 1 × Positive Valence, PT × Study 2 × Positive Valence, ECp × Study 1 × Positive Valence, ECp × Study 2 × Positive Valence, ECn × Study 1 × Positive Valence, ECn × Study 2 × Positive Valence), along with all lower-level interaction terms. A random intercept was included for each participant. FDR correction was applied to the four tests of social cognitive ability (i.e., Biological Motion, ER-40, RMET and EA). Specifically, we corrected the interaction between our main effect of interest (i.e., PT, ECp, and ECn) and the dummy variable coding for positive vs. negative valence; for this analysis, the correction was applied to the results at the "average study." A separate correction was applied to the three tests of study-specific effects.

Results. Of note, all *b* values presented hereon for these analyses have been reversed for presentation purposes, following the initial reverse coding of the EA data for GLMM analysis. At the average study, the valence interactions with PT, ECp, and ECn were not significant, $b = .003$, 95% CI = [-.006, .012], $t(5852) = 0.665$, $p = .506$ (FDR corrected, $p = .964$), $b = .002$, 95% CI = [-.009, .011], $t(5852) = 0.226$, $p = .821$ (FDR corrected, $p = .978$), and $b = .005$, 95% CI = [-.005, .015], $t(5852) = 0.949$, $p = .342$ (FDR corrected, $p = .456$), respectively.

There was a significant three-way interaction between ECp, Study 1-vs-2 and Positive Valence, as well as between ECp, Study 3-vs-2 and Positive Valence, $b = .035$, 95% CI = [.008, .062], $t(5852) = 2.587$, $p = .010$, and $b = .042$, 95% CI = [.014, .069], $t(5852) = 2.980$, $p = .003$, respectively. In Study 2, there was a significant two-way interaction between ECp and Positive Valence, $b = -.028$, 95% CI = [-.050, -.007], $t(5852) = 2.564$, $p = .010$, and this interaction remained after controlling for multiple tests (FDR corrected, $p = .03$). For stimuli of positive valence, ECp was associated with worse EA, whereas for stimuli of negative valence, ECp was not significantly associated with EA, *b* = -.025, 95% CI = [-.046, -.003], *t*(5852) = 2.269, *p* = .023, and *b* = .004, 95% CI = [-.014, .022], *t*(5852) = 0.409 *p* = .683, respectively. In Study 1 and Study 3, on the other hand, the interaction between ECp and Positive Valence was not significant, *b* = .007, 95% CI = [-.009, .022], *t*(5852) = 0.856, *p* = .392 (FDR corrected, *p* = .392), and *b* = .013, 95% CI = [-.003, .030], *t*(5852) = 1.559, *p* = .119 (FDR corrected, *p* = .179), respectively. This suggests that the association between ECp and EA did not differ by valence in either Study 1 or Study 3.

PT × EC interaction

Statistical Analysis. To examine whether PT interacted with EC (for both positive and negative emotions) in predicting EA performance, we used a similar GLMM as described previously for EA analyses. Video valence (negative=0, positive=1) and target expressivity were entered as Level 1 (within-person) predictors. The Level 2 (between-person) predictors included two dummy variables indicating which study participants completed (Study 3 was set as the reference group), PT, ECp, ECn, demographic covariates (age, gender, and SES), and alexithymia. Additionally, we included four three-way interaction terms (i.e., PT × ECp × Study 1, PT × ECp × Study 2, PT × ECn × Study 1, and PT × ECn × Study 3), along with all lower-level interaction terms. A random intercept was included for each participant. FDR correction was applied to the four tests of social cognitive ability (i.e., Biological Motion, ER-40, RMET and EA).

Results. Of note, all *b* values presented hereon for these analyses have been reversed for presentation purposes, following the initial reverse coding of the EA data for GLMM analysis. Our results revealed no study specific effects (i.e., all three-way interactions were not significant), as well as no significant interaction between PT and ECp, nor between PT and ECn, in predicting EA performance (at the average study), *b* = -.002, 95% CI = [-.009, .005], *t*(5857) = 0.616, *p* = .538 (FDR corrected, *p* = .619), and *b* = .003, 95% CI = [-.003, .010], *t*(5857) = 0.957, *p* = .338 (FDR corrected, *p* = .676), respectively.

