



OPEN

Gender differences in empathy, compassion, and prosocial donations, but not theory of mind in a naturalistic social task

Brennan McDonald & Philipp Kanske

Despite broad interest, experimental evidence for gender differences in social abilities remains inconclusive. Two important factors may have limited previous results: (i) a lack of clear distinctions between empathy (sharing another's feelings), compassion (a feeling of concern toward others), and Theory of Mind (ToM; inferring others' mental states), and (ii) the absence of robust, naturalistic social tasks. Overcoming these limitations, in Study 1 (N = 295) we integrate three independent, previously published datasets, each using a dynamic and situated, video-based paradigm which disentangles ToM, empathy, and compassion, to examine gender differences in social abilities. We observed greater empathy and compassion in women compared to men, but found no evidence that either gender performed better in ToM. In Study 2 (n = 226) we extend this paradigm to allow participants to engage in prosocial donations. Along with replicating the findings of Study 1, we also observed greater prosocial donations in women compared to men. Additionally, we discuss an exploratory, novel finding, namely that ToM performance is positively associated with prosocial donations in women, but not men. Overall, these results emphasize the importance of establishing experimental designs that incorporate dynamic, complex stimuli to better capture the social realities that men and women experience in their daily lives.

Taking others' perspective and sharing their emotions are crucial abilities underlying successful social interactions. Equally important are compassionate feelings for others, linked with the motivation to help¹. Neuroimaging and behavioral data have begun to show that the social affective states of *compassion* (caring feeling for another¹) and *empathy* (sharing emotions with another²), while conceptually connected, are discrete processes at the neural level^{1,3}. Further separated from social affect is the social cognitive process of mentalizing or *Theory of Mind* (ToM), involving the abstract, propositional representation of others' mental states^{4,5}.

A growing body of evidence shows that significant inter-individual differences exist between these distinct social abilities⁶. However, the evidence for inter-individual gender differences in these social capacities is inconclusive. Numerous studies have reported an advantage for women in emotional face recognition⁷, self-reports of empathy/compassion^{8–10}, and affective understanding of others^{11–13}. Conversely, other investigations yield ambiguous findings^{14–16}, observe extremely small effect sizes in large samples⁸, or find no gender differences in social abilities^{17–19}. Regarding prosocial behaviour, results again appear dependent on the sample, situation, or definitions of prosocial behaviour²⁰. Evidence does suggest women are often more prosocial compared to men^{21,22}. However, several studies also show men are more altruistic than women in situations involving strangers, requiring the helper's initiative rather than responding to a request for help^{23,24}.

A possible explanation for these inconsistent findings may be a narrow focus on isolated components of social processing (e.g. focusing on self-reports of empathy) or using unrelated measurement tools (e.g. comparing empathy and ToM measured by two unrelated tasks) leading many studies to inadequately distinguish between social affect and social cognition⁹. However, as these social processes are claimed to be separable^{25,26}, it is important that measurement is performed within the same sample and paradigm to allow for gender-specific differences to be distinguished. Additionally, the experimental designs used to evoke social responses play a crucial role in accurately capturing real-world social functioning and can strongly influence observed results. To date, the majority of research examining gender differences in social abilities rely on static measures, often requiring responses to isolated stimuli (e.g., videos/images of postures, eyes, faces, or individual body-parts),

Clinical Psychology and Behavioral Neuroscience, Faculty of Psychology, Technische Universität Dresden, Chemnitz Straße 46, 01187 Dresden, Germany. email: brennan.mcdonald@tu-dresden.de

self-report questionnaires, or tasks via a computer with no obvious social partner. By using more ecologically valid social stimuli that are dynamic, multimodal and situated – and thus better reflecting the complexity of real-world social behaviour – inter-individual differences in social affect and social cognition may be more accurately captured, including potential differences between men and women. For the current studies, we asked if gender differences in empathy, compassion, ToM and prosocial donations are observable during naturalistic, dynamic, video-based social situations.

For Study 1, we combined the results of three independent, previously published experimental datasets each using the EmpaToM task ($N = 295$), a validated assessment of social affect and social cognition³. The task involves 48 videos displaying either neutral or emotionally negative narrations. Following each video, participants' empathic responding, compassion, and ToM performance were measured. The EmpaToM task thus provides the unique advantage of allowing social affect and social cognition to be assessed near simultaneously with respect to the same stimuli. Additionally, because the EmpaToM task uses video narratives, participants are exposed to situated, dynamic and multimodal social stimuli, including the narrator's autobiographical story, body language, facial expressions and vocal inflections, increasing the ecological validity of the social measures.

For Study 2, we utilized an additional independent dataset ($N = 226$), which included an extended version of the EmpaToM task with a novel measure of prosociality/charitable donations. Along with allowing for a replication of Study 1, this additional measure allowed for empathy, compassion and ToM to be directly correlated with each participant's willingness to engage in prosocial behavior.

Based on previous evidence suggesting a social affect advantage for women, we hypothesized the following: women will show greater affect sharing (i.e., empathy) and report greater compassion during the emotionally negative narrations compared to men. Regarding ToM, the evidence is far less clear whether gender differences exist. Thus, we aimed to test whether a gender advantage exists in ToM performance. With respect to the prosocial donation measure in Study 2, we hypothesized based on previous evidence that women would show greater willingness to engage in prosocial donations compared to men. Finally, we examined if social affect and social cognition measures correlated with prosocial donations, both across the sample and for each gender. We then explored subsequent findings of interest by testing if significant measures were associated with participant's prosocial donations and if gender differences existed in this regard.

Results: study 1

Gender differences in empathy and compassion

While controlling for differences between the experimental locations (MRI scanner or laboratory) and participant's age, we analysed the effects of gender differences on empathy and compassion using a mixed design, repeated measures ANCOVA with gender (men vs. women) as the between-subject factor and video emotionality (emotionally negative vs. neutral videos) as the within-subject factor. Differences regarding empathy and compassion would emerge as an interaction effect between participants' gender and the emotionality of the video narrations. Effect size is reflected by the partial eta squared (η^2p).

The analyses showed a significant main effect of video emotionality for participants' ratings of the valence of their own affect ($F(1, 291) = 358.874$, $p < 0.001$, $\eta^2p = 0.55$) and for participant's compassion ratings ($F(1, 291) = 328.242$, $p < 0.001$, $\eta^2p = 0.53$), indicating that emotionally negative videos elicited stronger negative affect (i.e., empathic emotion sharing) and stronger compassion than neutral videos. Additionally, a main effect of participants' gender on affective valence ($F(1, 291) = 11.845$, $p < 0.001$, $\eta^2p = 0.039$) and compassion ($F(1, 292) = 10.382$, $p = 0.001$, $\eta^2p = 0.034$) was observed. Specifically, women reported stronger overall negative affect and compassion than men. However, we also observed significant interaction effects between participants' gender and video emotionality for both empathy ($F(1, 291) = 9.364$, $p = 0.002$, $\eta^2p = 0.031$) and compassion ($F(1, 291) = 10.382$, $p = 0.001$, $\eta^2p = 0.034$). Post-hoc, paired *t*-tests confirmed the hypotheses: Following emotionally negative narrations, empathy ($t(293) = 4.104$, $p < 0.001$, $d = 0.67$) and compassion ($t(293) = 5.126$, $p < 0.001$, $d = 0.80$) were significantly stronger in women than in men (see Fig. 1).

Gender differences in theory of mind

To test for gender differences in ToM performance, we analysed accuracy and reaction times in the multiple-choice questions, using a repeated measures ANCOVA with question type (factual-reasoning vs. ToM) as the within-subject factor and participants' gender (men vs. women) as the between-subject factor, whilst controlling for differences between the experimental locations (MRI scanner or laboratory) and participant's age. Gender differences in social cognition would emerge as an interaction effect between question type and participants' gender with increased accuracy and shorter reaction times indicating better ToM performance.

The analyses showed no significant main effect for question type with respect to accuracy ($F(1, 291) = 1.636$, $p = 0.20$, $\eta^2p = 0.006$). However, a significant main effect of question type was observed for reaction times ($F(1, 291) = 10.229$, $p < 0.002$, $\eta^2p = 0.034$) with participants demonstrating shorter reaction times for ToM questions, indicating that ToM questions were slightly easier than the factual reasoning questions, as observed previously for this task^{3,27}. No main effect of gender was observed for reaction times ($F(1, 291) = 4.801$, $p = 0.02$, $\eta^2p = 0.016$), however, a main effect of gender was observed for question accuracy ($F(1, 291) = 3.036$, $p = 0.08$, $\eta^2p = 0.01$) with women demonstrating greater overall accuracy for both question types. Finally, the analyses revealed no significant interaction effect between participants' gender and question type for both accuracy ($F(1, 291) = 1.282$, $p = 0.25$, $\eta^2 = 0.004$) and reaction times ($F(1, 291) = 0.565$, $p = 0.45$, $\eta^2p = 0.002$). Thus, we found no evidence for gender differences in ToM performance (see Fig. 1).

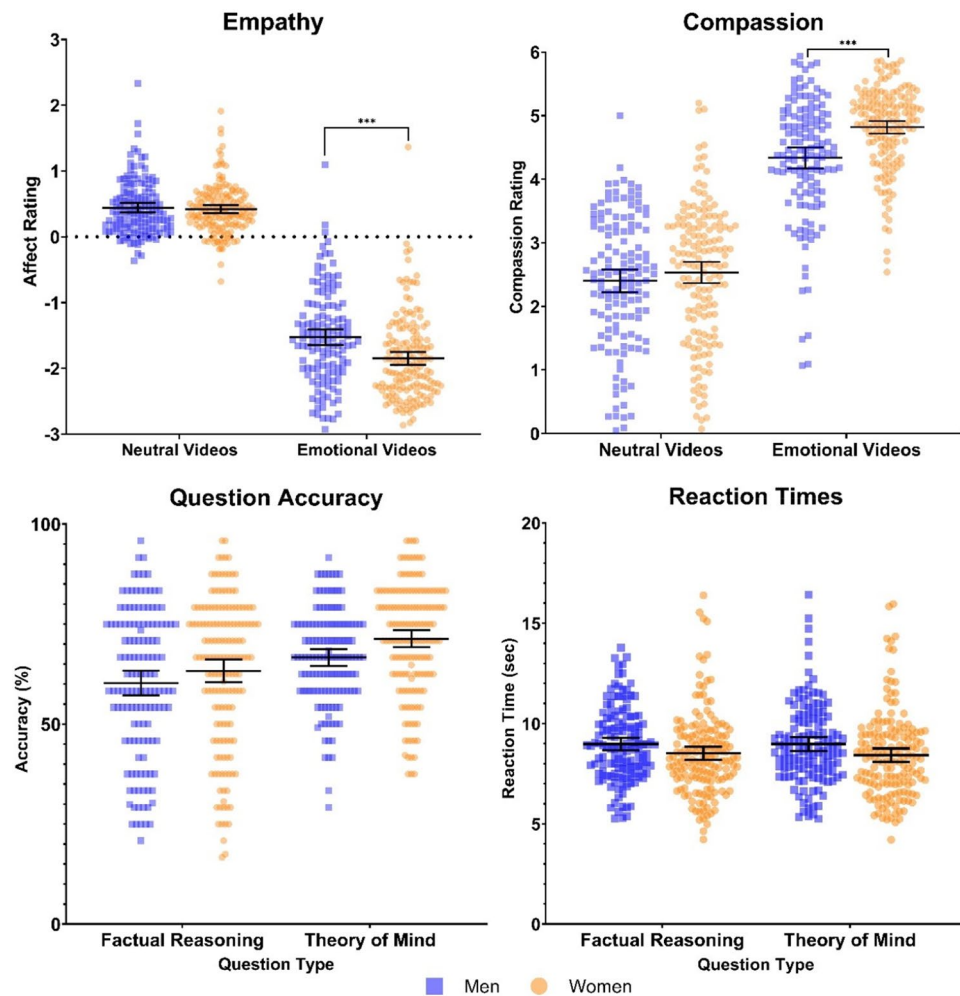


Figure 1. *** $p < 0.001$. Study 1: Empathy ratings showed that emotionally negative videos elicited stronger negative affect compared to affect ratings after neutral videos. Significant interaction effects were observed between emotional video narrations and participant's gender for empathy and compassion measures. Analysis of question accuracy and reaction times (on correctly answered trials) revealed no significant interaction effects among theory of mind questions and participants' gender. Bars indicate mean \pm 95% CI.

Results: study 2

Gender differences in empathy and compassion

We analysed the effects of gender differences on empathy and compassion using a repeated measures ANCOVA with gender (men vs. women) as the between-subject factor and video emotionality (emotionally negative vs. neutral videos) as the within-subject factor, controlling for participant's age. Differences regarding empathy and compassion would emerge as an interaction effect between participants' gender and the emotionality of the video narrations. The analyses for Study 2 showed a significant main effect of video emotionality for participants' ratings of the valence of their own affect ($F(1, 223) = 87.505$, $p < 0.001$, $\eta^2p = 0.28$) and for participant's compassion ratings ($F(1, 223) = 86.265$, $p < 0.001$, $\eta^2p = 0.27$), indicating that emotionally negative videos elicited stronger negative affect (i.e., empathic emotion sharing) and stronger compassion than neutral videos. Additionally, a main effect of participants' gender on affective valence ($F(1, 223) = 17.013$, $p < 0.001$, $\eta^2p = 0.07$) and compassion ($F(1, 223) = 9.579$, $p = 0.002$, $\eta^2p = 0.041$) was observed. Specifically, women reported stronger overall negative affect and compassion than men. However, we also observed significant interaction effects between participants' gender and video emotionality for both empathy ($F(1, 223) = 4.298$, $p < 0.03$, $\eta^2p = 0.019$) and compassion ($F(1, 223) = 5.235$, $p = 0.02$, $\eta^2p = 0.023$). Post-hoc, paired *t*-tests confirmed the hypotheses: Following emotionally negative narrations, empathy ($t(225) = 3.493$, $p < 0.001$, $d = 0.46$) and compassion ($t(225) = 3.948$, $p < 0.001$, $d = 0.52$) were significantly stronger in women than in men (see Fig. 2).

Gender differences in theory of mind

To test for gender differences in ToM performance, we analysed accuracy and reaction times in the multiple-choice questions, using a repeated measures ANCOVA with question type (factual-reasoning vs. ToM) as the within-subject factor and participants' gender (men vs. women) as the between-subject, controlling for

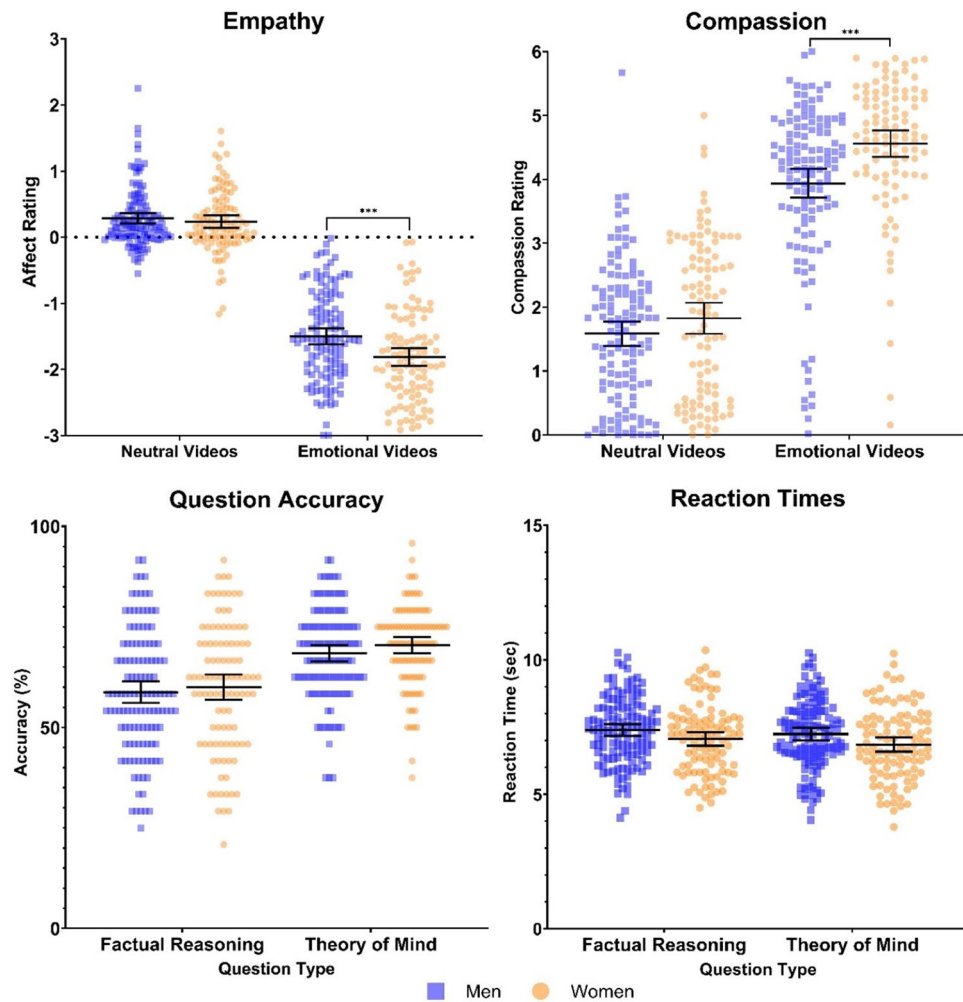


Figure 2. *** $p < 0.001$. Study 2: Replicating the results of Study 1 in an online environment, empathy ratings showed that emotionally negative videos elicited stronger negative affect compared to affect ratings after neutral videos. Significant interaction effects were observed between emotional video narrations and participant's gender for empathy and compassion measures. Analysis of question accuracy and reaction times (on correctly answered trials) revealed no significant interaction effects among theory of mind questions and participants' gender. Bars indicate mean \pm 95% CI.

participant's age. Gender differences in social cognition would emerge as an interaction effect between question type and participants' gender with increased accuracy and shorter reaction times indicating better ToM performance. The analyses showed a significant main effect for question type with participants demonstrating higher accuracy ($F(1, 223) = 12.065, p < 0.001, \eta^2p = 0.051$) for ToM questions, indicating that ToM questions were slightly easier than the factual reasoning questions, as observed previously for this task^{3,27}. No main effect for question type was observed with respect to reaction times ($F(1, 223) = 1.457, p = 0.22, \eta^2p = 0.05$). No main effect of gender was observed for question accuracy ($F(1, 223) = 1.364, p = 0.24, \eta^2p = 0.006$), while a main effect of gender was observed for question reaction times ($F(1, 223) = 5.504, p = 0.02, \eta^2p = 0.024$) with women demonstrating overall shorter reaction times for both question types. Finally, the analyses revealed no significant interaction effect between participants' gender and question type for both accuracy ($F(1, 223) = 0.185, p = 0.66, \eta^2p = 0.001$) and reaction times ($F(1, 223) = 1.457, p = 0.50, \eta^2 = 0.002$). Thus, we found no evidence for gender differences in ToM performance (see Fig. 2).

Gender differences in prosocial donations

Utilizing the prosocial donation data from Study 2 ($n = 226, 101$ women), we next tested if there were gender differences in prosocial donations. Due to the relatively large number of participants, who did not donate at all (men: $n = 54$, women: $n = 29$), the prosocial data was zero inflated with a right tailed distribution (men: skewness = 2.119, women: skewness = 1.304). Thus, the assumptions for an independent t-test analysis were not met. As alternatives that better fit the data distribution we employed a Mann–Whitney U test and a two-sample Kolmogorov–Smirnov test. The two-sample Kolmogorov–Smirnov test indicated that the cumulative distribution for donations by women differs from the cumulative distribution for donations by men ($D = 0.225, p = 0.007$,

see Fig. 3a and b). The Mann–Whitney U test revealed that women (Md=0.30, n=101) were willing to donate significantly more money on average to a charitable cause than men (Md=0.02, n=125), $U = 4859$, $z = -3.052$, $p = 0.002$, with a small effect size $r = 0.20$ (see Fig. 3c.).

Raw correlations between prosocial donations and empathy, compassion and theory of mind

To explore whether a possible relationship exists between social affect/cognition and prosocial donations we calculated correlation coefficients for Study 2, both for the total sample and for each gender (see Table 1). Because the prosocial donation data violated the normality assumption required for the calculation of Pearson’s r , the association between measures was examined using a nonparametric statistic (Spearman’s ρ). To enable the correlation of inter-individual differences, the following individual averages were used. The ability to empathically share others’ emotion is best reflected in participants’ reported affect after the emotionally negative narrations, capturing their affective state after viewing another person in distress. Because compassion and prosocial donations are positive social attributes regardless of video content, these measures reflect participants’ average value across all narrations (as done previously²⁸). ToM accuracy and factual-reasoning accuracy are the percentage of questions answered correctly for each question type. Reaction times reflect the time taken to answer either ToM or factual reasoning questions on those trials where the answer given was correct.

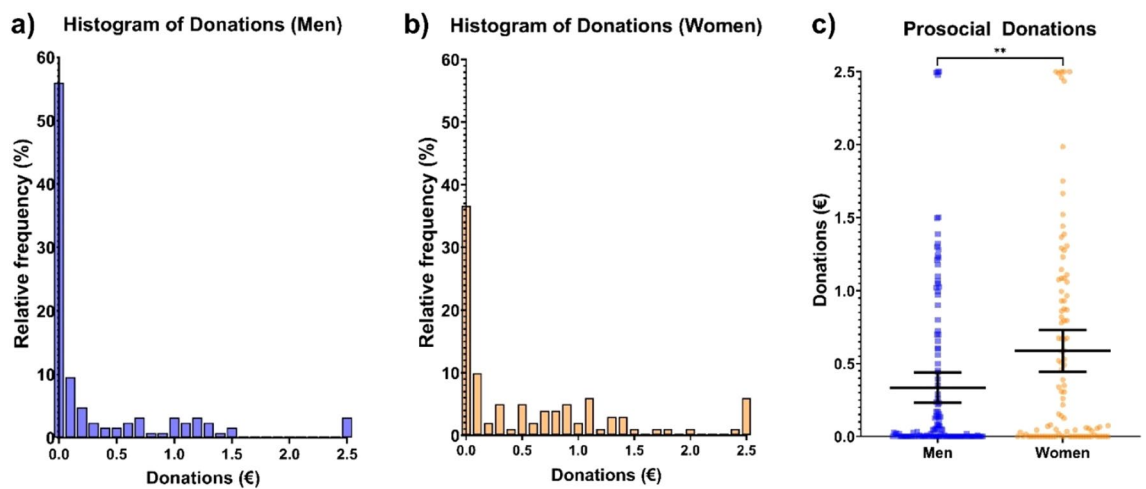


Figure 3. $**p < 0.01$. (a) and (b) Histograms showing the relative frequency of the average donations for men (n = 125) and women (n = 101) with a centre bin of 0.1. (c) Significant difference observed between the average donation amounts in women compared to men. Bars indicate mean \pm 95% CI.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|------------------------------------|---|-------------------------------------|------------------------------------|------------------------------------|----------------------------------|------------------------------------|-----------------------------------|
| 1. Empathy | – | -0.56* -0.49* (-0.60*) | -0.17* -0.06 (-0.22*) | -0.21* -0.14 (-0.20*) | -0.00 -0.02 (-0.09) | -0.25* -0.19 (-0.28*) | -0.05 -0.07 (-0.14) |
| 2. Compassion | | – | 0.33* 0.35* (0.25*) | 0.03 0.08 (-0.07) | 0.13* 0.21* (0.12) | 0.10 0.03 (0.15) | 0.17* 0.25* (0.17*) |
| 3. Prosocial donation | | | – | 0.00 0.23* (-0.21*) | 0.08 0.11 (0.09) | 0.06 0.14 (-0.03) | 0.11 0.09 (0.15) |
| 4. ToM accuracy | | | | – | -0.11 -0.19 (-0.02) | 0.46* 0.47* (0.43*) | -0.16* -0.16 (-0.15) |
| 5. ToM reaction time | | | | | – | -0.07 -0.19* (0.04) | 0.83* 0.87* (0.79*) |
| 6. Factual reasoning accuracy | | | | | | – | -0.11 -0.14 (-0.07) |
| 7. Factual reasoning reaction time | | | | | | | – |

Table 1. Inter-scale correlations (Spearman’s ρ) for measures of Social Affect, Social Cognition, and Prosocial Donations. $*p < 0.05$, Coefficients in bold represent the entire sample. Coefficients not in bold are for women only (n = 101). Coefficients for men reported in parentheses (n = 125).

Full sample

Both empathy and compassion correlated significantly with prosocial donations. Neither ToM accuracy/reaction times nor factual-reasoning accuracy/reaction times showed significant correlations with prosocial donations.

Gender differences

On inspection, the only correlation coefficients that clearly differed between men and women were that of ToM accuracy vs. prosocial donations with women showing a significant positive correlation and men showing a significant negative correlation (see Fig. 4.). To test if these correlation coefficients significantly diverged, we performed a Fisher's z-transformation. This method is sufficient for testing differences between Spearman's ρ coefficients²⁹. The analysis revealed that for ToM accuracy vs. prosocial donations the correlation coefficient for women differed significantly from that of men ($|\rho_{\text{diff}}| = 0.44$, $Z = 3.298$, $p < 0.001$). No other gender difference between correlation coefficients was statistically significant ($p > 0.05$).

Prosocial donations and ToM performance

Finally, to explore the possible factors influencing the observed divergent correlations between the average ToM performance and prosocial donations in men and women we used a generalized linear model with a gamma distribution and logarithmic mean function expand for repeated measures (as a robust alternative to a linear regression model with distributional assumptions that better fit the data). We also performed the analysis with the robust sandwich method to estimate the standard errors in these models avoiding the assumption that the model correctly specifies the variances of the residuals.

This data-driven model was used to test if the interaction between ToM accuracy and participants' gender was significantly associated with prosocial donations, controlling for participants' factual reasoning accuracy, age and highest level of education. Note: a Chi-square test revealed no gender difference in the level of education (X^2 [df = 2, N = 226] = 1.264, $p = 0.53$). Level of education was included as a categorical, nominal data in this model. Level of education was included in the model due to the positive association between educational attainment and earning potential³⁰. We also included the interaction term factual-reasoning accuracy x gender to control for the more general influence of question performance and participants' gender on donations.

The full results of the model are available in the supplementary data. We found a significant interaction effect for ToM accuracy vs. participants' gender ($\beta = 1.05$, $p = 0.013$) and no significant interaction for factual reasoning

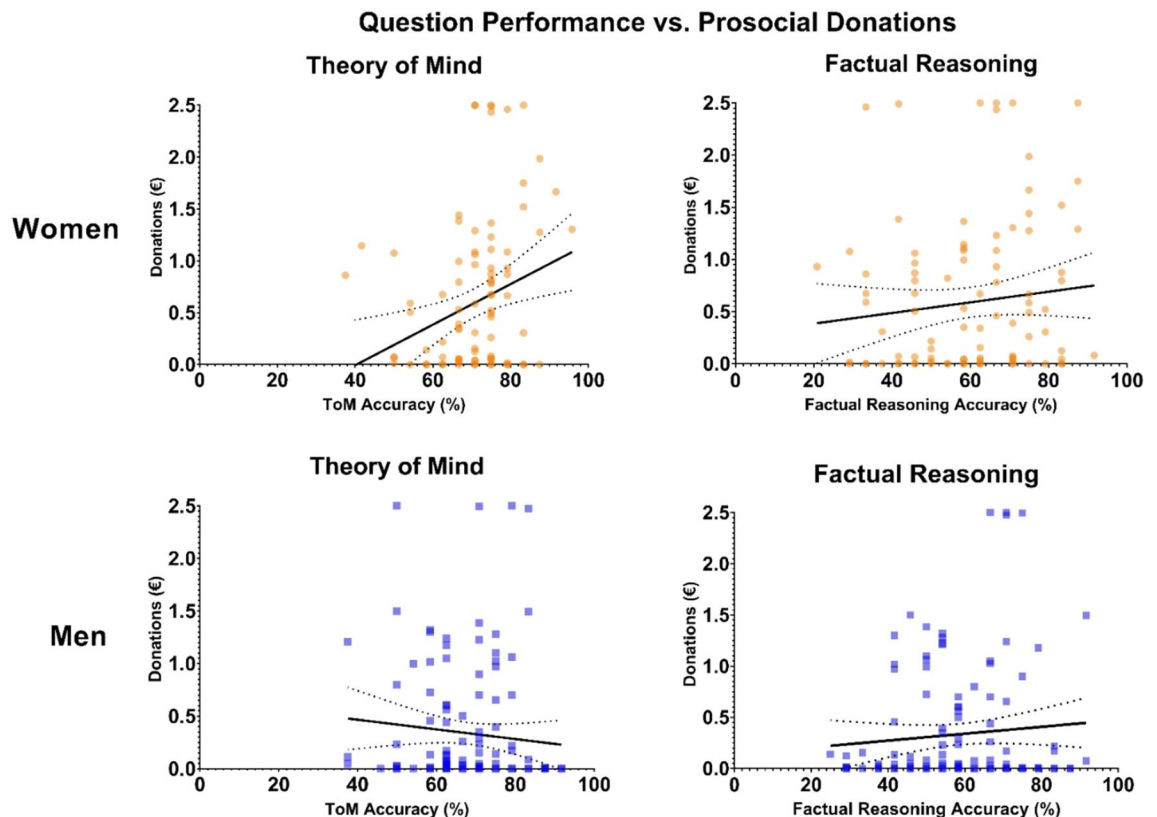


Figure 4. Scatterplots displaying linear correlations between ToM accuracy vs. average donation and factual reasoning accuracy vs. average donation for men ($n = 125$) and women ($n = 101$). Dotted curves indicate the 95% confidence bands of the best-fit line. ToM accuracy showed a significant positive correlation with prosocial donations in women ($\rho = 0.23$) and a significant negative correlation in men ($\rho = -0.21$). Factual accuracy did not correlate significantly.

accuracy vs. participants' gender ($\beta = 0.978$, $p > 0.05$). Specifically, ToM accuracy significantly predicted a 2.8% (95% CI = 0.1% to 5.5%) increase in donations per percentage point in women (see Fig. 5). We found no evidence that ToM performance significantly predicts donations in men (a non-significant 2.2% decrease in donations per ToM percentage point with 95% CI = -5.2% to 0.6%). To assess goodness of fit, Akaike's Information Criterion (AIC) was used to compare the full model and a reduced model (lacking the interaction term of interest: gender \times ToM accuracy). Following Arnold³¹, lower AIC coefficients were considered to indicate better model fit. We observed a lower AIC for the full model (AIC = 69.14) compared to the reduced model (AIC = 81.51), indicating that the full model better fit the data when including the interaction term of interest.

Discussion

The current studies investigated gender differences in social affect, social cognition and prosocial donations during naturalistic social narrations. We observed significantly greater empathy, compassion and prosocial donations in women compared to men. By contrast, we found no evidence that either gender performed better in ToM. However, ToM performance was positively associated with prosocial donations in women, but not men. These data expand on previous findings showing an advantage for women in social affective processes and prosocial behaviour, providing additional evidence for inter-individual variation in these separable social abilities.

Various explanations exist as to why social gender differences may occur. For example, evolutionary models propose that variation in reproductive investment generates sexual-selective pressures, differentially favouring the evolution of sex-specific social behaviours³². In contrast, gender socialization models posit that cultural beliefs about gender roles influence gender expectations, orienting women and men towards different social behavioural patterns^{33,34}. Importantly, both the existence and impact of such gender differences, whether social or biological, are a point of much heated debate^{35–39}. For the current studies, we do not commit to any ultimate cause for observed social gender differences. Rather, we focused on examining the possible presence of such differences by addressing several limitations of previous research. Broadly, our goal was to investigate gender differences in relation to two factors: i) the comprehensive, but distinct investigation of social affective and social cognitive processes and ii) the examination of these processes in a robust, naturalistic setting.

Research is beginning to converge on the finding that empathy, compassion and ToM are dissociable processes^{25,26}. Additionally, significant intra- and inter-individual differences in these social abilities exist⁶. Despite this, evidence for social gender differences remains unclear. Importantly, as these are claimed to be separable social processes, measurement within the same sample and paradigm is necessary to allow gender differences to be accurately distinguished. Thus, our first goal was to make clear distinctions between empathy, compassion, and ToM, utilizing high-level control conditions (i.e. neutral vs emotional videos and factual-reasoning vs ToM questions). Our second goal was to examine gender differences using stimuli combining numerous aspects of real-life social interactions, thereby overcome some of the limitations of previous research. Specifically, previous investigations examining social gender differences have produced conflicting results regarding whether such differences exist. A reason for these inconsistent findings may be that by over-isolating specific social abilities (e.g., recognizing emotions based on still images of eyes), or having no clear social partner (e.g., indirect behavior on a computer screen) researchers may limit themselves to investigating tiny effect sizes. Indeed, Baez and colleagues⁸ tested gender differences in empathy (i.e. harm of others) using short, animated scenarios of isolated body-parts in a very large sample ($n > 10,000$) and only found a miniscule effect. However, when taken

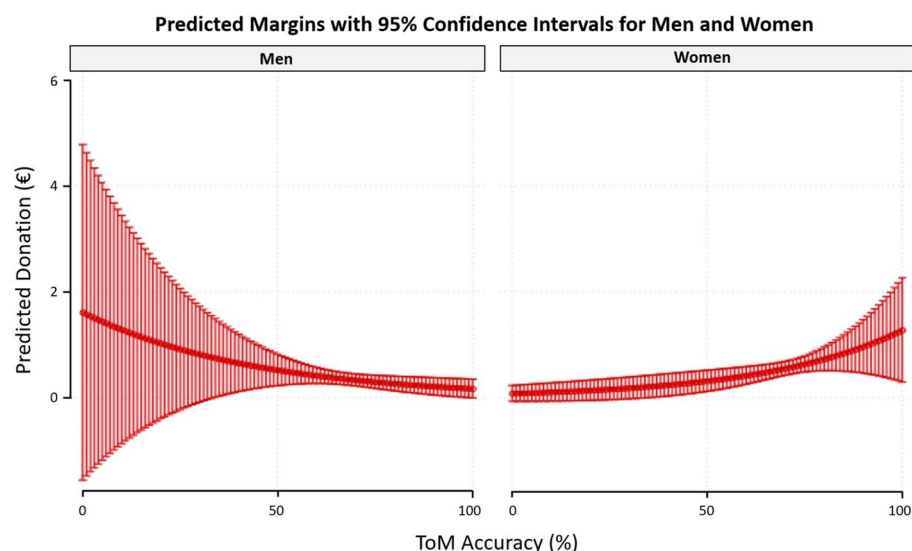


Figure 5. Predicted margins with 95% confidence intervals displaying the regression curves for ToM accuracy vs. average donation for men ($n = 125$) and women ($n = 101$), controlling for factual reasoning accuracy, age, and level of education. ToM accuracy significantly predicts donations in women and does not predict donations in men.

together, previous findings appear to suggest that women have a slight advantage in detecting affective facial expressions⁷, recognising vocal emotions⁴⁰, and understanding body language/emotional posture⁴¹. The present studies indicate that by naturalistically combining these discrete social facets in complex, multimodal stimuli of realistic people, robust differences in social affect between the genders can be detected. Specifically, our results exhibit that women have an advantage in empathy and compassion over men.

Regarding our social cognition findings, we did not observe any gender advantage in ToM ability over and above factual-reasoning ability. While evidence suggests that a female advantage in ToM may be present in childhood/adolescence^{42–44}, findings with adults are far more mixed, with some results indicating that women may show an advantage on ToM tasks involving an affective component compared to men^{11,13,45,46}, while other findings suggest no clear gender differences^{9,47–49}. Importantly, as with measures of social affect, these inconsistent results may result from a lack of clear distinctions between the separable processes of social affect and social cognition, thus creating a pronounced degree of overlap between measures. Moreover, this is made additionally confusing by the use of multiple different terms/synonyms in the literature referring to these processes, with social cognition also at times being referred to as cognitive ToM, mentalizing, or cognitive empathy, while empathy has been termed affective ToM or emotional empathy.

Here, we define ToM as the ability to understand and use abstract propositional knowledge about another's mental states. Although we observed a main effect of gender such that women were more accurate (Study 1) or had shorter reaction times (Study 2) for both question types, we did not observe an interaction effect between question type and gender. This indicates that while women may be better in general at answering the questions during the EmpaToM task, neither gender demonstrate a specific advantage for ToM questions. We interpret this lack of a gender difference as reflecting the complex, cognitive nature of ToM processing. Based on our definition, ToM performance during the EmpaToM task relies on a range of general cognitive abilities including attention, reasoning, memory, and language comprehension, with ToM questions differing from factual-reasoning control questions only in social content. As such, this multifaceted construct relies on the simultaneous application of diverse abilities that fall under the rubric of intelligence. While there exist well-established cognitive gender differences in certain specific tasks⁵⁰, there is no difference in average general intelligence between men and women^{51,52}. Our findings expand on this, demonstrating that when using factual-reasoning ability as a high-level control condition, men and women do not differ in their ToM ability. Put another way, the results of the present study provide evidence that neither men nor women have a cognitive advantage in abstract reasoning about another's mental states during naturalistic situations.

While we found no overall gender difference in social cognition, we did however observe differences in prosocial donations between men and women. Specifically, we found that women, on average, donated more money than men. This is consistent with a large body of work indicating that women are often more prosocial compared to men^{21,22}.

Subsequent to this finding, we then explored the linear correlations between prosocial donations and empathy, compassion, ToM and factual reasoning, both at the level of the full sample and between the genders. At the level of the full sample, we observed significant correlations for empathy with prosocial donations, while finding no significant correlations for ToM and factual reasoning. These findings provide support for the empathy-altruism hypothesis^{53,54} and further underline the separable nature of social affective and social cognitive processes⁵. The empathy-altruism hypothesis highlights the strong association between empathy and the motivation to increase the wellbeing of another person^{54,55}, with numerous studies supporting the connection between empathy and altruistic behaviour^{56–58}. Our results extend these findings by demonstrating that empathy correlates significantly with prosocial donations within the same naturalistic video-based task. Additionally, we also observe a significant association between compassion and prosocial donations, which is also to be expected. Compassion is defined as a feeling of warmth or loving-kindness towards others and is thus understood as an other-directed aspect of social affect, while empathy relates to one's own (negative) affect in response to the distress of another. In this regard, our data exhibit a stronger correlation for compassion with prosocial decision compared to empathy, consistent with previous data⁵⁸ and the strong motivational component towards others accredited to compassion⁵⁹.

Finally, at the level of the full sample, we observed no significant correlation with ToM and factual reasoning accuracy or RTs. This finding is consistent with the previous observations that affective states represent a more immediate activator of behavioural response than cognitive processes⁶⁰. However, in addition to these findings at the level of the full sample, we also observed gender differences in the linear correlations between prosocial donations and ToM accuracy, with women showing a significant positive correlation and men a significant negative correlation (see Table 1). To examine this result further, we performed an exploratory generalized linear model that better fit the data. This analysis revealed that ToM performance in women, but not men, was positively associated with prosocial donations. Put another way, although we found no evidence for a ToM advantage between men and women, this model suggests that women's average ToM accuracy is positively associated with prosocial donations, controlling for age, level of education, and factual-reasoning performance, with no such associated pattern observed for men.

We speculate that this exploratory, data-driven finding may relate to differences in men and women's interest in social stimuli. Previous evidence suggests that women tend to show a greater interest/responsiveness to social stimuli when compared to men^{7,40}. Additionally, research indicates that both social affective and social cognitive processes are involved in prosocial decision-making²⁷. We suggest that an interest in social stimuli may possibly favour the recruitment of social cognition during prosocial decision-making. Put another way, women's average greater interest in social stimuli compared to men may favour the observed relationship between the ability to take another's perspective (ToM performance) and respond to another's needs (prosocial donations). This social interest may also foster women's willingness to donate more when compared to men, as observed in the current studies. It is important, however, to highlight that this was a data-driven exploratory model based on a single sample and as such requires further replication and conformation before any strong claims can be

supported⁶¹. Future research may thus expand on this exploratory result by replicating this pattern between the genders and probing whether individual differences in social interest mediate the relationship between gender, ToM performance, and prosocial donations.

The current studies also have limitations that warrant mention. First, Study 1 utilized multiple previously published summary datasets, while Study 2 derived from a separate pre-registered investigation. Because the available datasets differed in how demographic data was collected and already had exclusion criteria applied, we chose not to apply our own additional exclusion of participants, beyond those of the original publications. This approach however partially limits the comparability of Study 1 and Study 2 as different approaches are used to define the samples. Second, Study 2 was conducted in an online environment and although participants were requested to undertake the experiment in a quiet environment alone, we cannot ensure this was the case. Third, an additional limitation of the current studies is the lack of videos with positive valence. Because the EmpaToM paradigm, to date, does not have validated video sets with positive valence, we were unable to test if gender differences in social affect and/or ToM are influenced by mirroring, sharing, or understanding positive emotions. However, this is an important facet for understanding social emotions, as previous data suggest that women self-report greater overall negative emotions than men, but show no difference in overall positive emotions⁶². Thus, a fruitful avenue of future investigation will be to extend our findings by including positive valence in social scenarios examining gender differences in empathy, compassion, prosocial donations, and ToM.

In conclusion, our studies examined gender differences using a naturalistic task that allows for clear distinctions between social affect and social cognition. We observed greater empathy, compassion and prosocial donations in women compared to men, but found no evidence that either gender performed better in ToM. However, ToM performance was positively associated with donations in women, but not men in our exploratory, data-driven model. Ultimately, our findings call for both the clear definition of social abilities as well as the use of naturalistic social tasks that incorporate multimodal, dynamic and situated stimuli to better capture the social realities that men and women experience in their daily lives.

Methods: study 1

Study 1 integrates three previously published, experimental datasets to examine gender differences in social abilities. For each experiment, a version of the EmpaToM task was performed. The datasets were selected based on their availability online with full summary EmpaToM and gender data. Differences and similarities between the datasets are summarized in Table 1. Dataset 1 (D1) is from a published functional imaging study examining differences in social abilities between young and elderly adults⁶³. The data for D1 are available at <https://doi.org/10.17605/OSF.IO/7EDBN>. Dataset 2 (D2) is from a published functional imaging study examining social abilities and personality⁶⁴. The data for D2 are available at <https://osf.io/f7zcr/>. Dataset 3 (D3) is from a published study in which the EmpaToM was a baseline measure to investigate meditation practice⁵⁸. The data for D3 is available at https://osf.io/tu2gj/?view_only=f285308c035d451ca2fce5f3788f97e3. The complete dataset and analysis scripts for Study 1 are available at https://osf.io/xgma6/?view_only=d0b7d3e1f10a415fa6c033678de5c9ff. The study protocol for each of the three studies was approved by the Ethics Committee of Technische Universität Dresden (D1: EK-486112015, D2: EK-133042018, D3: EK-180052018). All methods were performed in accordance with the relevant guidelines and regulations. Using G*Power 3⁶⁵, an a priori power analysis estimated a total sample of 200 participants ($\alpha = 0.05$, $1 - \beta = 0.8$) based on a small effect size ($f = 0.1$) would be sufficient to detect gender differences. Mann–Whitney U test and ANCOVA analyses were conducted using SPSS (IBM Corp., 2021, Version 28.0.1.0). Graphics were produced using Graphpad Prism (<https://www.graphpad.com>, Version 9.3.1). This study was not preregistered.

Participants

In total, 295 participants, consisting of 160 women (age mean = 31.21, SD = 16.84) and 135 men (age mean = 31.96, SD = 17.00), were included in the current analyses. A Mann–Whitney U test indicated that men and women did not differ significantly in age ($U = 11,772$, $p = 0.18$). As these previously published datasets were only available as summary data online, with differing demographic measures and already-applied exclusion criteria, no additional layers of exclusions were performed beyond those originally reported in the original publications.

For D1, reported exclusion criteria included excessive or illicit substance use, neurological or cognitive impairment and MRI related reasons. One hundred and one individuals were initially recruited and sixteen individuals were excluded from the final dataset due to MRI related issues (15) or EmpaToM values exceeding three standard deviations of the mean (1), leaving a final data set of 85 participants (43 women). For full details on participants see Stietz and colleagues⁶³.

For D2, 168 participants (age: < 18 or > 60 years) were originally invited. Exclusion criteria were MRI related reasons (27), severe mental disorder (3), and not providing sufficient Ecological Momentary Assessment data (16). Thus, 46 participants were excluded yielding a final sample of 122 participants. One participant was additionally excluded from the dataset due to incomplete EmpaToM data, leaving 121 participants (58 women) for the current analysis. For full details see Hildebrandt and colleagues⁶⁴.

For D3, participants were recruited through local advertisements via flyer or e-mail for a study examining meditation practice. 100 participants were initially recruited. Six participants were excluded from the final sample due to technical issues, leaving a final sample of 94 (69 women). For full details see Lehmann and colleagues⁵⁸.

Informed consent was obtained from all participants prior to the experiments and they were compensated for their time (see Table 2).

| | Dataset 1 | Dataset 2 | Dataset 3 |
|---|--|--|----------------------------------|
| Sample size | 85 (43 women) | 121 (58 women) | 89 (59 women) |
| Mean age (range) | YA: 24.0 (18–30) OA: 69.5 (65–77) | 25.5 (18–57) | 24.9 (18–65) |
| Original investigation | Age differences | Personality | Meditation/prosocial behaviour |
| Study design | Mixed | Within-subject | Within-subject |
| Previous publication | Stietz et al. ⁶³ | Hildebrandt et al. ⁶⁴ | Lehmann et al. ⁵⁸ |
| Experiment location | MRI Scanner | MRI Scanner | Laboratory |
| Rating bar duration* | 7 s | 4 s | 4 s |
| Question response duration [†] | 25 s | 15 s | 15 s |
| Additional measures during original investigation | Socio-demographic questionnaire ¹ Cognitive battery ^{2,4} | Socio-demographic questionnaire ¹ 14-day ecological momentary assessment ³ Unspecified online survey including trait measures of personality ³ Interpersonal Reactivity Index ³ | Meditation training ³ |
| Compensation | €8.50 | €120 | €30 or course credit |

Table 2. Differences between the three datasets in Study 1. YA younger adults, OA older adults. *Duration the rating bar appeared on the screen during the EmpaToM task (empathy and compassion measures). †Maximum duration for the ToM and factual reasoning questions during the EmpaToM task. ¹Measure taken prior to the EmpaToM task. ²Measures counterbalanced against the EmpaToM task. ³Measures taken after the EmpaToM task. ⁴Including the Trail Making Test A and B, Identical Pictures Test, Digit Span Backward Test, and Spot a Word Test.

EmpaToM

All three experiments used the validated EmpaToM task to measure empathy, compassion and ToM^{3,66}. Across several previous behavioral and fMRI studies, validation of the EmpaToM task's empathy, compassion, and ToM measures involved direct correlations and activation overlap with established empathy (Socio-affective Video Tasks⁶⁷) and ToM tasks (False Belief Task⁶⁸, Imposing Memory Task⁶⁹), as well as exhibiting overlap with previous meta-analytical findings^{67–70}.

The EmpaToM task utilizes a within-subject design and involves participants viewing a ~15 second video clip during each trial (48 in total), in which a narrator recounts an allegedly autobiographical experience with either neutral (e.g., moving into a new apartment) or negatively emotional content (e.g., relative with a disease). During each trial participants watched a video, after which they indicated the valence of their affective state on a continuous scale, with negative ratings after an emotionally negative narration indicating empathic sharing of the narrator's emotion (empathy measure). Another continuous scale on which participants indicated how much compassion they felt for the person in the video (compassion measure) directly followed this. Participants were then presented with a multiple-choice question (three choices, one correct answer) that required either ToM inference ("The person thinks that...") or factual-reasoning as a control condition ("It is correct that..."). Additionally in D3, participants reported their hypothetical willingness to help the narrator. However, due to the small sample size and, thus, insufficient statistical power, gender differences in this prosocial rating were not analysed for the current study. Before beginning the experiment, participants completed a practice round with instructions to familiarize themselves with the paradigm. The display timings of the rating bars and questions differed slightly between each dataset and are summarized in Table 2. For an illustrative example of the EmpaToM task, see Fig. 6 (note: informed consent was obtained to publish the individual's images depicted in the figure). For examples of the videos' content, see the supplementary material (S1: Example stories and questions for each experimental condition). Participants viewed 48 videos in total, across all datasets. The videos contained six different female and six different male narrators. The ordering of the videos were pseudo-randomized to control for frequency of male and female narrators, frequency of video's valence or question type, and frequently of the narrator. The version of the EmpaToM task used across the three datasets was programed in Presentation (<https://www.neurobs.com>) and took approximately 45 min to complete.

Procedure

The current study combines three summary datasets of the EmpaToM paradigm. The datasets were selected based on a review of the available, published online datasets for the EmpaToM task involving healthy adults. Three datasets fit this criteria and were available for download. The datasets were obtained from the online repository Open Science Framework. In Table 2, we summarize the datasets, including the additional measures that were not analysed in the current study. In each of the original investigations, participants completed the EmpaToM task alone, under laboratory conditions at the Technische Universität Dresden.

In brief, the procedures for the three original studies were as follows:

D1 (investigating age differences in social abilities): The experimental procedure involved two counterbalanced sessions in which participants either (i) completed the EmpaToM task during functional magnetic resonance imaging or (ii) undertook a battery of cognitive functioning tasks. At the beginning of the first session participants also completed a socio-demographic questionnaire⁶³. Participants were compensated €8.50 for their

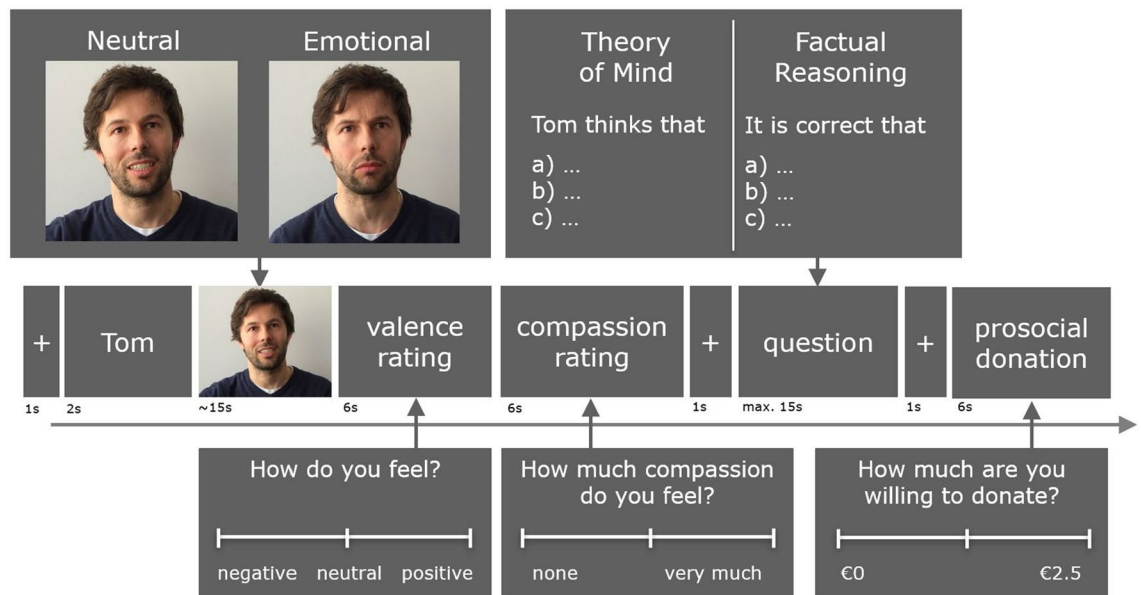


Figure 6. Example of an EmpaToM Task trial. The EmpaToM task utilizes a within-subject design. In each trial (48 in total), participants were shown either an emotionally negative or neutral video of a man or a woman describing an autobiographical experience. The content of the experience further required either factual-reasoning or ToM inference. Following the video, participants indicated on a continuous valence scale how they felt, that is, how much the participant shared the negative feelings expressed by the video's protagonist (empathy measure). This was followed by an additional continuous scale on which participants indicated how much compassion they felt (compassion measure). Participants were then presented with a multiple-choice question (three choices, only one correct answer) that required either ToM inference or factual reasoning. Last, participants indicated on a continuous scale how much money they were willing to donate to a charitable cause associated with the content of the video (prosocial donation measure, Study 2 only). Timings indicated are for EmpaToM version used for Study 2. Differences in timing between the dataset versions for Study 1 can be found in Table 2. Informed consent was obtained to publish the individual's images depicted in this figure.

time. The statistical analysis compared empathy, compassion, and ToM measures between younger and older adults in a mixed design.

D2 (investigating personality traits and social abilities): Participants were selected based on an online screening in which suitability for MRI scanning and demographics were assessed along with other questionnaires not reported in the original paper. The experimental procedure involved participants completing the EmpaToM task, combined with a subsequent 14-day ecological momentary assessment protocol on social interactions⁶⁴. Participants were compensated €120 for their time. Statistical analysis compared both behavioural and neuronal measures collected during the EmpaToM task with everyday social affect and social cognition measures collected during the ecological momentary assessment in a within-subject design.

D3 (investigating meditation practice, prosociality, and social abilities): The experimental procedure involved participants completing the EmpaToM task, with this data acquisition occurring at the start of a brief meditation training study⁵⁸. Participants were compensated €30 or course credit for their time. Lehmann and colleagues used this dataset to investigate statistical relationships between empathy, compassion, and theory of mind with the prosocial behaviour measure in a within-subject design.

Data

In the current study, we focus on the empathy, compassion and ToM measures of the EmpaToM task, comparing emotional with neutral videos and ToM with factual reasoning questions. Validation of these contrasts is described in detail by Kanske and colleagues³. The data used in the current study was summary data from the EmpaToM task. For each participant, an average was generated for each of the main measures. For empathy and compassion, this involved the average across the rating bar responses from the 24 emotional and 24 neutral videos. For the ToM measures, accuracy reflects the percentage of the 24 ToM and 24 factual reasoning questions correctly answered. RT is the corresponding average time taken to respond to these correctly answered ToM and factual reasoning questions.

Methods: study 2

Study 2 was conducted online. The Ethics Committee of Technische Universität Dresden (EK-98022021) approved this study. All methods were performed in accordance with the relevant guidelines and regulations. The power analysis for Study 2 was same as in Study 1. Mann–Whitney U and ANOVA analyses were performed using SPSS (IBM Corp., 2021, Version 28.0.1.0). Correlation analyses and graphic production were done using Graphpad Prism (<https://www.graphpad.com>, Version 9.3.1). Generalized linear model analysis was performed

using Stata (StataCorp, 2019, Version 15.1). The data and analysis scripts for Study 2 are available at https://osf.io/xgma6/?view_only=d0b7d3e1f10a415fa6c033678de5c9ff. The analysis for the current study was not preregistered.

Participants

The EmpaToM data used in the current study is derived from a separate investigation examining creativity and social behaviour. The pre-registration for this original study is available at <https://aspredicted.org/blind.php?x=2vk3a6>. Please note that the participant exclusion criteria for the current analyses does not deviate from the pre-registration for the original dataset investigating creativity and social behaviour. The original creativity study utilized a within-subject design.

Participants were recruited via Prolific (www.prolific.co) and tested in the context of an online study examining creativity and social abilities. To account for potential dropouts and other issues, 250 participants were invited to take part. Of these, 23 participants were excluded due to technical issues or not completing the experiment (7), using an unsupported browser or operating system (6), answering less than 33% of the questions correctly or answering on average in less than 1.5 s (6), limited or no rating bar movement (2), and experimenter error (2). In addition, multivariate outliers were screened for among social affect, social cognition, and demographic variables, generating Mahalanobis distance scores from a multiple linear regression. For the current dataset, there were 13 degrees of freedom, which equated to a critical Chi-square value of 34.529 (at $\alpha = 0.001$). The test revealed one multivariate outlier that was additionally removed from the analyses. Thus, 226 individuals (101 women) were included in the final dataset. A Mann–Whitney U test indicated that men (age mean = 30.02, SD = 7.75) and women (age mean = 32.59, SD = 12.53) did not differ significantly in age ($U = 6235$, $p = 0.87$). Participants reported their gender (man, woman, or diverse), age, and highest obtained level of education. None of the participants indicated that they identify as diverse. Highest obtained level of education was indicated based on 7 possible answers, conforming to the German education system, which was then concatenated into three categories: high school (German: Hauptschule, Realschule, Fachhochschule- oder allg. Hochschulreife), vocational training (German: Berufsausbildung), and university (German: Bachelor, Master, Promotion). In total, 14 participants reported high school education, 100 participants reported vocational training, and 112 participants reported a university education as their highest obtained level of education. Demographic data were collected via FormR⁷¹. Informed consent was obtained from participants prior to the experiment. Participants were debriefed as to the intention of the experiment at the end of data collection and received €12 reimbursement for their time.

EmpaToM

To allow for online testing, the original, offline version of the EmpaToM was translated into JavaScript using the PsychoJS library (<https://github.com/psychopy/psychojs>) and presented online via Pavlovia⁷². The online version of the EmpaToM takes approximately 40 min to complete. The empathy, compassion and ToM/factual reasoning measures of the EmpaToM task were the same in Study 2 as in Study 1 with the timing of each measure indicated in Fig. 1. In addition, participants in Study 2 answered a prosocial behaviour question integrated into the EmpaToM by combining it with the option to engage in a charitable donation (see^{73,27}). For the prosocial donation question, participants were endowed with €2.50 per trial. They could donate none, part, or this entire amount to a charity associated with each video. Specifically, during the instructions, participants were informed that a charity, thematically appropriate to the content of each video's narrative (e.g., *Red Cross* for a narrative discussing homelessness, *German Cancer Aid* for a narrative discussing a mother with cancer, etc.), had been selected and that they could donate up to €2.50 per trial to that given charity. Participants were then informed that a single trial would be randomly selected at the end of the experiment and the amount they chose to give during this single, random trial would be donated to the charity associated with the selected video. They were also told that the money they chose not to donate would be paid to them in addition to their compensation. Thus, the prosocial donation measure used in Study 2 was associated with a real-world cost for the participants.

Procedure

For Study 2, participants were invited online via Prolific to take part in a study exploring social behaviour. After initial invitation via Prolific, participants received a link to the FormR experiment page. The link was only valid with the access code and became invalid once a participant completed the study, thus limiting the possibility that participants would undertake the study twice. After accessing the link, participants were required to complete a series of questionnaires via FormR. These included a demographic questionnaire the Alternate Use Task and Instances Task⁷⁴, Inventory of Creative Activities and Achievements^{74,75}, Author Recognition Test⁷⁶, IQ measures (WAIS IV Matrix Reasoning Task and Mehrfachwahl-Wortschatz-Intelligenztest) and standard personality measures (Big Five Inventory and Interpersonal Reactivity Index). Only the demographic questionnaire was used in the current study. On finishing the final questionnaire, participants were then transferred to the Pavlovia experiment webpage where they undertook the online version of the EmpaToM paradigm. Participants were compensated with €12.50 for their time.

Data

The empathy, compassion and ToM measures used in Study 2 are same as described in the Data section in Study 1. For the prosocial donations measure, the average donation for each participant across the 24 emotional and 24 neutral videos was used.

Data availability

The data that support the findings of this study are openly available at the Open Science Framework: *Study 1* Dataset 1: <https://doi.org/10.17605/OSF.IO/7EDBN>. Dataset 2 <https://osf.io/f7zcr/>. Dataset 3 <https://osf.io/>

tu2gj/?view_only=f285308c035d451ca2fce5f3788f97e3. Study 2 (including full dataset and analysis scripts for both studies). https://osf.io/xgma6/?view_only=d0b7d3e1f10a415fa6c033678de5c9ff.

Received: 7 June 2023; Accepted: 17 November 2023

Published online: 25 November 2023

References

- Klimecki, O. M., Leiberg, S., Ricard, M. & Singer, T. Differential pattern of functional brain plasticity after compassion and empathy training. *Soc. Cogn. Affect. Neurosci.* **9**, 873–879 (2014).
- de Vignemont, F. & Singer, T. The empathic brain: How, when and why?. *Trends Cogn. Sci.* **10**, 435–441 (2006).
- Kanske, P., Böckler, A., Trautwein, F.-M. & Singer, T. Dissecting the social brain: Introducing the EmpaToM to reveal distinct neural networks and brain–behavior relations for empathy and Theory of Mind. *NeuroImage* **122**, 6–19 (2015).
- Frith, C. D. & Frith, U. The neural basis of mentalizing. *Neuron* **50**, 531–534 (2006).
- Kanske, P. The social mind: Disentangling affective and cognitive routes to understanding others. *Interdiscip. Sci. Rev.* **43**, 115–124 (2018).
- Stietz, J., Jauk, E., Krach, S. & Kanske, P. Dissociating empathy from perspective-taking: Evidence from intra- and inter-individual differences research. *Front. Psychiatry* <https://doi.org/10.3389/fpsy.2019.00126> (2019).
- McClure, E. B. A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. *Psychol. Bull.* **126**, 424–453 (2000).
- Baez, S. *et al.* Men, women...who cares? A population-based study on sex differences and gender roles in empathy and moral cognition. *PLOS ONE* **12**, e0179336 (2017).
- Di Tella, M., Miti, F., Ardito, R. B. & Adenzato, M. Social cognition and sex: Are men and women really different?. *Personal. Individ. Differ.* **162**, 110045 (2020).
- Salazar, L. R. The relationship between compassion, interpersonal communication apprehension, narcissism and verbal aggressiveness. *J. Happiness Well-Being* **4**, 1–14 (2016).
- Baron-Cohen, S. *et al.* The “reading the mind in the eyes” test: Complete absence of typical sex difference in ~400 men and women with autism. *PLOS ONE* **10**, e0136521 (2015).
- Fischer, A. L., O'Rourke, N. & Loken Thornton, W. Age differences in cognitive and affective theory of mind: Concurrent contributions of neurocognitive performance, sex, and pulse pressure. *J. Gerontol. Ser. B* **72**, 71–81 (2017).
- Krach, S. *et al.* Are women better mindreaders? Sex differences in neural correlates of mentalizing detected with functional MRI. *BMC Neurosci.* **10**, 9 (2009).
- Bradley, M. M., Codispoti, M., Sabatinelli, D. & Lang, P. J. Emotion and motivation II: Sex differences in picture processing. *Emotion* **1**, 300–319 (2001).
- Kempton, M. J. *et al.* The effects of gender and COMT Val158Met polymorphism on fearful facial affect recognition: A fMRI study. *Int. J. Neuropsychopharmacol.* **12**, 371–381 (2009).
- Rahman, Q., Wilson, G. D. & Abrahams, S. Sex, sexual orientation, and identification of positive and negative facial affect. *Brain Cogn.* **54**, 179–185 (2004).
- Barrett, L. F., Robin, L., Pietromonaco, P. R. & Eyssell, K. M. Are women the “more emotional” sex? Evidence from emotional experiences in social context. *Cogn. Emot.* **12**, 555–578 (1998).
- Derntl, B. *et al.* Multidimensional assessment of empathic abilities: Neural correlates and gender differences. *Psychoneuroendocrinology* **35**, 67–82 (2010).
- Grimshaw, G. M., Bulman-Fleming, M. B. & Ngo, C. A signal-detection analysis of sex differences in the perception of emotional faces. *Brain Cogn.* **54**, 248–250 (2004).
- Eagly, A. H. The his and hers of prosocial behavior: An examination of the social psychology of gender. *Am. Psychol.* **64**, 644–658 (2009).
- Rand, D. G., Brescoll, V. L., Everett, J. A. C., Capraro, V. & Barcelo, H. Social heuristics and social roles: Intuition favors altruism for women but not for men. *J. Exp. Psychol. Gen.* **145**, 389–396 (2016).
- Soutschek, A. *et al.* The dopaminergic reward system underpins gender differences in social preferences. *Nat. Hum. Behav.* **1**, 819–827 (2017).
- Eagly, A. H. & Crowley, M. Gender and helping behavior: A meta-analytic review of the social psychological literature. *Psychol. Bull.* **100**, 283–308 (1986).
- Johnson, R. C. *et al.* Cross-cultural assessment of altruism and its correlates. *Personal. Individ. Differ.* **10**, 855–868 (1989).
- Kanske, P., Böckler, A., Trautwein, F.-M., Parianen Lesemann, F. H. & Singer, T. Are strong empathizers better mentalizers? Evidence for independence and interaction between the routes of social cognition. *Soc. Cogn. Affect. Neurosci.* **11**, 1383–1392 (2016).
- Schurz, M. *et al.* Toward a hierarchical model of social cognition: A neuroimaging meta-analysis and integrative review of empathy and theory of mind. *Psychol. Bull.* **147**, 293–327 (2021).
- Tusche, A., Böckler, A., Kanske, P., Trautwein, F.-M. & Singer, T. Decoding the charitable brain: Empathy, perspective taking, and attention shifts differentially predict altruistic giving. *J. Neurosci.* **36**, 4719–4732 (2016).
- Reiter, A. M. F., Kanske, P., Eppinger, B. & Li, S.-C. The aging of the social mind—Differential effects on components of social understanding. *Sci. Rep.* **7**, 11046 (2017).
- Myers, L. & Sirois, M. J. Spearman Correlation Coefficients, Differences between. In *Encyclopedia of Statistical Sciences* (eds Myers, L. & Sirois, M. J.) (Wiley, 2006) <https://doi.org/10.1002/0471667196.ess5050.pub2>.
- Blundell, R., Dearden, L. & Sianesi, B. Evaluating the effect of education on earnings: Models, methods and results from the national child development survey. *J. R. Stat. Soc. Ser. A Stat. Soc.* **168**, 473–512 (2005).
- Arnold, T. W. Uninformative parameters and model selection using akaike's information criterion. *J. Wildl. Manag.* **74**, 1175–1178 (2010).
- Geary, D. C. Sexual selection and sex differences in social cognition. In *Biology, Society, and Behavior: The Development of Sex Differences in Cognition* (ed. Geary, D. C.) 23–53 (Ablex Publishing, 2002).
- Eagly, A. H., Wood, W. & Diekmann, A. B. Social role theory of sex differences and similarities: A current appraisal. In *The Developmental Social Psychology of Gender* (eds Eagly, A. H. *et al.*) 123–174 (Lawrence Erlbaum Associates Publishers, 2000).
- Thomas, G. & Maio, G. R. Man, I feel like a woman: When and how gender-role motivation helps mind-reading. *J. Pers. Soc. Psychol.* **95**, 1165–1179 (2008).
- Eliot, L. The trouble with sex differences. *Neuron* **72**, 895–898 (2011).
- Eliot, L., Ahmed, A., Khan, H. & Patel, J. Dump the “dimorphism”: Comprehensive synthesis of human brain studies reveals few male-female differences beyond size. *Neurosci. Biobehav. Rev.* **125**, 667–697 (2021).
- Hirnstain, M. & Hausmann, M. Sex/gender differences in the brain are not trivial—A commentary on Eliot *et al.* (2021). *Neurosci. Biobehav. Rev.* **130**, 408–409 (2021).
- McCarthy, M. M. Multifaceted origins of sex differences in the brain. *Philos. Trans. R. Soc. B Biol. Sci.* **371**, 20150106 (2016).

39. Eagly, A. H. & Wood, W. The nature-nurture debates: 25 years of challenges in understanding the psychology of gender. *Perspect. Psychol. Sci.* **8**, 340–357 (2013).
40. Lausen, A. & Schacht, A. Gender differences in the recognition of vocal emotions. *Front. Psychol.* <https://doi.org/10.3389/fpsyg.2018.00882> (2018).
41. Thompson, A. E. & Voyer, D. Sex differences in the ability to recognise non-verbal displays of emotion: A meta-analysis. *Cogn. Emot.* **28**, 1164–1195 (2014).
42. Charman, T., Ruffman, T. & Clements, W. Is there a gender difference in false belief development?. *Soc. Dev.* **11**, 1–10 (2002).
43. Walker, S. Gender differences in the relationship between young children's peer-related social competence and individual differences in theory of mind. *J. Genet. Psychol.* **166**, 297–312 (2005).
44. Calero, C., Salles, A., Semelman, M. & Sigman, M. Age and gender dependent development of Theory of Mind in 6- to 8-years old children. *Front. Hum. Neurosci.* <https://doi.org/10.3389/fnhum.2013.00281> (2013).
45. Xia, H., Wu, N. & Su, Y. Investigating the genetic basis of theory of mind (ToM): The role of Catechol-O-Methyltransferase (COMT) gene polymorphisms. *PLOS ONE* **7**, e49768 (2012).
46. Greenberg, D. M. *et al.* Sex and age differences in “theory of mind” across 57 countries using the English version of the “Reading the Mind in the Eyes” Test. *Proc. Natl. Acad. Sci.* **120**, e2022385119 (2023).
47. Frank, C. K., Baron-Cohen, S. & Ganzel, B. L. Sex differences in the neural basis of false-belief and pragmatic language comprehension. *NeuroImage* **105**, 300–311 (2015).
48. Navarra-Ventura, G. *et al.* Gender differences in social cognition: A cross-sectional pilot study of recently diagnosed patients with Schizophrenia and healthy subjects. *Can. J. Psychiatry* **63**, 538–546 (2018).
49. Sullivan, S. & Ruffman, T. Social understanding: How does it fare with advancing years?. *Br. J. Psychol.* **95**, 1–18 (2004).
50. Miller, D. I. & Halpern, D. F. The new science of cognitive sex differences. *Trends Cogn. Sci.* **18**, 37–45 (2014).
51. Halpern, D. F. & LaMay, M. L. The smarter sex: A critical review of sex differences in intelligence. *Educ. Psychol. Rev.* **12**, 229–246 (2000).
52. Ilescu, D., Ilie, A., Ispas, D., Dobrean, A. & Clinciu, A. I. Sex differences in intelligence: A multi-measure approach using nationally representative samples from Romania. *Intelligence* **58**, 54–61 (2016).
53. de Waal, F. B. M. Putting the altruism back into altruism: The evolution of empathy. *Annu. Rev. Psychol.* **59**, 279–300 (2008).
54. Eisenberg, N. & Fabes, R. A. Empathy: Conceptualization, measurement, and relation to prosocial behavior. *Motiv. Emot.* **14**, 131–149 (1990).
55. Eisenberg, N. & Miller, P. A. The relation of empathy to prosocial and related behaviors. *Psychol. Bull.* **101**, 91–119 (1987).
56. Bethlehem, R. A. I. *et al.* Does empathy predict altruism in the wild?. *Soc. Neurosci.* **12**, 743–750 (2017).
57. Klimecki, O. M., Mayer, S. V., Jusyte, A., Scheeff, J. & Schönenberg, M. Empathy promotes altruistic behavior in economic interactions. *Sci. Rep.* **6**, 31961 (2016).
58. Lehmann, K., Böckler, A., Klimecki, O., Müller-Liebmann, C. & Kanske, P. Empathy and correct mental state inferences both promote prosociality. *Sci. Rep.* **12**, 16979 (2022).
59. Singer, T. & Klimecki, O. M. Empathy and compassion. *Curr. Biol.* **24**, R875–R878 (2014).
60. Frijda, N. H., Kuipers, P. & ter Schure, E. Relations among emotion, appraisal, and emotional action readiness. *J. Pers. Soc. Psychol.* **57**, 212–228 (1989).
61. Höfler, M., Scherbaum, S., Kanske, P., McDonald, B. & Miller, R. Means to valuable exploration: I. The blending of confirmation and exploration and how to resolve it. *Meta-Psychology* <https://doi.org/10.15626/MP.2021.2837> (2022).
62. Fujita, F., Diener, E. & Sandvik, E. Gender differences in negative affect and well-being: The case for emotional intensity. *J. Pers. Soc. Psychol.* **61**, 427–434 (1991).
63. Stietz, J. *et al.* The ageing of the social mind: Replicating the preservation of socio-affective and the decline of sociocognitive-processes in old age. *R. Soc. Open Sci.* **8**, 210641 (2021).
64. Hildebrandt, M. K., Jauk, E., Lehmann, K., Maliske, L. & Kanske, P. Brain activation during social cognition predicts everyday perspective-taking: A combined fMRI and ecological momentary assessment study of the social brain. *NeuroImage* **227**, 117624 (2021).
65. Faul, F., Erdfelder, E., Lang, A.-G. & Buchner, A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* **39**, 175–191 (2007).
66. Tholen, M. G., Trautwein, F.-M., Böckler, A., Singer, T. & Kanske, P. Functional magnetic resonance imaging (fMRI) item analysis of empathy and theory of mind. *Hum. Brain Mapp.* **41**, 2611–2628 (2020).
67. Klimecki, O. M., Leiberg, S., Lamm, C. & Singer, T. Functional Neural Plasticity and Associated Changes in Positive Affect After Compassion Training. *Cereb. Cortex* **23**, 1552–1561 (2013).
68. Dodell-Feder, D., Koster-Hale, J., Bedny, M. & Saxe, R. fMRI item analysis in a theory of mind task. *NeuroImage* **55**, 705–712 (2011).
69. Kinderman, P., Dunbar, R. & Bentall, R. P. Theory-of-mind deficits and causal attributions. *Br. J. Psychol.* **89**, 191–204 (1998).
70. Bzdok, D. *et al.* Parsing the neural correlates of moral cognition: ALE meta-analysis on morality, theory of mind, and empathy. *Brain Struct. Funct.* **217**, 783–796 (2012).
71. Arslan, R. C., Walther, M. P. & Tata, C. S. formr: A study framework allowing for automated feedback generation and complex longitudinal experience-sampling studies using R. *Behav. Res. Methods* **52**, 376–387 (2020).
72. Peirce, J. *et al.* PsychoPy2: Experiments in behavior made easy. *Behav. Res. Methods* **51**, 195–203 (2019).
73. McDonald, B., Böckler, A. & Kanske, P. Soundtrack to the social world: Emotional music enhances empathy, compassion, and prosocial decisions but not theory of mind. *Emotion* **22**, 19–29 (2022).
74. Jauk, E., Benedek, M. & Neubauer, A. C. The Road to Creative Achievement: A Latent Variable Model of Ability and Personality Predictors. *Eur. J. Personal.* **28**, 95–105 (2014).
75. Diedrich, J. *et al.* Assessment of real-life creativity: The Inventory of Creative Activities and Achievements (ICAA). *Psychol. Aesthet. Creat. Arts* **12**, 304–316 (2018).
76. Mar, R. A., Oatley, K., Hirsh, J., dela Paz, J. & Peterson, J. B. Bookworms versus nerds: Exposure to fiction versus non-fiction, divergent associations with social ability, and the simulation of fictional social worlds. *J. Res. Personal.* **40**, 694–712 (2006).

Acknowledgements

The authors would like to thank Dr. Michael Höfler for his assistance with the generalized linear model of the prosocial data.

Author contributions

B.M. and P.K. contributed to the conception and writing of this work. B.M. prepared the draft for submission. Both authors contributed to the article and approved the submitted version.

Funding

Open Access funding enabled and organized by Projekt DEAL. This work was funded by Deutsche Forschungsgemeinschaft, (KA 4412/2-1, KA 4412/4-1, KA 4412/5-1, KA 4412/9-1, CRC 940/CO7, IRTG 2773/P4).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-47747-9>.

Correspondence and requests for materials should be addressed to B.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023