Supplemental Materials

Supplementary Methods

Strange Stories: This task consists of 24 short vignettes, each accompanied by a picture. The set of “Strange Stories” includes 12 types of stories depicting everyday situations where people say something that includes a non-literal meaning such as “Lie,” “White lie,” “Joke,” “Pretend,” “Misunderstanding,” “Persuade,” “Appearance/Reality,” “Figure of Speech,” “Sarcasm,” “Forget,” “Double Bluff,” and “Contrary Emotions” and two examples of each story type. Participants were asked to answer questions about the intention of the character described in each vignette. The percentage of correct answers was used as a measure of mentalizing abilities in this analysis.

Faux Pas test: The adult version of the Faux Pas test was developed by Stone et al. based on the child version of the Faux Pas test to assess pragmatics in children with autism. Like the Strange Stories task, the Faux Pas test consists of 10 short stories containing the occurrence of a faux pas, someone saying something awkward. After reading each story, participants were asked to answer a series of questions about the detection of the faux pas (e.g., “Did anyone say something awkward?”), understanding of the faux pas (e.g., “Who said something they shouldn’t have said”), and understanding mental state of other person (e.g., “Why did they say it?”). In a later session, researchers also included a question about empathic understanding (e.g., “How do you think X person felt?”). The percentage of correct answers was used as a measure of mentalizing abilities in this analysis.

Reading the Mind in the Eyes test (the Eyes): In the Eyes test, participants are asked to choose the best word that describes what the person in the photo was thinking or feeling following a forced-choice procedure. By including a glossary of all the mental state terms, the Eyes test is designed to reduce the demand of language comprehensive abilities. In this meta-analysis, both the original and revised versions of the Eyes test
were included. The number of correct responses made by participants was the dependent measure for this task.

2.4. Statistical analyses

To detect potential publication bias, first, we drew a funnel plot graphically. A funnel plot is a scatter plot of effect sizes against study size, which is used to detect bias or systematic heterogeneity. It is plotted with effect size on the X axis and the sample size or variance on the Y axis. In a symmetrical plot illustrating no publication bias, large studies appear toward the top of the graph and cluster around the mean effect size; small studies appear at the bottom and exhibit the expected greater variation. An asymmetrical funnel shape may indicate a sign of either publication bias or a systematic difference between smaller and larger studies. In cases where we found an asymmetrical funnel plot, the source of funnel plot asymmetry was further analyzed using Egger's regression test at \( \alpha = .05 \). Furthermore, we calculated a Fail Safe number (the number of missing studies necessary to make the group difference insignificant). Therefore, this approach provides adjusted effect size estimates for the funnel plot asymmetry.

Symptom severity estimation. To equate symptom severity in SCZ across studies, we rescaled the values of symptom scores and standardized across all studies on a scale from 0 (no symptoms) to 1 (maximal symptoms on a given scale) based on prior studies. To obtain this score, we divided the mean score reported for a sample by the maximum possible score on the scale (with an adjustment for scales with a minimum possible scores of 1 versus 0). To be specific, we calculated separate scores for (1) positive symptoms using Positive and Negative Syndrome Scale (PANSS) positive score or Scale for the Assessment of Positive Symptoms (SAPS) score, (2) negative symptoms using PANSS negative score or Scale for the Assessment of Negative
Symptoms (SANS \textsuperscript{12}) score, and (3) overall pathology using the Brief Psychiatric Rating Scale (BPRS) \textsuperscript{13} or PANSS total score. \textsuperscript{10} If none of these total scores were available, we used the average of the SAPS \textsuperscript{11} and SANS scores, \textsuperscript{12} average of PANSS positive and PANSS negative scores, or PANSS general psychopathology scale \textsuperscript{10} (see Table 5).

3. Results

Through the initial search, 43 studies were considered for inclusion. Of these, 6 studies were excluded, after reviewing the abstracts, for the following reasons: (a) studies did not include HC group (e.g., \textsuperscript{14,15}), (b) information could not be obtained about $M \ (SD)$ even after contacting the authors (e.g., \textsuperscript{16,17,18}), (c) a study that included the same participants that were reported in another study (e.g., \textsuperscript{4}). The final sample consisted of 37 studies.

Of these 37 studies, two included both individuals with SCZ and those with ASD (e.g., \textsuperscript{19,20}). Nine studies included both visual and cognitive-linguistic mentalizing tasks \textsuperscript{21-29}. In two studies, Baron-Cohen et al., \textsuperscript{4,30} one publication report \textsuperscript{4} was excluded as the data concerning the same participants had been reported in another study. \textsuperscript{30} Because demographic characteristics including mean age, percentage of males, mean score on full IQ scores, and the Eyes test between two studies were the same, we assumed the same patient samples had been used in those studies.
References:


