Current Research in Ecological and Social Psychology xxx (xxxx) 100108

Contents lists available at ScienceDirect

Current Research in Ecological and Social Psychology

journal homepage: www.elsevier.com

Listening in to a conversation enhances theory of mind

Emanuele Castano [a,*], Alison Jane Martingano [b], Gabriana Basile [c], Elly Bergen [d], Evelyn Hye Kyung Jeong [a]

[a] Department of Psychology and Cognitive Science, University of Trento, Italy
[b] Department of Psychology, University of Wisconsin, Green Bay, WI, United States of America
[c] Department of Psychology, The New School for Social Research, NY, United States of America
[d] New York Law School, NY, United States of America

ARTICLE INFO

Keywords:
Theory of mind
Listening in
Mentalizing
Conversation
Social cognition

ABSTRACT

Activities that require mentalizing, such as meditating or reading literary fiction, have been shown to enhance Theory of Mind (ToM) (e.g., Tan, Lo, and Macrae, 2014; Kidd & Castano, 2013). In this article, we conjecture that relatively greater mentalizing effort also occurs when individuals can only listen in on a conversation, compared to those who actively participate in it. Therefore, those who listen in should show better performance on subsequent ToM tasks. Participants (N = 77) were divided into triads and randomly assigned to be Interlocutors (a Director and a Matcher) or Listeners. In each triad, Interlocutors completed a collaborative figure matching task while talking to each other, while Listeners completed the same task while listening to live audio of the Interlocutors' conversation. All participants then completed two Theory of Mind measures. Multivariate analyses show that Listeners outperformed Interlocutors on both measures, but the pattern is significant only for one measure. These results complement existing theorizing and findings regarding the potential benefit of participating in activities that train ToM and may also help explain the often-observed stronger ToM performance of those (e.g., women) who, at least in certain contexts, have typically been discouraged from participating actively in conversations.

Introduction

Humans' capacity to reason about the minds of other humans is thought to be unparalleled in the animal world. It is also considered to be a key factor in the evolutionary success of Homo sapiens over their Neanderthal cousins because it greatly enhanced our capacity for cooperation (Culotta, 2010). Such a capacity to reason about the minds of others is known as mentalizing or Theory of Mind (ToM).

In its most basic form, ToM is the simple recognition that others have a mind that differs from our own, a question of some importance in the animal-human comparative work where the concept of ToM first originated (Premack and Woodruff, 1978). For our purposes here, however, what is of greater interest are advanced forms of ToM, which involve higher-level and the integration of cognitive processes (Goldman, 2006), and reflect the capacity to accurately infer and represent other people's thoughts, emotions, and complex mental states such as beliefs and intentions (Heyes and Frith, 2014). By underpinning social, emotional, and collective intelligence (Mayer et al., 2012; Vonk and Pitzen, 2017; Woolley et al., 2010), advanced forms of ToM aid social functioning.

Given the centrality of ToM processes for humans, it is unsurprising that we come into the world soft-wired for it. Newborns appear to be predisposed to develop social abilities related to ToM (e.g., Happé and Frith, 2013), and there is growing evidence that ToM processes are supported by a network of specific brain regions (Frith and Frith, 2006). As in all matters human, however, the development and deployment of ToM processes are influenced by experience.

Activities that require mentalizing enhance tom

Researchers have only recently started investigating what kind of experiences enhance ToM processes in neurotypical adults. Tan et al. (2014) found that a brief mindful meditation exercise enhances performance on one of the most commonly used tests of ToM, the Reading the Mind in the Eyes Test (RMET; Baron-Cohen et al., 2001). The purported mechanism is that meditation requires focusing on one's own bodily and psychological states, thus increasing interoceptive awareness. From

* Corresponding author at: Department of Psychology and Cognitive Science, University of Trento, Corso Bettini, 31 – 38068, Rovereto, Italy.
E-mail addresses: emanuele.castano@unitn.it, emanuele.castano@gmail.com (E. Castano).

https://doi.org/10.1016/j.cresp.2023.100108
2666-6227/

this perspective, meditation is essentially an exercise in self-directed mentalization, which then translates into greater performance on other-directed mentalization. More recent research, using longer training and a different measure of ToM, provided converging evidence that mental training that focuses on metacognition and perspective-taking can also improve ToM, through paths other than interoceptive awareness (Trautwein et al., 2020), and that such changes go hand in hand with changes in brain morphology in areas known to support ToM processes (Valk et al., 2017).

Other, seemingly unrelated tasks may also yield greater ToM by allowing people to practice mentalizing. For example, habitually reading narrative fiction is positively correlated with ToM performance (e.g., Castano et al., 2020; Kidd and Castano, 2017; Mar et al., 2006). This relationship may be a result of the fact that narrative fiction focuses on the intricacies of human personal and social relations. In order for readers to understand the plot, motivations, and relationships between fictional characters, they have to understand the inner life of characters — probably through a process of simulation (Gallese and Goldman, 1998; Oatley, 1999). Experimental evidence supports this causal hypothesis. In a series of experiments, Kidd and Castano (2013) assigned participants to read short stories (or excerpts of novels) categorized as either literary fiction or popular fiction, and subsequently assessed participants' performance on a variety of ToM tests, including the RMET. Results showed that reading literary (but not popular) fiction resulted in stronger performance on these measures (see also Kidd et al., 2016; Kidd and Castano, 2019; Pino and Mazza, 2016; Schwerin and Lenhart, 2022; van Kuijk, Verkoeijen, Dijkstra, and Zwaan, 2018). Kidd & Castano (2013) proposed that while popular fiction confirms pre-existing schemas, for instance in terms of social categories and personality types, by presenting characters who are coherent and transparent, the obscure characters of literary fiction defy easy categorization and understanding, and force the reader to engage in mentalizing. A similar point is made by cognitive literary theorist Lisa Zunshine: While all fiction requires understanding characters' embedded mental states, literary fiction "make[s] the reader infer implied mental states in addition to (and sometimes instead of) spelling some out" (Zunshine, 2019; p. 5).

Consistent with research on exposure to fiction in adults, other research has demonstrated a similar mechanism in children. Peskin and Astington (2004) read children fictional stories that varied, depending on the condition, with respect to the presence vs. absence of metacognitive language. Unexpectedly, yet consistent with the causal mechanism proposed here, it was the children who were read fiction without metacognitive language that subsequently scored higher on ToM tasks. In other words, having to infer characters' thoughts and feelings for themselves, instead of being told what these were (via metacognitive language), strengthened their mentalizing skills (Peskin and Astington, 2004).

Another set of findings emerging from an experiment focusing on film is also consistent with the idea that, when it comes to mental states language and ToM development, *less is more*. Building on theorizing developed for written fiction (Kidd and Castano, 2013), Castano (2021) reasoned that art films, similarly to literary fiction, imply the characters' mental states, while Hollywood films make them more explicit, and even "observable" to viewers through exaggerated facial expressions and close-ups. Consistent with this rationale, it was found that, compared to those who watched Hollywood films, participants who watched art films subsequently performed better on two measures of ToM.

Finally, other studies found enhanced ToM performance after showing participants award-winning TV dramas as opposed to documentaries (Black and Barnes, 2015b), by reading a series of metaphorical as opposed to non-metaphorical short stories (Bowes and Katz, 2015), and by asking participants to focus on the narrative aspect while playing a video game — as compared to no specific instructions or instructions to ignore the story (Bormann and Greitemeyer, 2015).

The research reviewed above suggests that the effect on ToM performance resulting from engaging in certain kinds of experiences (meditation, reading literary fiction, watching art films, and watching award-winning TV drama) is due to the fact that these experiences require greater mentalizing than the control tasks given in each experiment. Other mundane experiences also vary in the degree to which they require mentalizing. One of these experiences is conversation.

Listening in requires (Greater) mentalizing

It can be safely said that most, if not all, conversations require some mentalizing. The degree of mentalizing, however, can vary significantly depending on a host of factors, such as heterogeneity in prior knowledge of the topic or the characteristics and personal history of the participants in the conversation (Clark, 1985; Lewis, 1979; Stalnaker, 1978). The amount of required mentalizing effort can also be a function of the context in which the interaction occurs. Imagine an expert being interviewed for a podcast. The interviewer and expert are discussing a new product the expert has recently designed. The expert is drawing parallels to other well-known products, using analogies and detailed descriptions. Being actively engaged in the conversation, the interviewer conveys his level of understanding through non-verbal and verbal utterances (e.g., "Oh yes, I see!" or "Huh?"). As a consequence, the expert adjusts and adapts what she is saying. The interviewer can even improve his understanding of what the expert intends to say by asking direct questions. Back and forth questioning allows interlocutors to share the burden of achieving mutual understanding by relying on their partner to fill in any gaps in understanding (Clark and Wilkes-Gibbs, 1986; Goffman, 1976).

Now imagine the people who are listening to this podcast at home. They do not have this benefit: they cannot convey their understanding of the conversation, or lack thereof, nor can they ask for clarification. Furthermore, while the interlocutors only have to track each other's minds, listeners have to track the minds of both interlocutors. It should be stressed that having the same goal as the interlocutors means that listeners are not passively overhearing a conversation: they are required to engage in a difficult task. When it comes to developing an understanding of the topic, listeners are at a disadvantage (Schober and Clark, 1989). However, they may be benefiting from their status in a broader sense. We hypothesize that because they have to deploy their mentalizing abilities to a greater extent than interlocutors, individuals who can only listen in to a conversation will subsequently perform better than interlocutors on ToM tasks.

The current experiment

To test this hypothesis, we borrowed a tool from psycholinguistics, the referential communication task (Krauss and Weinheimer, 1966; as modified by Schober and Clark, 1989). This task typically involves groups of three people, each given a different role. Two are Interlocutors (in the role of Director and Matcher), and one is a Listener. In each triad, the Director gives instructions to the Matcher about the position of abstract (Tangram) figures within a grid (see Fig. 1) to get the Matcher to arrange six figures in a particular order. Interlocutors complete this task while talking to each other from different sides of a visual barrier. Listeners listen to the audio of this conversation and attempt to complete the same matching exercise as Matchers. Consistent with the rationale proposed above, research has shown that Listeners perform more poorly than Interlocutors on this task (Schober and Clark, 1989). We made use of the same paradigm, and we expect (1) to replicate the finding that Listeners perform more poorly than Interlocutors, and (2), most importantly, that Listeners will perform better than Interlocutors on ToM measures collected subsequently.

ToM is a complex, multifaceted construct that encompasses conceptual knowledge, cognitive processes, and social competence/motiva-

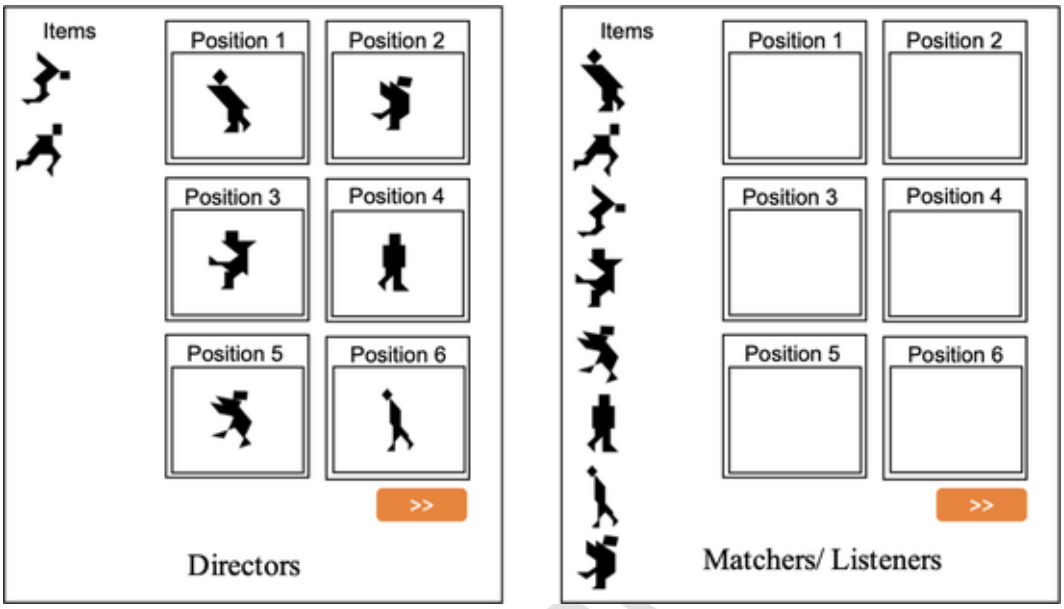

Fig. 1. Referential communication task hosted on Qualtrics™ platform.

tion (Apperly, 2012), and it has affective and cognitive dimensions (Apperly and Butterfill, 2009; Frith and Frith, 2008), which are captured by different measures. In this study, we thus used two measures of ToM: one, the Reading the Mind in the Eyes test (Baron-Cohen et al., 2001), which is considered more socio-perceptual and affective (Schurz et al., 2014, 2021); and one, the False Belief Understanding (Rowe et al., 2001), which is considered more socio-cognitive (Meinhardt-Injac et al., 2018).

Method

Participants

86 participants were recruited from the NYC area via craigslist.org and compensated $10 for participation. Participants were required to be strangers to each other. Upon arriving at the laboratory, they were randomly assigned to one of three roles (Matcher, Director, or Listener) by having them choose one of three folded slips of paper with the (hidden) word "matcher," "director," or "listener." Although we sought to have three participants per session, to avoid not being able to have the minimum of three participants, when possible, we over-recruited, scheduling 4–5 participants. In a few cases, this resulted in having two listeners, instead of only one. In those cases, for random assignment, we had four folded slips of paper, two of which had the word "listener." Listeners were in separate rooms and completed the task independently, listening to the interaction between Directors and Matchers. Nine participants were excluded from the analyses. Eight were Listeners and were excluded for the following reasons: microphone not working ($N = 1$); headphones not working ($N = 2$); did not follow instructions and went ahead of Interlocutors in their task completion ($N = 3$); took a phone call and rushed out ($N = 1$); withdrew from the study midway ($N = 1$). One was a Matcher, excluded because they went ahead of the Director in guessing. This did not affect the Director's action, but while it could have affected the Listener, this is moot since this specific Listener was one of the eight that were excluded on their own (de)merit. Although every session had at least one Listener, the final sample that we analyzed comprises 6 cases in which there was no Listener corresponding to the Director and Matcher. The decision not to delete the entire triad when one individual needed to be excluded follows from the fact that we are not looking at processes that emerge from the interaction of three individuals, but rather at the effect (on social cognition) that assuming a specific role in the task has on the individual assuming that role. The final sample was thus of 77 (27 Directors, 26 Matchers, and 24 Listeners). The sample consisted of 37 women and 40 men with the following racial breakdown: 34% Black, 32% White, 14% Asian, 12% Latin, 8% other. 60% of participants had a college/graduate-level education. The mean age was 39 years ($SD = 12$; *range* = 19–63). Informed consent for experimentation was obtained from all human participants, and the research was conducted in accordance with the Declaration of Helsinki. This research study was approved by the institutional review board of the New School for Social Research (protocol #2017–1012).

Procedure

Interlocutors sat at individual computers on different sides of a visual barrier and completed four rounds of the Tangram matching task. In each round, Directors were presented with six Tangram figures arranged in a 3 × 2 grid labeled 1 to 6, with two unused Tangram figures displayed next to the grid. Matchers were presented with eight Tangram figures alongside an empty 3 × 2 grid labeled 1 to 6 (see Fig. 1). Qualtrics software was used to present the material and carry out the task. The Director gave instructions to the Matcher, guiding them to place the figures into the grid as quickly and accurately as possible. Matchers and Directors were instructed to talk with each other freely. Listeners sat in a separate room and listened to a live audio stream of the Interlocutors' conversation and completed the same matching of the Tangram figures, using the same Qualtrics interface as Matchers. All participants then completed a series of tasks. While other measures were collected for different purposes, here we present the two ToM measures that were collected. No other measures were ToM measures and thus are not relevant to the hypothesis tested here.

Measures

Figure matching score

A score was computed as the percentage of correct matches over the four trials.

Reading the mind in the eyes (RMET)

This task asks participants to identify the emotion/mental state that people are expressing in 36 pictures of the eye region by choosing one

out of four options that describe emotions and other mental states, which differ from one picture to the next. For instance, one photo of the eye region is presented along with the four options joking, insisting, amused, relaxed, while another one has the four options serious, ashamed, bewildered, alarmed. For each of the 36 trials, only one response is considered correct (Baron-Cohen et al., 2001). Responses are coded 0 when wrong and 1 when correct, and a sum is computed. High scores indicate that participants can correctly infer the mental states of others. Participants can score from 0 to 36.

False belief understanding (FBU)

Participants completed a test of understanding false beliefs, which is considered a cognitive measure of ToM. While the false belief test is typically associated with research with children, whose ToM is still developing, the specific version used here has been developed for, and used with, adults (Rowe et al., 2001). In the present study, the scenario we used (taken from Rowe et al., 2001; p. 616) had participants listen to a short story in which the two protagonists are Richard and Ann, a couple engaged in the renovation of their spare bedroom. In the story, Richard begins the redecorating while Ann is at work, and he leaves a message with Ann's colleagues in which he informs her that he has begun the redecorating and asks her to call him. Ann gets the message, but she is in a rush and does not call Richard. When she arrives home, Richard says, "You're going to be surprised when I tell you what I've been doing this afternoon." Participants are then asked why Richard says this. Scoring for correct answers followed the same rules as Rowe et al. (2001). To be correct, participants had to show more than simply a superficial first-order false belief, such as "Richard doesn't know Anne has gotten his message" (A doesn't know Y). Instead, participants have to show they understand the rationale behind the belief, "Richard thinks Anne doesn't know he's had the afternoon off" (A thinks B doesn't know X; a second-order false belief). In the study by Rowe et al. (2001), what they define as normal adults responded correctly to the second-order false belief approximately 65–70% of the time. This FBU task was chosen because it has been used in previous studies that look at factors that may improve ToM performance in adults (Pino et al., 2017), and because it complements the more socio-perceptual, affective measure that is the RMET (Schurz et al., 2014, 2021). The FBU is considered more of a socio-cognitive measure of ToM (Meinhardt-Injac et al., 2018). Together, these two tasks capture the multifaceted, complex construct of ToM.

Results

SAS (SAS Institute Inc., 2014) was used for statistical analyses. We conducted a post-hoc power analysis to determine our achieved level of statistical power. Our hypothesis did not differentiate between our two measures of ToM, so we carried out our power analysis using a composite score obtained by averaging the two ToM measures (after standardization). In G*power (Faul et al., 2007), we used a T-Test, Post hoc estimator, $\alpha = 0.05$, with the actual effect size obtained by comparing Interlocutors and Listeners on the composite score (Cohen's $d = 0.74$). This revealed 0.84 power. First, we checked that men and women were equally distributed across the three roles, $\chi 2$ (2, $N = 77$) = 2.84, $p = .24$. Second, we looked at the performance on the figure matching task ($M = 80\%$; $SD = 23.82$). Since Directors do not have a performance score, the comparisons are only between Matchers and Listeners. It was found that while Matchers ($M = 85.41\%$) scored higher than Listeners ($M = 74.13\%$), the difference was not significant, $F(1, 48) = 2.91$, $p = .09$, $\eta^2 = 0.06$. Third, we tested the effects on ToM measures, namely the RMET ($M = 24$, $SD = 4.5$) and the FBU ($M = 0.52$, $SD = 0.50$). For FBU, three independent coders of participants' responses came to a high level of agreement in deciding whether the response indicated second-order ToM understanding (Fleiss' kappa = 0.84), and any disagreements were resolved by consensus. Since our hypothesis was that Listeners perform better than Interlocutors (Directors and Matchers) on the ToM measures, we report this specific contrast, followed by the omnibus F for the main effect of Role.

Our rationale did not differ for the two measures of ToM. Thus, we first conducted a MANOVA test (DVs: FBU and RMET). The specific contrast was significant, $F(2, 73) = 4.48$, $p = .01$ [omnibus effect: $F(4, 146) = 2.19$, $p = .07$]. Means are reported in Table 1, and graphically represented in Fig. 2. Looking at the two ToM measures separately revealed a significant effect for RMET, $F(1, 74) = 6.69$, $p = .01$, $\eta^2 = 0.04$ (omnibus $F = 3.36$, $p = .04$) but a non-significant one for FBU, $F(1, 74) = 3.02$, $p = .08$, $\eta^2 = 0.03$ (omnibus $F = 1.52$, $p = .22$). When to the same analysis we added gender as a covariate, the effects described above did not change: significant effects remained significant and non-significant effects remained non-significant. The MANOVA test for gender was not significant, $F(2, 70) = 1.90$, $p = .15$. Looking at the two ToM measures separately revealed a significant effect of gender for the RMET ($M = 25.05$ vs. 23.02; $t(71) = 1.96$, $p = .05$) but not for the FBU ($M = 0.54$ vs. 0.50; $t(71) = 0.22$, *n.s.*).

An alternative way to assess whether Role had the same effect on the two ToM measures consists of testing a model which included the Interlocutors vs. Listeners contrast, a within-participants factor with two levels (RMET vs. FBU), and their interaction. We thus standardized RMET and FBU scores and carried out this analysis. The specific contrast testing our hypothesis was significant, $F(1, 75) = 8.78$, $p = .01$ [omnibus effect: $F(2, 74) = 4.35$, $p = .02$], but the interaction effect was not significant. The absence of an interaction conveys that the hypothesized effect is not significantly different for the RMET and FBU. See Table 1 and Fig. 2 for scores.

Because we had the same hypothesis for both ToM measures, we reported above the results from multivariate analysis of variance, even though FBU is a dichotomous variable, the most appropriate analytical strategy for which is a frequency analysis. This revealed that while only 45% of the Interlocutors (44% and 46% for Directors and Matchers, respectively) responded correctly, 67% of the Listeners did, $\chi 2$ (1, $N = 77$) = 3.03, $p = .08$.

Discussion

Research suggests that Theory of Mind (ToM) abilities can be boosted through experiences that require individuals to exert mentalizing effort (e.g., Kidd and Castano, 2013). Most research to date has focused on experiences with the fictional worlds of novels, films, video games, and TV series. Here, we propose that ToM can be boosted by mundane communication experiences that also require mentalizing effort. Specifically, we hypothesized that compared to Interlocutors (i.e., people who talk) in a conversation, Listeners (i.e., people who can only listen to it) are required to exert greater mentalizing effort, which then results in stronger performance on ToM tasks.

To test this hypothesis, we relied on the referential communication task, in which two Interlocutors can communicate to solve a task, while a third person, the Listener, has to solve the same tasks but can only listen to the conversation without being able to partake in it. Early research (Schober and Clark, 1989) found that Listeners performed more

Table 1
RMET and FBU scores, as a function of Role, and for the whole sample.

	Interlocutors		Overhearers ($N = 24$)	Whole Sample ($N = 77$)
	Directors ($N = 27$)	Matchers ($N = 26$)		
RMET, mean (SD)	**23.04 (4.56)**	**23.23 (4.80)**	**25.92 (3.60)**	**24.06 (4.32)**
FBU,% correct responses	**44%**	**46%**	**67%**	**52%**

Note. RMET = Reading the Mind in the Eyes Test; FBU = False Belief Understanding.

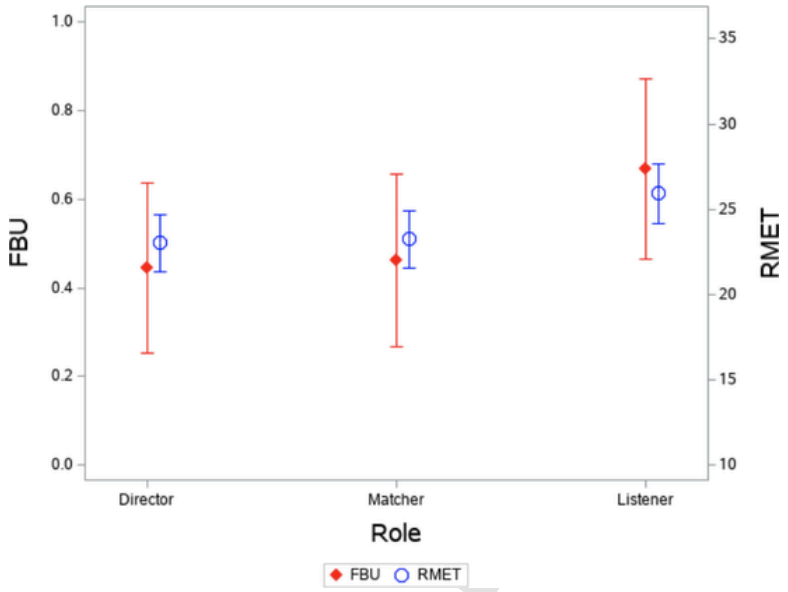

Fig. 2. FBU (False Belief Understanding) and RMET (Reading the Mind in the Eyes Test) mean scores (and standard errors) as a function of the role of the participant.

poorly than Matchers on the Figure Matching Task. Here, the means are in the same direction, but the effect is not significant. At least two reasons can explain this difference. First, the set of figures used here was chosen in collaboration with one of the authors of the Schober and Clark (1989) article to be more difficult than the set used in the original article. Consistently with this choice, both Matchers and Listeners perform more poorly in the present study than in the original one. Second, the present study has half as many Listeners as in the study by Schober and Clark (1989). It should be noted, however, that the level of difficulty of the task, within limits of not being overly simple or impossible to solve (which could lead to frustration), does not have any obvious relevance for the goal of the present experiment, which is to investigate the hypothesis that listeners would perform better than interlocutors on subsequent ToM tasks.

The advantage on ToM measures expected for Listeners, compared to Interlocutors, emerged clearly for the RMET but less so for the FBU. The significant MANOVA effect (which specifically tests the effect over multiple measures, as opposed to a single one) indicates that the pattern is consistent for both measures of ToM. Furthermore, the mixed model analysis revealed a significant effect of the contrast testing the hypothesis (and of the omnibus effect of Role), but no interaction with the within-participants factor, thus also suggesting that the effect was similar across the two measures. Yet, since analysis of variance is not the most appropriate analytical strategy for the FBU data, this should be interpreted with sufficient skepticism and together with the results emerging from the Chi-square analysis for the FBU variable. Overall, therefore, Listeners displayed better ToM than Interlocutors when this is assessed via the RMET, while for the FBU results are unclear.

One could ask whether the listening experience is actually training ToM ability (that is, enhancing an individual's ability in a real sense) or whether it is priming or activating ToM capacities that the individual already possesses. This is an interesting question, which is at times discussed in experiments that report manipulation effects on ToM measures (e.g., Castano, 2021). Our experiment is not set up to answer such a question, and, in fact, we are not sure that this is a question that can easily find an empirical answer. But we offer some relevant speculation below, based on Apperly's (2012) proposition that ToM can and should be understood as tripartite: as conceptual knowledge, cognitive processes, and social competence/motivation. It seems unreasonable to suggest that the results presented here, and those obtained in previous studies that find effects of experimental manipulations on ToM performance, could be due to the expansion of conceptual knowledge. It is, however, possible that exerting mentalizing effort acts as a prime, making mind-related concepts more accessible. This enhanced accessibility may thus translate into greater performance on ToM tasks. The manipulation used in this and other experiments may also serve as a sort of procedural or mindset priming (Taylor and Gollwitzer, 1995), activating not only mind concepts but also the set of cognitive processes that are part and parcel of ToM (see Apperly, 2012). From a neuroscience perspective, these experiences may be activating the specific network of brain areas that research has shown to be involved when we process the contents of another person's mind (Saxe and Kanwisher, 2013). Such activation would result in a facilitation on subsequent, unrelated tasks assessing ToM, like those that have been used in the present experiment.

Repeated exposure to literary fiction, art films, or award-winning TV drama, or frequently finding oneself in the position of having to make inferences about other people's minds without the possibility of directly asking about them (as in our experiment), may result in a chronically higher likelihood of deployment of ToM processes. This relates to the third aspect of ToM proposed by Apperly (2012), social competence or motivation, which refers to the "propensity for ToM that is not reducible to […] cognitive capacity for ToM or to […] motivation in general." (p. 835). Just as evidence of the association between lifetime exposure to literary fiction and ToM (e.g., Kidd and Castano, 2017) complements experimental findings that brief exposure may enhance ToM performance (e.g., Kidd and Castano, 2013), it is possible that an association exists between a lifetime of listening in and ToM. That is, individuals who are more likely to find themselves regularly in a position to listen, but not contribute, to conversations, should possess ToM skills to a greater extent.

One finding in support of this conjecture is that people with less power are more likely to engage in perspective-taking, a close cousin to ToM (see Galinsky et al., 2016, for a review). Other convergent evidence comes from studies showing that compared to men, women are less talkative (Leaper and Ayres, 2007), more empathic (e.g., Baron-Cohen, 2010; McClure, 2000), and generally perform better on ToM tasks (e.g., Baron-Cohen and Hammer, 1997; McClure, 2000). In the current experiment, we also found evidence for this advantage of women over men on the RMET - although not for the FBU. This result is consistent with previous findings that women perform better than men on the RMET (e.g., Dorris et al., 2022) but only at a very young age on the FBU (Charman et al., 2002). While the exact nature of gender differences on ToM is contested and may be the consequence of a host of factors, including hormone levels (e.g., van Honk et al., 2011), such differ-

ences may also be in part due to differences in social experiences and roles. For instance, women may be more likely to find themselves in positions where they need to listen, understand, and empathize with others, which in turn could contribute to their stronger ToM abilities.

The research presented here has various limitations. First, it is a single study, which clearly needs to be replicated before strong conclusions can be drawn. Complementary evidence could come from a neuroimaging study in which Interlocutors' and Listeners' activation of the ToM brain network is compared. Second, although we obtained a diverse sample of Americans, our sample was from a WEIRD (Western, Educated, Industrialized, Rich, and Democratic) society (Henrich et al., 2010), which limits its representativeness (Rad et al., 2018). To the extent that ToM skills differ by culture (Lillard, 1998), it would be unwise to generalize outside of this population without further research in other cultural contexts.

Future studies using other ToM measures, and particularly testing different types of conversations, are also needed. We used the referential communication paradigm because in it, Interlocutors and Listeners have the same goal and, likely, the same motivation to pursue such a goal. This allows for a meaningful comparison between the two groups, in a setting in which people are equally motivated to understand what the other(s) are thinking. Conversations vary, however, in the extent to which this is the case. We would expect that overhearing a conversation in which Interlocutors' reciprocal mentalizing is not as important/present, may not activate ToM processes in the Listener any more than it does in the Interlocutors, and perhaps even less. However, there are regular day-to-day experiences wherein people find themselves motivated to listen in, such as when attending group meetings, panel discussions, listening to interviews, and various interpersonal settings where multiple people are present. Our finding is likely limited to these settings. That is, settings wherein those listening in find themselves either practically or socially unable to actively participate, but are still motivated to mentalize to understand the conversation. Furthermore, it would be of interest to investigate whether overhearing a conversation among several individuals, as opposed to only a pair, may result in an even stronger ToM advantage. Such investigations would also allow us to shed light on whether, aside from a general improvement in performance in ToM used here, reading about or listening to multiple people's perspectives may enhance the ability to embed mental representations inside other mental representations — e.g., "Sam thought that Henry believed that Sam wanted to buy a Tax Disc" (Kinderman et al., 1998; p. 204). This experiment would also help elucidate what mediates the effect on ToM reported here. Initially, our rationale revolved primarily around the idea that listeners have to mentalize more because they cannot partake in the conversation, asking questions and clarifications. Another, non-mutually-exclusive possibility, is that the effect is due to the fact that listeners have to track two minds, while each interlocutor has to track only one. There is, of course, an upper limit to the number of minds that an individual can track at once, but evidence that listening to three or four individuals speaking results in greater mentalizing effort than listening to only two would support this conjecture.

In conclusion, notwithstanding the limitations discussed above, the experiment presented here provides, to our knowledge, the first evidence of the beneficial effects of listening in on Theory of Mind. While ToM is important for referential communication (Sidera et al., 2019) and explicitly asking one's partner to clarify their position may afford a better understanding of the topic at hand (Eyal et al., 2018), these findings suggest that there is a benefit to exerting mentalizing effort. Listening to a conversation, more so than taking part in it, appears to sharpen this all-important social cognition skill.

Uncited references

, .

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

Apperly, I.A., 2012. What is “theory of mind”? Concepts, cognitive processes and individual differences. Q. J. Exp. Psychol. 65 (5), 825–839. https://doi.org/10.1080/17470218.2012.676055.

Apperly, I.A., Butterfill, S.A., 2009. Do humans have two-systems to track beliefs and belief-like states? Psychol. Rev. 116, 953–970. https://doi.org/10.1037/a0016923.

Baron-Cohen, S., 2010. Empathizing, systemizing, and the extreme male brain theory of autism. Prog. Brain Res. 186, 167–175.

Baron-Cohen, S., Hammer, J., 1997. Is autism an extreme form of the” male brain”? Adv. Infancy Res. 11, 193–218.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., Plumb, I., 2001. The “Reading the Mind in the Eyes” Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. The J. Child Psychol. Psychiatry Allied Discipl. 42 (2), 241–251. https://doi.org/10.1111/1469-7610.0071.

Black, J.E., Barnes, J.L., 2015. The effects of reading material on social and non-social cognition. Poetics 52, 32–43.

Bormann, D., Greitemeyer, T., 2015. Immersed in virtual worlds and minds: effects of in-game storytelling on immersion, need satisfaction, and affective theory of mind. Soc. Psychol. Personal. Sci. 6 (6), 646–652.

Bowes, A., Katz, A., 2015. Metaphor creates intimacy and temporarily enhances theory of mind. Mem. Cognit. 43 (6), 953–963. https://doi.org/10.3758/s13421-015-0508-4.

Castano, E., 2021. Art films foster theory of mind. Human. Soc. Sci. Commun. 8 (1). https://doi.org/10.1057/s41599-021-00793-y.

Castano, E., Martingano, A.J., Perconti, P., 2020. The effect of exposure to fiction on attributional complexity, egocentric bias and accuracy in social perception. PLoS One 15 (5), e0233378. https://doi.org/10.1371/journal.pone.0233378.

Charman, T., Ruffman, T., Clements, W., 2002. Is there a gender difference in false belief development? Soc. Dev. 11 (1), 1–10.

Clark, H.H., 1985. Language use and language users. In: Lindsey, G., Aronson, E. (Eds.), Handbook of Social Psychology. Random House, New York, pp. 179–232. Vol. 2.

Clark, H.H., Wilkes-Gibbs, D., 1986. Referring as a collaborative process. Cognition 22, 1–39. https://doi.org/10.1016/0010-0277(86)90010-7.

Culotta, E., 2010. Did Modern Humans Get Smart Or Just Get Together? Science 328 (5975), 164. https://doi.org/10.1126/science.328.5975.164. . –164. .

Dorris, L., Young, D., Barlow, J., Byrne, K., Hoyle, R., 2022. Cognitive empathy across the lifespan. Developmental Medicine & Child Neurology 64 (12), 1524–1531.

Eyal, T., Steffel, M., Epley, N., 2018. Perspective mistaking: accurately understanding the mind of another requires getting perspective, not taking perspective. J. Pers. Soc. Psychol. 114 (4), 547–571. https://doi.org/10.1037/pspa0000115.

Faul, F., Erdfelder, E., Lang, A.-.G., Buchner, A, 2007. G*Power 3: a flexible statistical power analysis for the social, behavioral, and biomedical sciences. Behav. Res. Methods 39, 175–191.

Frith, C.D., Frith, U., 2006. The neural basis of mentalizing. Neuron 50, 531–534.

Frith, C.D., Frith, U., 2008. Implicit and explicit processes in social cognition. Neuron 60, 503–510. https://doi.org/10.1016/j.neuron.2008.10.032.

Galinsky, A.D., Rucker, D.D., Magee, J.C., 2016. Power and perspective-taking: a critical examination. J. Exp. Soc. Psychol. 67, 91–92.

Goffman, E., 1976. Replies and responses. Language Soc. 5 (3), 257–313. https://doi.org/10.1017/S0047404500007156.

Gallese, V., Goldman, A., 1998. Mirror neurons and the simulation theory of mind-reading. Trends Cogn. Sci. 2, 493–501.

Goldman, A., 2006. Simulating minds: The philosophy, psychology, and Neuroscience of Mindreading. Oxford University Press, Oxford, UK.

Happé, F., Frith, U., 2013. Annual research review: towards a developmental neuroscience of atypical social cognition. J. Child Psychol. Psychiatry 55 (6), 553–577. https://doi.org/10.1111/jcpp.12162.

Henrich, J., Heine, S.J., Norenzayan, A., 2010. Most people are not WEIRD. Nature 466 (7302), 29. https://doi.org/10.1038/466029a. . –29. .

Heyes, C.M., Frith, C.D., 2014. The cultural evolution of mind reading. Science 344 (6190), 1243091. https://doi.org/10.1126/science.1243091.

Kidd, D.C., Castano, E., 2013. Reading literary fiction improves theory of mind. Science 342 (6156), 377–380. https://doi.org/10.1126/science.1239918.

Kidd, D.C., Castano, E., 2017. Different stories: how levels of familiarity with literary and genre fiction relate to mentalizing. Psychol. Aesthet. Creat. Arts 11 (4), 474–486. https://doi.org/10.1037/aca0000069.

Kidd, D., Castano, E., 2019. Reading literary fiction and theory of mind: three preregistered replications and extensions of Kidd and Castano (2013). Soc. Psychol. Personal. Sci. 10 (4), 522–531. https://doi.org/10.1177/2F1948550618775410.

Kidd, D., Ongis, M., Castano, E., 2016. On literary fiction and its effects on theory of mind. Sci. Study Lit. 6 (1), 42–58. https://doi.org/10.1075/ssol.6.1.04kid.

Kinderman, P., Dunbar, R., Bentall, R.P., 1998. Theory-of-mind deficits and causal

