



# Adaptation of the virtual assessment of mentalizing ability and evaluation of its utility and psychometric properties in Chinese individuals on the schizophrenia spectrum

Yuan Cao<sup>1,6</sup>, Winnie W. Y. So<sup>2,6</sup>, Yi Wang<sup>3,4</sup>, Ding-ding Hu<sup>3,4</sup>, Yi Xie<sup>3,4</sup>, Jie Gu<sup>3,4</sup>, Raymond C. K. Chan<sup>3,4</sup> and David H. K. Shum<sup>2,5</sup>✉

Schizophrenia is associated with impairments in theory of mind (ToM), the ability to understand and attribute mental states. However, the nature of the deficits across ToM subconstructs (viz., first- and second-order cognitive ToM, and first- and second-order affective ToM) remains unclear, partly due to assessment issues. The current study aimed to first adapt an ecologically valid ToM assessment tool, namely, the Virtual Assessment of Mentalizing Ability (VAMA), for use among healthy Chinese individuals (Study 1). We then compared 39 schizophrenia patients with 37 controls, and 48 individuals with high social anhedonia (SA) with 54 individuals with low SA (Study 2) using the adapted version of VAMA and Yoni Task. Results of Study 1 showed that the adapted VAMA demonstrated acceptable reliability (item-total correlation and test-retest reliability for total score,  $r = 0.731$ ,  $p_{FDR} < 0.01$ ) and construct validity (main effect of Order and Type). In Study 2, schizophrenia patients performed significantly worse than controls, making more “hypermentalizing” and “no mentalizing” ToM errors. Moreover, hypermentalizing error in the patient group was found to be significantly associated with their negative symptoms ( $r = 0.388$ ,  $p_{FDR} < 0.05$ ). Interestingly, the VAMA results differed from those of the Yoni Task, possibly due to differences in task complexity. Finally, individuals with high SA showed impairments in second-order cognitive ToM compared to the controls. Overall, our findings suggest that the VAMA can be adapted for use in China, and is sensitive to ToM impairments in clinical and at-risk groups. Limitations on the psychometric properties were discussed.

*Schizophrenia* (2025)11:43; <https://doi.org/10.1038/s41537-025-00594-1>

## INTRODUCTION

Theory of mind (ToM), the ability to attribute mental states (emotion, intention, and thoughts) to other individuals, is useful in explaining and predicting human behavior<sup>1</sup>. It is a multidimensional construct classified according to types (cognitive vs. affective) and orders (first vs. second)<sup>2,3</sup>. Affective ToM (inferring others' emotions) requires more personal engagement than cognitive ToM (inferring others' thoughts and intentions)<sup>2,4</sup>. Second-order ToM (reasoning of how another person thinks about a third person's mental states) requires high-level cognitive processes and is more complex than first-order ToM (inferring the thoughts of another person)<sup>3,4</sup>.

ToM impairment has been found in people with schizophrenia spectrum disorders<sup>5</sup>. Schizophrenia is a serious chronic psychiatric disorder with a global prevalence of 1%<sup>6</sup>. During social interactions, people with schizophrenia commonly present difficulties with emotion recognition, ToM, and empathy<sup>5</sup>. In particular, accumulating evidence has consistently shown that ToM is impaired across different stages of schizophrenia spectrum disorders, including in at-risk individuals with schizotypal traits or social anhedonia (SA) (SA, a trait whereby individuals are unable to take pleasure from social and/or interpersonal relationships)<sup>7–13</sup>. Notably, negative symptoms, such as SA and asociality are strongly associated with ToM impairment compared with other symptoms of schizophrenia<sup>14,15</sup>. Thibaut et al.<sup>14</sup> suggested that clinical symptoms affect how an individual perceives and interprets information, thereby affecting the underlying mechanisms of ToM. Research<sup>16,17</sup> has demonstrated

that individuals with high levels of SA tend to score significantly lower on the Hinting Task, a mentalizing measurement that assesses participants' understanding of indirect speech<sup>18</sup>, but inconsistent results<sup>19</sup> have been observed when using the Reading Mind in the Eyes Test, another mentalizing assessment tool that requires the participants to infer a person's emotional status by studying the photographs of their eyes<sup>20</sup>.

In recent years, researchers have attempted to use a more sophisticated approach to assess ToM<sup>21,22</sup>. Frith introduced different types of ToM errors, i.e., “hypermentalizing” (attributing a mental state when none is present), “reduced ToM” (failing to identify a mental state), and “no ToM” (a complete absence of mental inference)<sup>20</sup>. These distinctions are crucial, as evidence suggests significant associations between error types and clinical symptoms<sup>4,22–24</sup>. However, most ToM instruments do not assess the four subconstructs and error types of ToM. In addition, researchers have questioned the validity of the classic measures of ToM, particularly in ecological validity<sup>1,25</sup>, making it challenging to understand the level of ToM impairment in schizophrenia patients or individuals with high SA.

The Virtual Assessment of Mentalizing Ability (VAMA) is an ecological ToM assessment tool recently developed in Australia. It was designed to assess the four subconstructs of ToM separately and included error scores for the three types of ToM errors<sup>26</sup>. The VAMA showed acceptable to satisfactory reliability and validity (sensitive to ToM impairment in people with early and chronic schizophrenia)<sup>4,26,27</sup>. However, considering the cultural specificity of social cognitive processes, the VAMA, which was developed for

<sup>1</sup>Department of Social Work and Social Administration, The University of Hong Kong, Hong Kong, China. <sup>2</sup>Department of Rehabilitation Sciences, The Hong Kong Polytechnic University, Hong Kong, China. <sup>3</sup>Neuropsychology and Applied Cognitive Neuroscience Laboratory, CAS Key Laboratory of Mental Health, Institute of Psychology, Chinese Academy of Sciences, Beijing, China. <sup>4</sup>Department of Psychology, University of Chinese Academy of Sciences, Beijing, China. <sup>5</sup>Mental Health Research Center, The Hong Kong Polytechnic University, Hong Kong, China. <sup>6</sup>These authors contributed equally: Yuan Cao, Winnie W. Y. So. ✉email: david.shum@polyu.edu.hk

The primary aim of this study was to adapt and validate VAMA for use in the Chinese context. We hypothesized that the adapted version would demonstrate psychometric properties comparable to the original Australian version<sup>26</sup>. Specifically, we hypothesized this version would have: 1a) satisfactory test–retest reliability, 1b) comparable mean accuracy of the four subconstructs with the original version, 1c) strong correlation between items and their corresponding subscales to ensure internal consistency, 1d) no floor nor ceiling effects, 1e) significant main effects of Order and Type, where better performance on first-order than second-order ToM and cognitive ToM better than affective ToM (construct validity) will be shown. For Study 2, we hypothesized that schizophrenia patients would 2a) show poorer performance in the subconstructs of VAMA, 2b) make more errors in all three types of ToM errors than the healthy controls and 2c) we expected that associations would be found between schizophrenia clinical symptoms and the error types of VAMA.) For the individuals with a high level of SA, we also hypothesized that 2d) they would show poorer performance in the subconstructs of VAMA and 2e) make more errors in all three types of ToM errors than those with low SA.

**Participants.** A total of 100 healthy young participants were recruited in Beijing, China ( $M_{\text{age}} = 23.49$  years,  $SD = 1.13$ , range = 22–28;  $M_{\text{years of education}} = 16.85$  years,  $SD = 0.82$ , range = 13–19; 50% male, see Table 1). The exclusion criteria were as follows: (1) a medical history of brain injury or psychiatric or neurological disorders; (2) a history of substance abuse or addiction; (3) a family history of psychiatric or neurological disorder; and (4) inability to read or understand Chinese. Participants were recruited from the general community through convenience sampling and snowballing techniques. Of the 100 participants, 30 were randomly selected to participate in re-testing. The participants received RMB110 after completing the study. This study was approved by the Human Subjects Ethics Committee of the Institute of Psychology, Chinese Academy of Sciences in Beijing (reference number: H20045). Written informed consent was obtained from each participant before they enrolled in the study.

*The adapted version of the VAMA:* The VAMA version used in this study is an adaptation of the original VAMA<sup>26</sup> for use in the Chinese context. While keeping the characters' basic needs, framework, and task administration the same as in the original VAMA, we decided to change the perspective from the "first-person" to a "third-person" (i.e., outside observer) and replace the virtual game environment navigation with two-dimensional (2D) non-immersive virtual reality (VR) videos for two reasons: (1) to be more in line with the mechanism of ToM. Vogeley and his colleagues<sup>32</sup> investigated the neural mechanisms underlying first-person perspective and third-person perspective. Although common activations for both perspectives were found in the occipital, parietal, and prefrontal areas, the "third-person

**Table 1.** Demographic characteristics and clinical information for samples in Studies 1 and 2.

Table 1. Demographic characteristics and clinical information for samples in Studies 1 and 2.												
Study 1		Study 2										
		Sample 1: Clinical and Healthy Control					Sample 2: High and Low SA					
Healthy Young Adults (N = 100)		Individuals with Schizophrenia (N = 39)	Healthy Matched Controlled (N = 37)	$t/\chi^2$	p	Cohen's d	Low SA (N = 54)	High SA (N = 48)	$t/\chi^2$	p	Cohen's d	
Male/Female	50/50	5/34	8/29	1.04	0.31	/	22/32	25/23	1.32	0.25	/	
Age (SD)	23.49 (1.13)	29.13 (9.11)	27.57 (8.04)	0.79	0.43	0.18	19.81 (0.70)	19.83 (0.72)	0.131	0.90	0.03	
Range	22–28	18–51	19–50				18–21	18–21				
Education (years)	16.85 (0.82)	14.10 (3.07)	16.03 (2.14)	3.19**	0.002	0.72	14.57 (0.50)	14.65 (0.53)	0.71	0.48	0.14	
Range	13–19	8–20	12–21				14–15	14–16				
CSAS							5.65 (2.64)	23.06 (5.15)	21.11	0.000	4.34	
Estimated IQ (SD)	/	109.79 (28.83)	139.80 (20.03)	5.29***	<0.001	1.20						
Illness duration (years)	/	5.937 (6.04)	/									
PANSS_Negative	/	20.29 (7.14)	/									
PANSS_Positive	/	14.61(6.48)	/									
PANSS_General	/	27.92 (8.84)	/									

SA social anhedonia, CSAS Revised Social Anhedonia Scale, PANSS Positive and Negative Syndrome Scale for Schizophrenia.  
\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\* $p < 0.001$ .

SA social anhedonia, CSAS Revised Social Anhedonia Scale, PANSS Positive and Negative Syndrome Scale for Schizophrenia.

$$^*p < 0.05, ^{**}p < 0.01, ^{***}p < 0.001.$$

perspective" showed increased activity in the precuneus, right superior parietal, and right premotor cortex, while the "first-person perspective" showed an increased activity in the mesial prefrontal cortex, posterior cingulate cortex, and superior temporal cortex bilaterally. Both "first-person perspective" and "third-person perspective" rely on egocentric reference frames. However, the "third-person perspective" requires an additional translocation of the egocentric viewpoint, leading to differential neural processes. Therefore, we believe that there will be differences between the original VAMA and the current adapted version. Brown<sup>33</sup> also showed that the "first-person" perspective is about the self, known as egocentric, but a "third-person" perspective involves the perception of others from an external frame of reference, regarded as "all-centric." Since ToM is the process of adopting the third-person perspective to infer the mental state of others, we decided to adopt the "third-person" instead of the "first-person" perspective in this study. (2) The original VAMA involves a shopping task that places considerable demands on the executive function of the participants, thereby increasing their cognitive load. This heightened cognitive demand can detract from the primary focus of our study, which is to assess mentalizing ability. To align with the main objective of our research, we adopted a methodology similar to that used in other well-known ToM tasks, e.g., the Movie for the Assessment of Social Cognition (MASC)<sup>34</sup>. Consequently, our study exclusively used ToM video clips and corresponding ToM questions. This approach ensured that the participants' cognitive resources were directed toward evaluating their mentalizing abilities, rather than being diverted by tasks that require extensive executive functioning. By focusing on ToM-specific stimuli, we aimed to obtain a more accurate assessment of the participants' mentalizing capabilities.

**New scenarios and ecological validity:** The original storyline of the VAMA was developed within a Western cultural framework, which may not be entirely suitable for the Chinese context. As Van de Vijver<sup>35</sup> highlighted, norm-driven adaptations are essential to accommodate cultural differences in norms, values, and practices, as specific cultural contexts may not universally apply. Social norms influence not only individual actions but also the expectations others have of those actions, potentially affecting the inference of mental states. For instance, cultural variations are evident in how faux pas are perceived; mistaking a customer for a waiter is universally considered a faux pas by English participants (100%), but not by Canadians (65.4%). Similarly, while 100% of English participants view giving up a seat to an older passenger on a bus as normal behavior, 21.2% of Chinese participants perceive it as a faux pas, despite it being a control item with no faux pas intended. These variations underscore the need for concept-driven adaptations rather than literal translations of neuropsychological assessments<sup>36</sup>.

As mentioned, the storyline of the VAMA<sup>26</sup> was developed based on Western culture, which may not be appropriate for the Chinese context. Thus, after seeking advice from two subject-matter investigators (YC and DS, both are co-authors of the current study and neuropsychologists), some videos that were not culturally suitable were removed (#1, #3, #4, #6, #7, #9, and #10; from the original VAMA)<sup>26</sup>, for example, in scenario #3 from the original VAMA, the story is about the test-taker and character B meets character A at the hairdresser, where she has just had her hair cut. Character B positively jokes about Character A's hair, saying "You need to get a new hairdresser. That looks awful." While this scenario is considered normal/appropriate in Western countries, it is unusual that friends will tell this kind of joke between friends in China. Some new scenarios (control #3, control #4, #3, #4, #5, #6, #7, #8, and #10; from the current adapted version of VAMA) were created based on the VAMA framework, capturing the Chinese collectivistic cultures of obedience to authority, filial piety, ancestral worship, conservatism, and perseverance<sup>37,38</sup>. To ensure the created scenarios are culturally sensitive and

ecologically valid, pilot data were obtained from 23 volunteers ( $M_{age} = 23.57$  years,  $SD = 2.46$ ), asking for comments and opinions on the scenarios.

Ecological validity refers to the extent to which outcomes derived from controlled experimental conditions correspond to the performance observed in real-world settings<sup>39</sup>. However, as mentioned by Caviere et al.<sup>39</sup>, few ToM assessments of schizophrenia have explicitly defined ecological validity. Thus, to ensure that the scenarios relate to young adults' daily life events, and to improve the verisimilitude and ecological validity of the stimuli, three questions, one each on familiarity, understanding, and relatedness, were designed. The volunteers were recruited to read all 10 scenarios and then asked to provide comments on the scripts and rate each scenario based on the level of familiarity (Question: "How familiar are you with this scenario?" rated from 1 = "Not familiar at all" to 7 = "Very familiar"), understanding (Question: "To what extent do you believe you understand the thoughts or emotions of these story characters at that time?" rated from 1 = "Do not understand at all" to 7 = "Understand very well"), and more specifically, relatedness to determine how closely the scenarios mirror real-life situations and ensure that the scenarios are relevant to their (or their peers') real-life experiences (Question: "How closely does this scenario align with the lives of young people (aged 18–30)?" rated on a 7-point Likert scale from 1 = "Not close at all" to 7 = "Very closely aligned"). The volunteers reported that most of the scenarios were familiar to them ( $M = 4.37$ ,  $SD = 1.04$ ), easy to understand ( $M = 4.93$ ,  $SD = 1.13$ ), and close to their daily lives ( $M = 4.70$ ,  $SD = 1.10$ ). Scripts were also provided to four subject-matter experts (YC, YW, RC, and DS) to validate the content of the Chinese version. They are the co-authors of the current study, are experienced in developing assessment tools, conducting social cognition studies and have published papers in social cognition and mental disorders. After receiving all of the comments, appropriate revisions were made, and the final scripts were used in the video filming. Appendix A (Supplementary Materials) shows an example scenario of the VAMA, and Appendix B (Supplementary Materials) shows one of the sample questions.

A professional technical production team was employed to carry out the filming at The Hong Kong Polytechnic University. The first version of the scenarios were recorded on a green screen and then transformed into a digital format. Editing and postproduction were completed using Adobe Premiere Pro® (Adobe®, USA). However, different from the original version<sup>26</sup>, the current version did not develop the virtual shopping environment. Considering the verisimilitude and authenticity which may bias participants' judgments<sup>40</sup>, we decided to employ a real-world environment to develop the videos instead of building the virtual shopping center environment in the current study. In addition, in view of the usability in other Asian countries in the future, we decided to use this setting in the current study. This is because the virtual background that we adopted is comparatively easier to change into another background while the virtual shopping environment (Canty's) is developed based on the Australian shopping mall that people from other countries may not be familiar with.

After obtaining the permission of a local shopping mall, the second co-first author and two programmers took photos of the mall and built the video background. Thereafter, the research team used E-Prime 3.0 to develop the stimuli. The filmed segments, including the video clips and ToM questions, were merged, and their quality was checked before data collection. Different from the original version, this adapted version did not involve any shopping tasks like the original version<sup>26</sup>. It is believed that the original VAMA involves a shopping task that requires significant executive function, thereby increasing participants' cognitive load. However, the primary aim of this study was to assess mentalizing ability. To focus on this, we adopted a methodology similar to that of another widely used ToM

assessment tools (e.g., MASC)<sup>34</sup>, using only ToM video clips and questions in the current study. This approach ensures that participants' cognitive resources are devoted to their mentalizing abilities, rather than being distracted by tasks demanding executive functioning. The participants completed the task (i.e., watching the video clips and answering the ToM questions) on a laptop, which had the tool installed. To enhance the experimental experience and minimize disturbances from external noise, we provided all participants with over-ear headphones. These headphones are designed to fit snugly over the head and ears, offering superior sound isolation and noise-canceling capabilities. The intimate auditory environment allowed the participants to immerse themselves fully in the task at hand. Thereafter, to ensure the feasibility and effectiveness of the experiment, we conducted a pilot study in mainland China to collect preliminary data. Pilot data were obtained from 66 volunteers in Weifang (a prefecture-level city in China's central Shandong province;  $n = 35$ ,  $M_{\text{age}} = 20.45$  years,  $SD = 0.86$ ) and Beijing (the capital of China;  $n = 31$ ,  $M_{\text{age}} = 22.97$  years,  $SD = 1.25$ ). This initial phase allowed us to test our experimental design, procedures, and materials in a real-world setting, providing valuable insights into potential challenges and areas for refinement. Although we did not perform data analysis on these pilot data sets, the process was instrumental in assessing the practicality of our approach and identifying necessary adjustments before proceeding to the full-scale study.