Affect Sharing Task Exploratory Analyses

Gender-Moderation

Statistical Analysis. To examine whether PT interacted with EC (for both positive and negative emotions) in predicting affect sharing, we used a similar GLMM as described previously for affect sharing analyses. Video valence (negative=0, positive=1) and target expressivity were entered as Level 1 (within-person) predictors. The Level 2 (between-person) predictors included a dummy variable indicating which study participants completed (Study 3 was set as the reference group), our main variables of interest (i.e., PT, ECp, ECn), demographic covariates (age, gender, and SES), and alexithymia. Additionally, we included two interaction terms: PT × ECp, and PT × ECn). Given that we did not observe any study-specific effects in our primary analyses with affect sharing, we did not test for study-specific effects in these exploratory analyses. For similar reasons, we did not include an interaction between video valence and our main effects of interest. A random intercept was included for each participant. Of note, all *b* values presented hereon for these analyses have been reversed for presentation purposes, following the initial reverse coding of the affect sharing data for GLMM analysis.

Results. Our results revealed no significant interaction between PT and Female Gender, $b = -.007$, 95% CI = [-.036, .022], $t(3892) = 0.492$, $p = .623$, between ECp and Female Gender, $b = -.001$, 95% CI = [-.033, .031], $t(3892) = 0.068$, $p = .946$, nor between ECn and Female Gender, $b = .015$, 95% CI = [-.017, .047], $t(3892) = 0.907$, $p = .365$.

PT × EC interaction

Statistical Analysis. To examine whether participant gender moderated any of the associations between self-reported empathic tendencies and affect sharing, we conducted a similar GLMM to our primary analyses on affect sharing. As with our primary analyses, affect sharing scores were first reversed, and +2 was added to these scores. We used a GLMM with γ distribution and a log linking function, with participants' congruence for each stimuli as the outcome variable. Video valence (negative=0, positive=1) and target expressivity were entered as Level 1 (within-person) predictors. The Level 2 (between-person) predictors included a dummy variable indicating which study participants completed (Study 3 was set as the reference group), our main variables of interest (i.e., PT, ECp, ECn), demographic covariates (age, gender, and SES), and alexithymia. Additionally, we included three gender-moderation interaction terms: PT × Female Gender, ECp × Female Gender, and ECn × Female Gender. Given that we did not observe any study-specific effects in our primary analyses with affect sharing, we did not test for study-specific effects in these exploratory analyses. For similar reasons, we did not include an interaction between video valence and our main effects of interest. A random intercept was included for each participant. Of note, all *b* values presented hereon for these analyses have been reversed for presentation purposes, following the initial reverse coding of the affect sharing data for GLMM analysis.

Results. Our results revealed no significant interaction between PT and ECp, nor between PT and ECn, in predicting affect sharing, $b = -.006$, 95% CI = [-.021, .008], $t(3893) = 0.817$, $p = .414$, and $b = .006$, 95% CI = [-.008, .021], $t(3893) = 0.889$, $p = .374$, respectively.

Identifying Problematic Participants

Open-ended questions. In order to identify potentially problematic participants (e.g., "bots" and "farmers") multiple open-ended questions were incorporated throughout the online survey. A free-response question was placed at the end of the Qualtrics survey, prompting participants to list two topics that had been addressed in the survey. In addition, for a subset of videos throughout the online EA video task, participants were prompted to describe the topic discussed in the video; specifically, for each participant, two videos (out of 9 videos) were selected as the target of the prompt (the first following the 1st video; the second after a video in a random position). Of note, however, the open-ended questions (for both the Qualtrics survey and the video task) were added as of November 1st, 2019, therefore a small subset of participants completed the surveys without answering these questions.

Response coding. Following data collection, a co-author (C.S.) reviewed all responses to these open-ended questions and flagged any potential "bot-based/farmer" responding. Specifically, for each participant, responses to the Qualtrics survey open-ended question and to the three open-ended video prompts were evaluated; based on these responses, participants were flagged as problematic if their answers were completely irrelevant to the question or to the video (e.g., "the topic was very useful for us," "nice and good survey," "very good speech"), markedly ungrammatical or non-sensical (e.g., "good behaviour was that person," "it was problem has taken at the past"), or clearly copy-and-pasted. These participants were classified as "problematic," whereas all other participants were classified as "acceptable." Of note, for a subset of 50 participants, no response to the open-ended questions could be coded: some participants did not respond to any of the open-ended questions (i.e., they left their answers blank), despite being presented the prompts, while others, as mentioned previously, had completed the survey prior to the addition of the open-ended questions. Based on acceptable internal consistency estimates for the Big Five Personality traits among the subsample (all αs >.80 except agreeableness, $\alpha = .66$), these participants were also included in the final dataset.

Sample selection. Of the 1,116 participants who completed the survey, 414 participants were classified as "problematic" (whereas 747 participants were deemed "acceptable"). To examine the consistency with which "problematic" (vs. "acceptable") participants responded to items of the same construct, Cronbach's alpha was obtained for each of the Big Five Inventory scales (BFI; John et al., 2008). According to Chmielewski and Kucker (2020), internal consistency estimates of the BFI are often in the mid .80s for MTurk samples. Internal consistency estimates from the current sample are presented in Table S6 below (Cronbach's αs <.60 are displayed as orange-colored cells, while αs >.80 are presented in green). In particular, the subset of "problematic" participants demonstrated extremely low internal consistency for each of the BFI scales. As a result, these "problematic" participants ($N = 414$) were removed from future analyses. Of particular interest, the vast majority of these "problematic" participants (84%) correctly responded to at least six of the eight attention check items; indeed, of the 414 participants, only 68 participants (17%) failed two or more attention check items. These numbers raise concerns regarding the utility of standard attention check items to identify potential "bots" and/or "farmers" within MTurk samples.

Supplemental Table 4. Internal Consistency Estimates (Cronbach's Alpha) for the BFI Scales Within the Group of "Problematic" Participants and the Group of "Acceptable" Participants

	Sample Size	O	C	E	A	N
Problematic	414	.51	.02	–.11	–.09	.13
Acceptable	747	.81	.86	.84	.80	.87

Note: O = openness, C = conscientiousness, E = extroversion, A = agreeableness, N = neuroticism.

Obtaining final sample. Finally, two duplicate cases were removed from the dataset, resulting in a final sample of 745 participants. For both these participants, the entry from the later timepoint was removed to avoid practice effects. Our final sample was further screened for cases of no/very low slider movement (i.e., they did not rate the music or video stimuli but rather left the slider at the "neutral" position throughout). We excluded 108 participants who provided no slider movement or no slider movement after the fifth data point, across all stimuli on both the music and social tasks. Our rationale for identifying participants with no slider movement after the fifth position was that individuals who only responded in the first few seconds of a clip were likely not motivated to respond honestly throughout.

Intercorrelations of Tasks

Table 1. Summary of Study Data Available for Each Task
(As shown in main text)

	Study 1	Study 2	Study 3
Biological Motion Task	✓	✓	
ER-40 Task		✓	
RMET	✓		
Empathic Accuracy	✓	✓	✓
Affect Sharing		✓	✓

Note: participants completed either the EA or the Affect Sharing task in Study 2 and Study 3, therefore it was not possible to examine the association between EA and Affect Sharing scores.

Scores on the Biological Motion task were positively correlated with RMET scores in Study 1, $r(1470) = .298$, $p < .001$, and ER-40 scores in Study 2, $r(281) = .137$, $p = .021$.

As with our primary analyses, we used multilevel modeling estimate the associations with EA scores (b estimates below reversed for presentation purposes), controlling for target expressivity and video valence. Scores on the Biological Motion task were positively associated with EA scores in Study 1 and Study 2, $b = .038$, $t(2457) = 1.97$, $p = .049$, and $b = .201$, $t(1200) = 2.73$, $p = .006$, respectively. By contrast, RMET scores (in Study 1) were not significantly associated with EA scores, $b = .015$, $t(2457)$, $p = .332$, and neither were ER-40 scores (in Study 2), $b = .001$, $t(1200) = .387$, $p = .699$.

For Affect Sharing (in Study 2), neither ER-40 scores, nor Biological Motion scores, were associated with total Affect Sharing scores, nor with valence-specific Affect Sharing (all p's >.05).