attributions. Br. J. Psychol. 89 (2), 191–204.
Krauss, R.M., Weinheimer, S., 1966. Concurrent feedback, confirmation, and the encoding of referents in verbal communication. J. Pers. Soc. Psychol. 4 (3), 343.
Leaper, C., Ayres, M.M., 2007. A meta-analytic review of gender variations in adults' language use: talkativeness, affiliative speech, and assertive speech. Pers. Soc. Psychol. Rev. 11 (4), 328–363. https://doi.org/10.1177/1088868307302221.
Lewis, D., 1979. Scorekeeping in a language game. In: Bäuerle, R., Egli, U., von Stechow, A. (Eds.), Semantics from Different Points of View. Springer Series in Language and Communication. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-67458-7_12. vol 6.
Lillard, A., 1998. Ethnopsychologies: cultural variations in theories of mind. Psychol. Bull. 123 (1), 3–32. https://doi.org/10.1037/0033-2909.123.1.3.
Mar, R.A., Oatley, K., Hirsh, J., dela Paz, J., Peterson, J.B., 2006. 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 (5), 694–712. https://doi.org/10.1016/j.jrp.2005.08.002.
Mayer, J.D., Salovey, P., Caruso, D.R., 2012. The Validity of the MSCEIT: additional analyses and evidence. Emotion Rev. 4 (4), 403–408. https://doi.org/10.1177/1754073912445815.
McClure, E.B., 2000. A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. Psychol. Bull. 126 (3), 424.
Meinhardt-Injac, B., Daum, M.M., Meinhardt, G., Persike, M., 2018. The two-systems account of theory of mind: testing the links to social-perceptual and cognitive abilities. Front. Hum. Neurosci. 12, 25.
Oatley, K., 1999. Why fiction may be twice as true as fact: fiction as cognitive and emotional simulation. Rev. Gen. Psychol. 3 (2), 101–117.
Peskin, J., Astington, J.W., 2004. The effects of adding metacognitive language to story texts. Cogn. Dev. 19 (2), 253–273. https://doi.org/10.1016/j.cogdev.2004.01.003.
Pino, M.C., Mazza, M., 2016. The use of "literary fiction" to promote mentalizing ability. PLoS One 11 (8), e0160254.
Pino, M.C., Mazza, M., Mariano, M., Peretti, S., Dimitriou, D., Masedu, F., Valenti, M., Franco, F., 2017. Simple mindreading abilities predict complex theory of mind: developmental delay in autism spectrum disorders. J. Autism Dev. Disord. 47 (9), 2743–2756. https://doi.org/10.1007/s10803-017-3194-1.
Premack, D., Woodruff, G., 1978. Does the chimpanzee have a theory of mind? Behav. Brain Sci. 1 (4), 515–526. https://doi.org/10.1017/S0140525X00076512.
Rad, M.S., Martingano, A.J., Ginges, J., 2018. Toward a psychology ofHomo sapiens: making psychological science more representative of the human population. Proc. Natl Acad. Sci. 115 (45), 11401–11405. https://doi.org/10.1073/pnas.1721165115.
Rowe, A.D., Bullock, P.R., Polkey, C.E., Morris, R.G., 2001. Theory of mind impairments and their relationship to executive functioning following frontal lobe excisions. Brain 124 (3), 600–616. https://doi.org/10.1093/brain/124.5.1062.
SAS Institute Inc, 2014. SAS® OnDemand for Academics: User's Guide. SAS Institute Inc, Cary, NC.
Saxe, R., Kanwisher, N., 2013. People thinking about thinking people: the role of the temporo-parietal junction in "theory of mind, Social Neuroscience. Psychology Press, pp. 171–182.
Schober, M.F., Clark, H.H., 1989. Understanding by addressees and overhearers. Cogn. Psychol. 21 (2), 211–232. https://doi.org/10.1016/0010-0285(89)90008-X.
Schurz, M., Radua, J., Aichhorn, M., Richlan, F., Perner, J., 2014. Fractionating theory of mind: a meta-analysis of functional brain imaging studies. Neurosci. Biobehav. Rev. 42, 9–34.
Schurz, M., Radua, J., Tholen, M.G., Maliske, L., Margulies, D.S., Mars, R.B., Kanske, P., 2021. Toward a hierarchical model of social cognition: a neuroimaging meta-analysis and integrative review of empathy and theory of mind. Psychol. Bull. 147 (3), 293.
Schwerin, J., Lenhart, J., 2022. The effects of literariness on social-cognitive skills: examining narrative engagement, transportation, and identification as moderators. Psychology of Aesthetics, Creativity, and the Arts.
Sidera, F., Perpiñà, G., Serrano, J., Rostan, C., 2019. Why is theory of mind important for referential communication? Curr. Psychol. 37 (1), 82–97. https://doi.org/10.1007/s12144-016-9492-5.
Stalnaker, R.C., 1978. Assertion. In: Cole, P. (Ed.), Pragmatics. Brill. https://doi.org/10.1163/9789004368873_013.
Tan, L.B., Lo, B.C., Macrae, C.N., 2014. Brief Mindfulness Meditation Improves Mental State Attribution and Empathizing. PLoS One 9 (10), e110510. https://doi.org/10.1371/journal.pone.0110510.
Taylor, S., Gollwitzer, P., 1995. Effects of mindset on positive illusions. J. Pers. Soc. Psychol. 69 (2), 213–226.
Trautwein, F.M., Kanske, P., Böckler, A., Singer, T., 2020. Differential benefits of mental training types for attention, compassion, and theory of mind. Cognition 194, 104039.
Van Honk, J., Schutter, D.J., Bos, P.A., Kruijt, A.W., Lentjes, E.G., Baron-Cohen, S., 2011. Testosterone administration impairs cognitive empathy in women depending on second-to-fourth digit ratio. Proc. Natl Acad. Sci. 108 (8), 3448–3452.
Van Kuijk, I., Verkoeijen, P., Dijkstra, K., Zwaan, R.A., Vazire, S., Vul, E., 2018. The effect of reading a short passage of literary fiction on theory of mind: a replication of Kidd and Castano (2013). Collabra: Psychology 4 (1).
Valk, S.L., Bernhardt, B.C., Trautwein, F.M., Böckler, A., Kanske, P., Guizard, N., Singer, T., 2017. Structural plasticity of the social brain: differential change after socio-affective and cognitive mental training. Sci. Adv. 3 (10), e1700489.
Vonk, J., Pitzen, J., 2017. Believing in other minds: accurate mentalizing does not predict religiosity. Pers. Individ. Dif. 115, 70–76. https://doi.org/10.1016/j.paid.2016.06.008.
Woolley, A.W., Chabris, C.F., Pentland, A., Hashmi, N., Malone, T.W., 2010. Evidence for a Collective Intelligence Factor in the Performance of Human Groups. Science 330 (6004), 686–688. https://doi.org/10.1126/science.1193147.
Zunshine, L., 2019. What Mary Poppins knew: theory of mind, children's literature, history. Narrative 27 (1), 1–29. https://doi.org/10.1353/nar.2019.0000.