**VAMA scoring:** Following the original VAMA version, the participants' responses were scored dichotomously: each "Accurate mentalizing" response was rated as 1, and others were rated as 0. In addition, "No ToM," "Hypermentalizing," and "Reduced ToM" were scored dichotomously to compare the frequencies of these error types. Scores for each subscale and error type ranged from 0 to 10, and the total accurate score ranged from 0 to 40.

**Procedures.** After providing informed consent, eligible participants were invited to complete the computer-based VAMA and the questionnaires via WenJuanXing (Changsha Ranxing Information Technology Co., Ltd., Hunan, China), which is a data science company that specializes in managing a personal information database containing details, such as age, gender, and residence, for over 2.6 million Chinese residents. The platform enables research teams to distribute questionnaires online and recruit specific participants based on chosen sampling methods. WenJuanXing has been utilized in many studies to gather representative samples and conduct cross-sectional online surveys<sup>41,42</sup>. To examine test-retest reliability, 30 participants were randomly

selected and asked to complete the VAMA again approximately 4 weeks (i.e.,  $\pm 1$  day) after the first testing session.

**Data analyses.** IBM SPSS 29.0.0.0 was used to analyze the collected data. The item difficulty (by calculating the mean accuracy of each item) and item-total correlation for each item were calculated to examine the internal consistency of the adapted version of the VAMA, and Pearson's correlation coefficients were calculated to examine its test-retest reliability. To evaluate construct validity, a  $2 \times 2$  repeated-measures analysis of variance (ANOVA) was used to investigate the effects of Order (viz., first- vs. second-order ToM) and Type (viz., cognitive vs. affective ToM) on ToM performance. All data (de-identified and anonymous), analysis codes, and research materials are available upon request from the corresponding author.

## RESULTS

### Mean scores, item difficulty, and item-total correlation

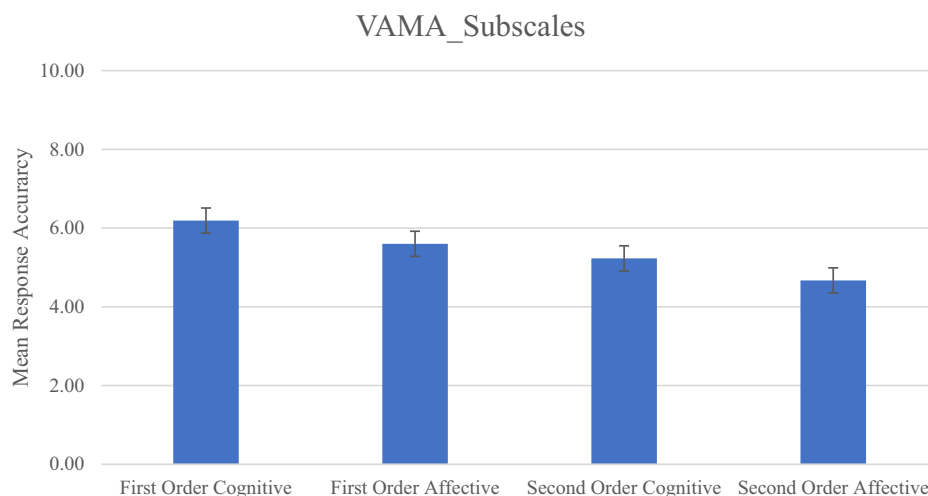
The descriptive statistics for each VAMA subscale are shown in Fig. 1 and Table 2. Item difficulty with the four subscales was as follows: first-order cognitive ToM = 0.40–0.84, first-order affective ToM = 0.24–0.89, second-order cognitive = 0.11–0.81, second-order affective = 0.13–0.65. The item-total correlations were statistically significant, with either  $p_{FDR} < 0.05$  or  $p_{FDR} < 0.01$ . These results suggest that all of the items measured their respective subscales (see Table 2).

### Construct Validity

A significant main effect of Order was found ( $F(1, 99) = 39.04$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.28$ ), where the participants showed better performance in first-order ToM ( $M = 5.90$ ,  $SD = 1.19$ ) than in second-order ToM ( $M = 4.95$ ,  $SD = 1.23$ ). The results also revealed a significant main effect of Type ( $F(1, 99) = 20.76$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.17$ ), where the participants performed better in cognitive ToM ( $M = 5.71$ ,  $SD = 1.16$ ) than in affective ToM ( $M = 5.14$ ,  $SD = 1.11$ ). The 2-way Order  $\times$  Type interaction was not statistically significant ( $F(1, 99) = 0.008$ ,  $p > 0.93$ ).

### Test-retest reliability

The adapted version of the VAMA was readministered to a randomly selected subsample of 30 individuals 4 weeks after the initial test. Moderate to strong positive correlations were found between the test-retest scores: first-order cognitive ToM:



**Fig. 1** Mean response accuracy on the subscales of the VAMA. Error bars represent standard error.



**Table 2.** Table of item difficulty and item-total correlation of VAMA in Study 1.

	First-order Cognitive		First-order Affective		Second-order Cognitive		Second-order Affective	
	<i>M</i>	<i>r</i>	<i>M</i>	<i>r</i>	<i>M</i>	<i>r</i>	<i>M</i>	<i>r</i>
Item 01	0.84	0.251*	0.89	0.141	0.55	0.367**	0.61	0.306**
Item 02	0.40	0.319**	0.76	0.273**	0.49	0.308**	0.38	0.303**
Item 03	0.63	0.498**	0.66	0.346**	0.59	0.240*	0.55	0.435**
Item 04	0.80	0.336**	0.69	0.430**	0.11	0.178	0.43	0.267**
Item 05	0.44	0.324**	0.74	0.346**	0.69	0.240*	0.13	0.330**
Item 06	0.68	0.382**	0.24	0.431**	0.76	0.447**	0.45	0.403**
Item 07	0.58	0.237*	0.64	0.288**	0.81	0.419**	0.60	0.373**
Item 08	0.84	0.438**	0.32	0.328**	0.43	0.348**	0.30	0.449**
Item 09	0.51	0.386**	0.34	0.363**	0.29	0.410**	0.65	0.420**
Item 10	0.50	0.388**	0.33	0.360**	0.54	0.424**	0.62	0.377**

ToM Theory of mind.  
\* $p_{FDR} < 0.05$ , \*\* $p_{FDR} < 0.01$ .

**Table 3.** Table of test-retest reliability in Study 1.

	First-order Cognitive_T1	First-order Affective_T1	Second-order Cognitive_T1	Second-order Affective_T1	Total Score _T1
First-order Cognitive	0.568**	-	-	-	-
First-order Affective	-	0.326	-	-	-
Second-order Cognitive	-	-	0.296	-	-
Second-order Affective	-	-	-	0.380	-
Total score	-	-	-	-	0.731***

ToM Theory of mind.  
\*\* $p_{FDR} < 0.01$ , \*\*\* $p_{FDR} < 0.001$ .

$r = 0.568$ ,  $p_{FDR} < 0.01$ ; first-order affective ToM:  $r = 0.326$ ,  $p_{FDR} > 0.05$ ; second-order cognitive ToM:  $r = 0.296$ ,  $p_{FDR} > 0.05$ ; second-order cognitive ToM:  $r = 0.380$ ,  $p_{FDR} < 0.05$ ; and total score:  $r = 0.731$ ,  $p_{FDR} < 0.01$  (see Table 3).

## STUDY 2

### Method

#### Participants

**Sample 1: Schizophrenia patients and healthy matched controls:** To examine the clinical utility of the adapted version of the VAMA, 39 schizophrenia patients ( $M_{age} = 29.13$  years,  $SD = 9.11$ , age range = 18–51) were recruited from Peking University Sixth Hospital, and 37 healthy matched controls ( $M_{age} = 27.57$  years,  $SD = 8.04$ , age range = 19–50) were recruited from the community in Beijing, China. This clinical group included 29 inpatients and 10 outpatients who met the diagnostic criteria for schizophrenia according to the Fifth Edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5)<sup>43</sup> with no specified limit for illness duration. The exclusion criteria for all participants were those with an IQ below 70, a history of substance abuse or drug addiction, a record of brain trauma or neurological disorders, and a history of electroconvulsive therapy within the past 2 months. For the healthy matched control group, only those lacking psychiatric disorders and any family history of psychiatric or neurological disorders were recruited. The Chinese version of the Mini International Neuropsychiatric Interview (MINI)<sup>44</sup> was used to screen individuals with psychiatric disorders for this group.

Table 1 shows the demographic information of the two groups. Nonsignificant between-group differences were found in gender

and age ( $p > 0.05$ ), while significant between-group differences were found in education years ( $p < 0.01$ ) and estimated IQ scores ( $p < 0.001$ ). This study was approved by the Ethics Committee of the Peking University Sixth Hospital (reference number: 2021–24). Written informed consent was obtained from all participants before data collection.

**Sample 2: Individuals with high and low levels of SA:** An extreme group design was adopted to explore whether the adapted version of the VAMA can detect ToM impairment in people with high levels of SA. The participants were divided into two groups: those with high and low SA scores. The initial pool of 13,709 individuals consisted of students who participated in a large-scale survey activity in Shandong, China, during which they completed the Revised SA Scale (CSAS)<sup>43</sup>. This pool of people was not specifically recruited for our experiment but served as “potential participants.” According to previous studies’ sample size (which also adopted the high and low SA group design)<sup>44,45</sup>, we randomly invited 174 individuals from this pool and only 126 of them were willing to participate in this study. Upon arrival for the study, participants were asked to complete the CSAS again to ensure that their scores still met the inclusion criteria for the high or low SA groups. The exclusion criteria were identical to those in Study 1. Only those who continued to meet these criteria were included in the final sample of 102 participants. In accordance with the criteria of a previous study on SA<sup>46</sup>, individuals were allocated into the high SA group ( $n = 48$ ) if their CSAS score was higher than 1 standard deviation ( $SD = 6.67$ ) equal to or above the mean score ( $M = 10$ , cut off score  $\geq 17$ ), whereas individuals whose CSAS score was equal to or lower than the mean score were allocated into the low SA group ( $n = 54$ , cut-off score  $\leq 10$ ). The high SA group had a

**Table 4.** Pearson's correlations between IQ, years of education, duration of illness, VAMA, and Yoni subconstructs for patients with Schizophrenia and healthy controls in Study 2 (sample 1).

	1	2	3	4	5	6	7	8	9	10	11	12
1. Age	/											
2. IQ (Estimate)	−0.054	/										
3. Years of Education	0.040	0.486***	/									
4. Duration of illness	0.240	−0.330	−0.470**	/								
5. VAMA_1st Cognitive	−0.160	0.545***	0.449***	−0.283	/							
6. VAMA_1st Affective	−0.100	0.185	0.145	0.114	0.316**	/						
7. VAMA_2nd Cognitive	−0.067	0.282*	0.146	−0.016	0.213	0.400***	/					
8. VAMA_2nd Affective	0.004	0.327**	0.173	−0.121	0.322**	0.263	0.345**	/				
9. YONI_1st Cognitive	−0.175	0.202	0.138	−0.252	0.323**	0.070	0.278*	0.010	/			
10. YONI_1st Affective	−0.211	0.218	0.144	−0.262	0.339**	0.165	0.282*	0.034	0.849***	/		
11. YONI_2nd Cognitive	−0.205	0.489***	0.088	−0.211	0.349**	0.180	0.344**	0.087	0.722***	0.637***	/	
12. YONI_2nd Affective	−0.162	0.503***	0.230	−0.265	0.477***	0.318**	0.389**	0.145	0.791***	0.748***	0.850**	/

ToM Theory of mind, VAMA Virtual Assessment of Mentalizing Ability, 1st first-order, 2nd second-order.

\*Represents  $p_{FDR} < 0.05$ , \*\*represents  $p_{FDR} < 0.01$ , \*\*\* represents  $p_{FDR} < 0.001$ .

mean age of 19.83 years ( $SD = 0.72$ ) and consisted of 25 male and 23 female participants. The low SA group had a mean age of 19.81 years ( $SD = 0.70$ ) and consisted of 22 male and 32 female participants (see Table 1). No significant between-group differences were found in gender, age, and education years ( $p > 0.05$ ). This study was approved by the Ethics Committee of the Weifang Medical University (reference number: H20045). We obtained written informed consent from all participants before data collection. All participants received RMB110 after completing the whole study.

#### Measurements

**VAMA task:** The adapted version of the VAMA task, as described in Study 1, was also used in Study 2.

**Clinical assessment.** The severity of clinical symptoms in schizophrenia patients (sample 1) was assessed using the Positive and Negative Syndrome Scale (PANSS) developed by Kay et al.<sup>47</sup>. The Chinese version of the PANSS was used in the current study<sup>48</sup>. Estimated IQ was measured using the Wechsler Adult Intelligence Scale-Revised Chinese version (WAIS-RC)<sup>49</sup>.

**Revised SA scale:** Revised Social Anhedonia Scale (CSAS)<sup>50</sup>. The CSAS was administered to the participants (sample 2) to measure their hedonic experience. The CSAS assesses the reduction in the level of hedonic experience in interpersonal relationships and communications using 40 true–false items in clinical or nonclinical settings. Higher scores indicate more severe SA. The Chinese version of the CSAS has been used and validated in China and has shown good internal consistency (Cronbach's  $\alpha = 0.86$ )<sup>43</sup>.

**Chinese version of the Yoni task:** To examine the group difference between patients with schizophrenia and healthy matched controls (sample 1), another ToM assessment tool, the Chinese version of the computerized Yoni task<sup>51,52</sup> was used. In total, 98 trials were conducted for this task, comprising 32 and 66 first-order and second-order ToM questions, respectively. The participants were asked to infer the thoughts (cognitive ToM), feelings (affective ToM), and position (physical condition, control condition) of “Yoni” (Chinese name is Xiaoming) according to his facial expression. The Chinese version has been applied to schizophrenia patients in previous studies<sup>53</sup>.

**Procedures.** A psychiatrist evaluated the eligibility of all participants in sample 1 after they had signed the informed consent form. The doctor also interviewed the healthy participants to rule

out mental illness. This study comprised three parts: a semi-structured interview, including the PANSS (for the clinical group) or MINI (for healthy participants), self-reported surveys and experimental tasks (VAMA and Yoni). Finally, they were administered the two ToM tasks (VAMA and Yoni task) by trained research assistants, which took around 1.5–2 hrs. The participants in sample 2 were divided into the high- and low-level SA groups based on their CSAS scores. Eligible participants were then invited to complete the adapted version of the VAMA only. The entire process took 30 mins to 1 h.

**Data analysis.** Studies have reported associations between premorbid intelligence and ToM, in addition to significant correlations between the estimated IQ scores and the VAMA and Yoni task (see Table 4). Therefore, analyses of covariance (ANCOVAs) were adopted to analyze the data. Two 2 (Schizophrenia vs. Healthy Matched Control group)  $\times$  2 (First- vs. Second-order ToM)  $\times$  2 (cognitive versus affective ToM) mixed ANCOVAs were then conducted for the VAMA and Yoni task. Independent samples  $t$ -tests were also conducted to examine between-group differences in the four subconstructs (for the VAMA and Yoni task) and three error types. As we proposed a directional hypothesis, we evaluated the one-tailed  $p$ -value rather than the two-tailed  $p$ -value<sup>54</sup>. For sample 2, similar data analyses were adopted for the adapted VAMA. Finally, Pearson's correlations were calculated to examine the association between clinical symptoms of schizophrenia patients and the frequency of error types of the adapted version of the VAMA. All data, analysis code, and research materials are available upon request from the corresponding author.

## RESULTS

### Sample 1: Schizophrenia versus healthy matched control group

**VAMA Task.** As expected, a significant main effect of Group was found ( $F(1, 73) = 37.23$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.34$ ), where overall, the clinical group performed ( $M = 4.23$ ,  $SD = 1.50$ ) significantly poorer than the healthy matched control group ( $M = 5.33$ ,  $SD = 1.55$ ), as shown in Appendix C. Non-significant main effects of Order ( $F(1, 73) = 1.51$ ,  $p > 0.05$ ) and Types ( $F(1, 73) = 1.92$ ,  $p > 0.05$ ) were also found. Notably, a significant two-way interaction between Order  $\times$  Type ( $F(1, 73) = 4.41$ ,  $p < 0.05$ ,  $\eta_p^2 = 0.06$ ) was observed. Particularly, in cognitive ToM, participants performed better in

**Table 5.** VAMA (adapted version) and Yoni Task performance for the Schizophrenia and Healthy Control group in Study 2.

	Schizophrenia ( <i>n</i> = 39)		Healthy Matched Control ( <i>n</i> = 37)		<i>t</i>	<i>p</i> (one-tailed)	df	Cohen' <i>d</i> (95% C.I.)
	Mean	SD	Mean	SD				
VAMA_1st Cognitive	4.31	1.49	6.32	1.65	5.60***	<0.001	74	1.28 [0.79,1.78]
VAMA_1st Affective	4.18	1.41	5.70	1.58	4.44***	<0.001	74	1.02 [0.54,1.49]
VAMA_2nd Cognitive	3.92	1.77	5.05	1.60	2.92**	0.002	74	0.67 [0.21,1.13]
VAMA_2nd Affective	3.87	1.94	4.92	1.64	2.54**	0.007	74	0.58 [0.12,1.04]
VAMA_hypermentalizing	4.78	2.63	2.75	1.86	3.87***	<0.001	68.53	0.88 [−1.35,−0.41]
VAMA_reduced_ToM	10.86	2.44	11.64	2.78	1.04	0.152	74	0.24 [−0.21,0.69]
VAMA_No_ToM	4.92	2.49	3.11	2.19	3.57***	<0.001	74	0.81 [−1.29,−0.35]
Yoni_1st Cognitive	0.80	0.29	0.93	0.17	2.20*	0.016	58.26	0.51 [0.04,0.98]
Yoni_1st Affective	0.80	0.30	0.92	0.16	2.07*	0.02	55.86	0.48 [0.01,0.95]
Yoni_2nd Cognitive	0.62	0.26	0.81	0.18	3.55***	<0.001	64.41	0.83 [0.35,1.30]
Yoni_2nd Affective	0.68	0.20	0.86	0.13	4.32***	<0.001	61.76	1.01 [0.52,1.49]

VAMA Virtual Assessment of Mentalizing Ability, ToM Theory of mind, 1st first-order, 2nd second-order.

\*Represents  $p < 0.05$ , \*\* represents  $p < 0.01$ , \*\*\* represents  $p < 0.001$ .

the first-order ToM ( $M = 5.31$ ,  $SD = 1.47$ ) than the second-order ToM ( $M = 4.49$ ,  $SD = 1.68$ ,  $F(1, 73) = 10.23$ ,  $p = 0.002$ , and  $\eta_p^2 = 0.12$ ), while in affective ToM, participants also performed better in first-order ( $M = 4.94$ ,  $SD = 1.50$ ) than second order ( $M = 4.39$ ,  $SD = 1.77$ ,  $F(1, 73) = 5.24$ ,  $p = 0.03$ , and  $\eta_p^2 = 0.07$ ). In both first- and second-order ToM, non-significant main effects were found between types ( $p > 0.05$ ). The three-way interaction between Order  $\times$  Type  $\times$  Group ( $F(1, 73) = 0.50$ ,  $p > 0.05$ ) and the two-way interactions between Order  $\times$  Group ( $F(1, 73) = 3.68$ ,  $p > 0.05$ ), and Type  $\times$  Group ( $F(1, 73) = 0.01$ ,  $p > 0.05$ ) were not statistically significant.

Planned comparisons using *t*-tests were conducted to examine between-group differences in all four subconstructs of the adapted version of the VAMA. The clinical group showed significantly worse performance on all subconstructs [viz., first-order cognitive ( $t(74) = 5.60$ ,  $p_{\text{one-tailed}} < 0.001$ ,  $d = 1.28$ ), first-order affective ( $t(74) = 4.44$ ,  $p_{\text{one-tailed}} < 0.001$ ,  $d = 1.02$ ), second-order cognitive ( $t(74) = 2.92$ ,  $p_{\text{one-tailed}} = 0.002$ ,  $d = 0.67$ ), second-order affective ToM ( $t(74) = 2.54$ ,  $p_{\text{one-tailed}} = 0.007$ ,  $d = 0.58$ ; also see Appendix C and Appendix D]. In addition, patients in the schizophrenia group showed significantly more “hypermentalizing” errors ( $t(68.53) = 3.87$ ,  $p_{\text{one-tailed}} < 0.001$ ,  $d = 0.88$ ) and “no mentalizing” errors ( $t(74) = 3.57$ ,  $p_{\text{one-tailed}} < 0.001$ ,  $d = 0.81$ ) than did the healthy control group, while a non-significant between-group difference was found in “reduced ToM” errors ( $t(74) = 1.04$ ,  $p_{\text{one-tailed}} = 0.152$ ; see Table 5).

Table 6 summarizes the correlations between clinical symptoms and mentalizing error types in patients with schizophrenia. A significant positive association was found between negative symptoms and “hypermentalizing” errors ( $r = 0.388$ ,  $p_{\text{FDR}} < 0.05$ ), but none of the other symptom–error type associations were found to be statistically significant ( $p > 0.05$ ).

### Yoni task

The mixed ANCOVA results showed significant main effects and a significant interaction effect (also see Appendix E). First, a significant main effect of Order was found ( $F(1, 70) = 17.25$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.20$ ), where the participants showed better performance in first-order ToM ( $M = 0.86$ ,  $SD = 0.24$ ) than in second-order ToM ( $M = 0.74$ ,  $SD = 0.18$ ).

A non-significant main effect of Type was also found ( $F(1, 70) = 1.08$ ,  $p > 0.05$ ). Surprisingly, non-significant main effect of the clinical group was observed on Yoni task performance ( $F(1,$

**Table 6.** Pearson's correlations between clinical symptoms and mentalizing error types for patients with schizophrenia.

	Positive Symptoms	Negative Symptoms	General Symptoms
VAMA_hypermentalizing	0.212	0.388*	0.215
VAMA_reduced_ToM	0.205	−0.118	0.029
VAMA_No_ToM	0.043	0.243	0.070

ToM Theory of mind.

\*Represents  $p_{\text{FDR}} < 0.05$ .

$70) = 3.03$ ,  $p > 0.05$ ). Similar to the VAMA results, a significant two-way interaction effect between Order  $\times$  Type was found ( $F(1, 70) = 4.35$ ,  $p < 0.05$ ,  $\eta_p^2 = 0.06$ ); specifically, in cognitive ToM, the participants performed better in first-order ToM ( $M = 0.87$ ,  $SD = 0.24$ ) than in second-order ToM ( $M = 0.71$ ,  $SD = 0.22$ ,  $F(1, 70) = 57.36$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.45$ ), while in affective ToM as well, the participants also performed better in first-order ToM ( $M = 0.86$ ,  $SD = 0.24$ ) than in second-order ToM ( $M = 0.77$ ,  $SD = 0.17$ ,  $F(1, 70) = 21.49$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.24$ ). In second-order ToM, the participants performed better in affective ToM ( $M = 0.77$ ,  $SD = 0.13$ ) than in cognitive ToM ( $M = 0.71$ ,  $SD = 0.13$ ,  $F(1, 70) = 13.92$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.17$ ), while non-significant simple main effects of Type were observed on first-order ToM. Similar to the VAMA results, the three-way interaction between Order  $\times$  Type  $\times$  Group ( $F(1, 70) = 0.31$ ,  $p > 0.05$ ) and the two-way interactions between Order  $\times$  Group ( $F(1, 70) = 0.013$ ,  $p > 0.05$ ) and Type  $\times$  Group ( $F(1, 70) = 0.00$ ,  $p > 0.05$ ) were not found to be statistically significant. In particular, four independent samples *t*-tests were performed to compare the four subconstructs of ToM between the two groups. The clinical group showed significantly worse performance in all of the subconstructs [viz., first-order cognitive ( $t(58.26) = 2.20$ ,  $p_{\text{one-tailed}} = 0.016$ ,  $d = 0.51$ ), first-order affective ( $t(55.86) = 2.07$ ,  $p_{\text{one-tailed}} < 0.02$ ,  $d = 0.48$ ), second-order cognitive ( $t(64.41) = 3.55$ ,  $p_{\text{one-tailed}} < 0.001$ ,  $d = 0.83$ ), and second-order affective ToM ( $t(61.76) = 4.32$ ,  $p_{\text{one-tailed}} < 0.001$ ,  $d = 1.01$ ; also see Appendix E] than the healthy matched control group.

### Sample 2: High level versus low level of SA

**VAMA Task.** The mixed ANOVA results revealed two significant main effects (see Appendix F, Figs. 1 and 2). First, a main effect of

**Table 7.** VAMA (adapted version) performance for high SA and low SA in Study 2.

	High SA ( <i>n</i> = 48)		Low SA ( <i>n</i> = 54)		<i>t</i>	<i>p</i> (one-tailed)	df	Cohen's <i>d</i> (95% C.I.)
	Mean	SD	Mean	SD				
VAMA_1st Cognitive	5.83	1.34	5.94	1.48	0.40	0.35	100	0.08[−0.31,0.47]
VAMA_1st Affective	5.50	1.37	5.31	1.52	0.65	0.26	100	0.13[−0.52,0.26]
VAMA_2nd Cognitive	5.00	1.44	5.56	1.72	1.75*	0.04	100	0.35[−0.05,0.74]
VAMA_2nd Affective	4.62	1.84	4.87	1.47	0.75	0.23	100	0.15[−0.24,0.54]
VAMA_hypermentalizing	4.17	2.73	3.81	2.49	0.68	0.25	100	0.14[−0.52,0.26]
VAMA_reduced_ToM	11.63	2.37	11.15	2.52	0.98	0.16	100	0.20[−0.58,0.20]
VAMA_No_ToM	2.73	1.77	2.59	1.56	0.41	0.34	100	0.08[−0.47,0.31]

VAMA Virtual Assessment of Mentalizing Ability, SA social anhedonia  
 \*Represents  $p < 0.05$

order was observed ( $F(1, 100) = 20.28$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.17$ ); specifically, the participants performed significantly better in first-order ( $M = 5.65$ ,  $SD = 1.12$ ) than in second-order ToM ( $M = 5.01$ ,  $SD = 1.30$ ). Second, a significant main effect of Type was also found ( $F(1, 100) = 14.26$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.13$ ), where the participants performed significantly better in cognitive ToM ( $M = 5.58$ ,  $SD = 1.25$ ) than in affective ToM ( $M = 5.08$ ,  $SD = 1.13$ ). The main effect of Group ( $F(1, 100) = 0.88$ ,  $p = 0.35$ ) and the interaction effects of Order  $\times$  Group ( $F(1, 100) = 2.40$ ,  $p = 0.124$ ), Order  $\times$  Types ( $F(1, 100) = 0.04$ ,  $p = 0.85$ ), Types  $\times$  Group ( $F(1, 100) = 1.28$ ,  $p = 0.26$ ), and Order  $\times$  Types  $\times$  Group ( $F(1, 100) = 0.00$ ,  $p = 0.98$ ) were not significantly different.

To examine whether the two groups differed in the four subconstructs of ToM, planned comparisons using *t*-tests were performed. A significant between-group difference was found in second-order cognitive ToM ( $t(100) = 1.75$ ,  $p_{\text{(one-tailed)}} = 0.04$ ,  $d = 0.35$ ), where the participants with a high level of SA ( $M = 5.00$ ,  $SD = 1.44$ ) scored lower than their low SA counterparts ( $M = 5.56$ ,  $SD = 1.72$ ; see Table 7). Finally, non-significant differences were observed in the mentalizing error types between the two groups ["hypermentalizing":  $t(100) = 0.68$ ,  $p_{\text{(one-tailed)}} = 0.25$ ,  $d = 0.14$ ; "reduced ToM":  $t(100) = 0.98$ ,  $p_{\text{(one-tailed)}} = 0.16$ ,  $d = 0.20$ ; "no ToM":  $t(100) = 0.41$ ,  $p_{\text{(one-tailed)}} = 0.34$ ,  $d = 0.08$ ]; see Table 7].

## GENERAL DISCUSSION

In the current study, we hypothesized that the adapted version would demonstrate psychometric properties comparable to the original Australian version<sup>26</sup>, including satisfactory test-retest reliability, the comparable mean accuracy of the four subconstructs, strong item-total correlations, neither floor nor ceiling effect would be observed, and participants would perform better in first-order than in second-order ToM and cognitive than affective ToM. Results from Study 1 provided evidence to support the adaptation of the VAMA for the Chinese culture, with acceptable psychometric properties. To further support the validity, we also hypothesized that schizophrenia patients and individuals with high levels of SA would show poorer performance in VAMA and make more ToM errors. In addition, we also expected that associations between clinical symptoms and the error types of VAMA in schizophrenia patients would be found. Results from Study 2 demonstrated that schizophrenia patients showed impairment on all four subconstructs of ToM, and they were found to make more hypermentalizing and no mentalizing ToM errors. On the other hand, individuals with high levels of SA, only showed impairment in one of the subconstructs of ToM (viz., second-order-cognitive ToM) and these individuals did not make more ToM error types compared to low levels of SA.

In light of the findings from Study 1, the adapted Chinese version of the VAMA is somehow similar in psychometric

properties to the original Australian version<sup>26</sup> (See Appendix G in Supplementary materials). We removed some scenarios that appeared more specific for the Western culture and added some new scenarios and items that were appropriate for the Chinese context, and still found acceptable internal consistency for this adapted version. Consistent with our hypothesis, the main effect of Order and Types were observed from the ANOVA results, where first-order ToM was found to be easier than second-order ToM, and cognitive ToM was easier than affective ToM. These results are consistent with previous studies<sup>11,26</sup>, supporting the construct validity of this adapted version of VAMA. However, it should be noted that the test-retest reliability for the four subconstructs requires further studies. Nevertheless, it should be pointed out that the test-retest reliability findings were similar to those of other ToM tools, like The Awareness of Social Inferences Test (TASIT, test-retest: 0.534)<sup>55</sup>, Hinting Task ( $r_{\text{test-retest}}$ : 0.509)<sup>18</sup>, and Social Attribution Task—Multiple Choice (SAT-MC,  $r_{\text{test-retest}}$ : 0.554)<sup>56–58</sup>. Although most of the ToM tools were found to have adequate to satisfactory known-group validity, their internal consistency and test-retest reliability in the non-clinical populations were not "satisfactory" as suggested by the classical test theory<sup>59</sup>. The possible reason is that the complex format of the constructs reduces their reliability<sup>58</sup>. Thus, it is not surprising that the second-order ToM in our current study, requires a higher and more complex reasoning skills, has a relatively low reliability when compared to the simpler first-order ToM. Another reason may be that the test was constructed with items of diverse levels of difficulty (two/three/four characters in a scenario), which could have contributed to reduced internal consistency<sup>60</sup>. Previous story-based, film-based studies also showed a similar level of internal consistency, i.e., the "Strange Stories Film Task" (Cronbach's  $\alpha = 0.58$  (Intention), 0.42 (Mental state language question), 0.73 (Interaction question)<sup>60</sup>, and VAMA (First-order cognitive: Cronbach's  $\alpha = 0.69$ )<sup>26</sup>.

In Study 2, the adapted version of VAMA revealed ToM impairments in clinical and subclinical populations. In general, schizophrenia patients performed significantly worse than the healthy controls on the VAMA (both total and subconstruct scores) and the Yoni task (subconstruct scores), which is in line with our expectations<sup>10–13</sup>. ToM impairment in schizophrenia patients could be explained by the deactivation of different brain regions. A recent systematic review<sup>61</sup> suggested that, compared to healthy controls, schizophrenia patients demonstrated deactivation of brain areas during ToM tasks, such as the inferior parietal gyrus, anterior cingulate cortex, and posterior cingulate cortex, impacting their emotional processing, empathy, and executive functions. Moreover, unlike controls, the patients showed no activation of the cerebellum and thalamus<sup>62</sup>. In addition to neurobiological deficits, childhood adversity, and emotional neglect have been identified as significant predictors of ToM impairment in



schizophrenia patients<sup>54,63,64</sup>. The attachment-developmental-cognitive hypothesis suggests that key psychological features of schizophrenia, such as ToM deficits and negative self-and-other attributions, are attributable to disruptions or impairments in attachment relationships during the first decade of life<sup>65</sup>. Consequently, it is plausible that the psychological mechanisms underlying ToM deficits in schizophrenia are related to factors associated with disrupted attachments, such as reduced communication quality and insufficient emotional nurturance in early emotional neglect<sup>64</sup>. Additionally, other studies<sup>54,63</sup> have reported adverse childhood experiences impact ToM impairments by examining the association between childhood trauma and brain function during ToM tasks in schizophrenia patients. The findings underscore the critical role of early stress in shaping ToM-related brain networks differently between healthy controls and schizophrenia patients, partially explaining the clinical and behavioral outcomes of the disorder. However, the current study did not test the roles of childhood adversity, further examination is needed. Taken together, our results confirm that, similar to the Western counterparts, Chinese schizophrenia patients show clear impairments in all four subconstructs of ToM<sup>66</sup>.

However, some results from the Yoni task in the current study were unexpected and inconsistent with previous studies. The overall score of schizophrenia patients was not significantly poorer than healthy controls after controlling the IQ. This could be due to the task complexity. Compared to VAMA, the Yoni task focuses on basic ToM skills and incorporates visual stimuli, such as simple cartoon faces, eye-gaze direction, and the mouth shape of the character<sup>67</sup>. In contrast, VAMA involves more facial perception, emotion recognition, and social information processing which may better discern differences in ToM performance between healthy controls and schizophrenia patients. Consequently, this result demonstrated that VAMA is more sensitive in identifying the ToM impairment among schizophrenia patients than the Yoni task.

Another highlight of the current study is the associations between clinical symptoms and error types. In particular, negative symptoms showed a significant correlation with “hypermentalizing” errors in schizophrenia patients, which is consistent with the results of previous studies<sup>14,68</sup>. Pelletier-Baldelli<sup>68</sup> proposed a model suggesting that disruptions in lower-level processes can have a “bottom-up” impact on social cognition and behavior. This model posits that specific changes in sensorimotor processes and social cognitive functions influence certain negative symptoms more than others. For instance, abnormalities in the parietofrontal cortical pathway, which is involved in monitoring personal space, could affect both the expression and interpretation of social cues. This disruption might negatively impact mentalisation abilities and, consequently, social motivation, while having minimal effect on speech production or facial expressions. Additionally, abnormalities in nearby parietofrontal circuits involved in eye movements and gestures may lead to different behavioral outcomes, contributing to various negative symptoms. Therefore, we speculated that the sensorimotor experiences of schizophrenia patients might heighten their propensity for overthinking or affect their interpretation of other’s mental states from social cues. Future studies could examine other subclinical groups, such as those with avolition or alogia<sup>68,69</sup>.

Notably, non-significant associations were observed between positive symptoms and other error types, which differs from previous studies’ findings<sup>4,22</sup>. This difference may be explained by cultural differences. Beck et al.<sup>70</sup> studied cross-cultural differences in ToM impairment between schizophrenia patients from China and Denmark, suggesting that the complexity of ToM tasks, such as those involving multiple individuals discussing the mental states of a third person, is likely to be influenced by cultural factors, particularly in complex scenarios where language aids in developing the cognitive representations necessary for such

reasoning<sup>71,72</sup>. Notably, language abilities encompass semantics, syntax, and pragmatics, with pragmatic skills being essential for the appropriate use and interpretation of language in communicative contexts, typically evaluated through measures derived from naturalistic conversations<sup>72</sup>. Similarly, Quesque et al. also found that differences between countries accounted for more than 20% of the variance in the results of the Faux Pas Test<sup>73</sup> and the Facial Affect Recognition Test<sup>74</sup>. Their findings highlighted notable distinctions between Chinese and Western participants, suggesting cultural variations in perspective-taking. Chinese individuals often prioritize understanding others’ perspectives over self-related ones, unlike Western participants. Our results provide new insights into how the clinical symptoms of Chinese schizophrenia patients relate to error types, particularly negative symptoms, but not positive symptoms. Additional research is needed to validate these findings.

The performance of individuals with high levels of SA in the subconstructs of ToM was lower than that of individuals with low levels of SA. Although this suggests that SA affects ToM, it should be noted that statistically significant between-group difference was only found in second-order cognitive ToM. A study conducted by Lui et al.<sup>75</sup> indicated that unaffected relatives of schizophrenia patients exhibiting negative schizotypal features were impaired in several cognitive abilities, such as visual—spatial memory, working memory, and verbal fluency. The VAMA task engages multiple cognitive functions—including memory, executive functioning, and social cognition where these cognitive processes are closely linked to ToM ability, especially higher-order ToM<sup>12</sup>. Hence, individuals with a high level of SA may also have some difficulties with mentalizing (i.e., second-order cognitive ToM) when a higher level of reasoning skills is required, due to difficulties in general higher-order cognitive abilities. This suggests that second-order cognitive ToM is more sensitive in identifying ToM impairment among at-risk individuals. This result is in line with another recent systematic review<sup>76</sup>, which also demonstrated the association between ToM abnormalities and negative schizotypy (Cohen’s  $d = -0.15$ ). Although this result differed from those studies that found significant deficits in ToM and/or emotion recognition performance among the SA extreme groups<sup>16,77,78</sup>, this could be related to our small sample size in Study 2 and the different measurements used in previous studies. Future studies with larger sample size are needed to replicate this study and better understand ToM impairment in this subclinical population.

This study has some limitations. First, confirmatory factor analysis (CFA) was not performed. Due to the small sample size of healthy young adults ( $n = 100$ ), we were unable to meet the minimum sample requirement for conducting the CFA (i.e., typically five to ten participants per item)<sup>79</sup>. Further research is needed to validate this version of VAMA. Second, the sample size of at-risk individuals (viz., those with high levels of SA) was small. Given their more subtle and less severe social cognitive impairments, a larger sample size is needed to clarify the nature of ToM impairment in this group. Third, only two groups of schizophrenia patients were included in this study. Future studies should include other groups (e.g., individuals with positive schizotypy, first-degree relatives, and ultra-high-risk individuals) to provide a more comprehensive picture of the extent of ToM impairments in schizophrenia spectrum disorders. Third, although the ToM impairments found in this study are likely to be related to deficiencies in the underlying neural mechanisms in schizophrenia patients and those with high levels of SA, our study only collected and reported behavioral data. To understand the neural mechanisms underlying ToM impairments in our participants, neuroimaging techniques, such as fMRI or fNIRS, are needed. Fourth is the exclusion of participants with a history of substance abuse or drug addiction. While this exclusion criterion was implemented to enhance the internal validity of the study by reducing confounding variables, it has the potential to inadvertently limit the generalizability of our findings. By excluding these individuals, our

sample may probably not fully represent the broader population of people with psychotic disorders. Future research should consider including participants with a history of substance abuse to capture the diversity and to refine the applicability of VAMA. Fifth, the current study did not collect convergent validity measures as in Canty et al.'s study<sup>26</sup>. Instead, we focus on examining the theory consistent known group difference as criterion validity, demonstrating that schizophrenia patients performed worse than healthy controls in VAMA. However, this approach did not fully rule out alternative explanations for the group difference, such as general cognitive deficits in schizophrenia. Future research should incorporate additional measures to establish the convergent validity of the adapted version VAMA. Last but not least, ToM deficits are a transdiagnostic issue beyond schizophrenia. More research is needed to understand how these deficits manifest across different clinical disorders.

Notably, the study has some implications. The adapted version of VAMA demonstrates potential as a tool for distinguishing between individuals with high SA and those with schizophrenia based on ToM impairments. In our study, schizophrenia patients exhibited ToM impairments across all subconstructs of VAMA, while individuals with high SA showed differences in only one subconstruct. This pattern suggests that VAMA can effectively capture the broader ToM deficits characteristic of schizophrenia, while also detecting more subtle impairments associated with high SA. VAMA could assess at-risk individuals by comparing their scores to standardized mean scores derived from clinical and nonclinical populations. Such comparisons could facilitate the early identification of at-risk individuals can inform targeted interventions and support. The VAMA's sensitivity to varying degrees of ToM impairment underscores its utility as a discriminative tool in both research and clinical practice. In addition, it is important to acknowledge the potential influence of dimensionality on the assessment of ToM constructs. As noted by previous researchers<sup>80,81</sup>, the novelty of employing VR and three-dimensional (3D) in therapeutic settings often highlights its potential motivational effects. However, the reported experiences with these technologies are mixed<sup>80</sup>. Future research examining the effects of 3D and 2D representations could provide valuable insights into how dimensionality impacts the user experience and cognitive engagement.

In conclusion, the newly adapted Chinese version of the VAMA is a reliable and ecologically valid tool. To the best of our knowledge, it is the only available tool in Chinese that can assess the four subconstructs of ToM and the three types of ToM errors in people with schizophrenia spectrum disorders. Understanding these specific impairments of ToM subconstructs is crucial as it can provide insights into the nuanced ways in which schizophrenia affects social cognition. Further research on the VAMA in other clinical or subclinical populations is needed to clarify the utility of the adapted VAMA in identifying at-risk individuals (i.e., psychosis or psychotic disorder) for early assessment and prevention.

## DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Received: 20 October 2024; Accepted: 4 March 2025;

Published online: 17 March 2025

## REFERENCES

1. Quesque, F. & Rossetti, Y. What do theory-of-mind tasks actually measure? theory and practice. *Perspect. Psychol. Sci.* **15**, 384–396 (2020).
2. Shamay-Tsoory, S. G., Harari, H., Aharon-Peretz, J. & Levkovitz, Y. The role of the orbitofrontal cortex in affective theory of mind deficits in criminal offenders with psychopathic tendencies. *Cortex* **46**, 668–677 (2010).
3. Perner, J. & Wimmer, H. "John thinks that Mary thinks that" attribution of second-order beliefs by 5- to 10-year-old children. *J. Exp. Child Psychol.* **39**, 437–471 (1985).
4. Canty, A. L., Neumann, D. L. & Shum, D. H. K. Using virtual reality to assess theory of mind subprocesses and error types in early and chronic schizophrenia. *Schizophr. Res. Cogn.* **10**, 15–19 (2017).
5. Green, M. F. et al. Social cognition in Schizophrenia: an NIMH workshop on definitions, assessment, and research opportunities. *Schizophr. Bull.* **34**, 1211–1220 (2008).
6. Saha, S., Chant, D., Welham, J. & McGrath, J. A systematic review of the prevalence of Schizophrenia. *PLOS Med.* **2**, e141 (2005).
7. Frith, C. D. Schizophrenia and theory of mind. *Psychol. Med.* **34**, 385–389 (2004).
8. Kwapil, T. R. Social anhedonia as a predictor of the development of schizophrenia-spectrum disorders. *J. Abnorm. Psychol.* **107**, 558–565, (1998).
9. Pieslinger, J. F., Wiskerke, J. & Igelström, K. Contributions of face processing, social anhedonia and mentalizing to the expression of social autistic-like traits. *Front. Behav. Neurosci.* **16**, 1046097 (2022).
10. Ku, Y.-L. & Lin, M.-T. Facial emotion recognition and its relationships with neurocognition and theory of mind in Schizophrenia. *Chin. J. Psychol.* **62**, 73–97 (2020).
11. Huang, Y.-L., Chen, T.-T., Dziobek, I. & Tseng, H.-H. Mentalizing in a movie for the assessment of social cognition (MASC): the validation in a taiwanese sample. *BMC Psychol.* **11**, 287 (2023).
12. Thibaut, E., Achim, A. M., Parent, C., Turcotte, M. & Cellard, C. A meta-analysis of the associations between theory of mind and neurocognition in schizophrenia. *Schizophr. Res.* **216**, 118–128 (2020).
13. Brüne, M. Theory of Mind" in Schizophrenia: A Review of the Literature. *Schizophr. Bull.* **31**, 21–42 (2005).
14. Thibaut, E., Rae, J., Raucher-Chéné, D., Bougeard, A. & Lepage, M. Disentangling the relationships between the clinical symptoms of Schizophrenia spectrum disorders and theory of mind: a meta-analysis. *Schizophr. Bull.* **49**, 255–274 (2023).
15. Andrzejewska, M., Wójcik, P., Domowicz, K. & Rybakowski, J. Emotion recognition and theory of mind in chronic schizophrenia: association with negative symptoms. *Arch. Psych. Psychother.* **19**, 7–12 (2017).
16. Monestès, J. L., Villatte, M., Moore, A., Yon, V. & Loas, G. Décisions en situation probabiliste et théorie de l'esprit dans la schizotypie. *L'Encéphale* **34**, 116–122 (2008).
17. Villatte, M., Monestès, J.-L., McHugh, L., Freixa i Baqué, E. & Loas, G. Assessing deictic relational responding in social anhedonia: a functional approach to the development of theory of mind impairments. *Int. J. Behav. Consult. Therapy* **4**, 360–373 (2008).
18. Corcoran, R., Mercer, G. & Frith, C. D. Schizophrenia, symptomatology and social inference: investigating "theory of mind" in people with schizophrenia. *Schizophr. Res.* **17**, 5–13 (1995).
19. Gooding, D. C., Johnson, M. & Peterman, J. S. Schizotypy and altered digit ratios: a second look. *Psychiatr. Res.* **178**, 73–78 (2010).
20. Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y. & Plumb, I. The "Reading the Mind in the Eyes" Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *J. Child Psychol. Psychiatr. Allied Discipl.* **42**, 241–251 (2001).
21. Abu-Akel, A. & Shamay-Tsoory, S. G. in *Social Cognition in Schizophrenia: From Evidence to Treatment* (eds. David L. Roberts & David L. Penn) 0 (Oxford University Press, 2013).
22. Fretland, R. A. et al. Theory of mind in schizophrenia: error types and associations with symptoms. *Schizophr. Res.* **162**, 42–46 (2015).
23. Vaskinn, A., Andersson, S., Østefjells, T., Andreassen, O. A. & Sundet, K. Emotion perception, non-social cognition and symptoms as predictors of theory of mind in schizophrenia. *Compr. Psychiatr.* **85**, 1–7 (2018).
24. Andrade-González, N. et al. Mentalizing errors in patients with Schizophrenia who received psychosocial rehabilitation: a case-control study. *Psychiatr. Quart.* **92**, 947–959 (2021).
25. Kessler, K. & Rutherford, H. The two forms of visuo-spatial perspective taking are differently embodied and subserve different spatial prepositions. *Front. Psychol.* **1**, 213 (2010).
26. Canty, A. L., Neumann, D. L., Fleming, J. & Shum, D. H. K. Evaluation of a newly developed measure of theory of mind: The virtual assessment of mentalising ability. *Neuropsychol. Rehabil.* **27**, 834–870 (2017).
27. Grainger, S. A., Rakunathan, V., Adams, A. G., Canty, A. L. & Henry, J. D. An assessment of age differences in theory of mind using the virtual assessment of mentalizing ability. *Aging Neuropsychol. Cogn.* **28**, 97–107 (2021).
28. Kobayashi, C., Glover, G. H. & Temple, E. Cultural and linguistic effects on neural bases of 'Theory of Mind' in American and Japanese children. *Brain Res.* **1164**, 95–107 (2007).
29. Wu, S. & Keysar, B. The effect of culture on perspective taking. *Psychol. Sci.* **18**, 600–606 (2007).

30. Punsoda-Puche, P., Barajas, A., Mamano-Grande, M., Jiménez-Lafuente, A. & Ochoa, S. Relationship between social cognition and premorbid adjustment in psychosis: a systematic review. *Schizophrenia* **10**, 36 (2024).
31. Sahl, A. C. et al. Theory of mind in schizophrenia: a comparison of subgroups with low and high IQ. *Nordic J. Psychiatr.* **77**, 329–335 (2023).
32. Vogeley, K. et al. Neural correlates of first-person perspective as one constituent of human self-consciousness. *J. Cogn. Neurosci.* **16**, 817–827 (2004).
33. Brown, S. The “Who” system of the human brain: a system for social cognition about the self and others. *Front. Hum. Neurosci.* **14**, 224 (2020).
34. Dziobek, I. et al. Introducing MASC: a movie for the assessment of social cognition. *J. Autism Dev. Disord.* **36**, 623–636 (2006).
35. Van de Vijver, F. J. R. *Advanced Research Methods and Statistics for the Behavioral and Social Sciences.*, (Cambridge University Press., 2019).
36. Quesque, F. et al. Does culture shape our understanding of others’ thoughts and emotions? An investigation across 12 countries. *Neuropsychology* **36**, 664–682 (2022).
37. Yang, K. S. “The psychological transformation of the Chinese people as a result of societal modernization,” in *The Handbook of Chinese Psychology*. 479–498 (Oxford University Press, 1996).
38. Hu, X. et al. Do Chinese traditional and modern cultures affect young adults’ moral priorities? *Front. Psychol.* **9**, 1799 (2018).
39. Cavieres, A., Acuña, V., Arancibia, M. & Escobar, C. Advances in the ecological validity of research on social cognition in Schizophrenia: a systematic review of the literature. *Schizophr. Res. Cogn.* **39**, 100333 (2025).
40. Jarsch, M. et al. Development of the basel version of the awareness of social inference test – theory of mind (BASIT-ToM) in healthy adults. *J. Neuropsychol.* **17**, 125–145 (2023).
41. Wang, J. et al. Acceptance of COVID-19 Vaccination during the COVID-19 Pandemic in China. *Vaccines* **8**, 482 (2020).
42. Liu, J., Liu, M., Zhang, S., Ma, Q. & Wang, Q. Intent to have a second child among Chinese women of childbearing age following China’s new universal two-child policy: a cross-sectional study. *BMJ Sex. Reprod. Health* **46**, 59 (2020).
43. Chan, R. C. K. et al. A Study of trait anhedonia in non-clinical Chinese samples: evidence from the Chapman scales for physical and social anhedonia. *PLOS One* **7**, e34275 (2012).
44. Zhang, Y.-j et al. Altered neural mechanism of social reward anticipation in individuals with schizophrenia and social anhedonia. *Eur. Arch. Psychiatr. Clin. Neurosci.* **273**, 1029–1039 (2023).
45. Pu, J. et al. Differential manifestations of anhedonia in people with social anhedonia and subsyndromal depression. *Asian J. Psychiatr.* **100**, 104188 (2024).
46. Zhang, R.-t et al. Affective forecasting in individuals with social anhedonia: The role of social components in anticipated emotion, prospection and neural activation. *Schizophr. Res.* **215**, 322–329 (2020).
47. Kay, S. R., Opler, L. A. & Lindenmayer, J.-P. The positive and negative syndrome scale (PANSS): rationale and standardisation. *Brit. J. Psychiatr.* **155**, 59–65 (1989).
48. Shu L., S. T., Li X., Kong Q. The reliability, validity of PANSS and its implication. *Chin. Mental Health J.*, **18**, 45–47 (2004)
49. Dai, X., Ryan, J. J., Paolo, A. M. & Harrington, R. G. Factor analysis of the mainland chinese version of the wechsler adult intelligence scale (WAIS-RC) in a brain-damaged sample. *Int. J. Neurosci.* **55**, 107–111 (1990).
50. Eckblad, M. L., Chapman, L. J., Chapman, J. P., Mishlove, M. *The Revised Social Anhedonia Scale. Unpublished Test*, (University of Wisconsin, 1982).
51. Ho, K. K. Y. et al. Theory of mind impairments in patients with first-episode schizophrenia and their unaffected siblings. *Schizophr. Res.* **166**, 1–8 (2015).
52. Shamay-Tsoory, S. G., Aharon-Peretz, J. & Levkovitz, Y. The neuroanatomical basis of affective mentalizing in schizophrenia: Comparison of patients with schizophrenia and patients with localized prefrontal lesions. *Schizophr. Res.* **90**, 274–283 (2007).
53. Wang, Y.-y et al. Theory of mind impairment and its clinical correlates in patients with schizophrenia, major depressive disorder and bipolar disorder. *Schizophr. Res.* **197**, 349–356 (2018).
54. Vai, B. et al. Adverse childhood experiences differently affect theory of mind brain networks in schizophrenia and healthy controls. *J. Psychiatr. Res.* **172**, 81–89 (2024).
55. McDonald, S., Flanagan, S., Rollins, J. & Kinch, J. TASIT: a new clinical tool for assessing social perception after traumatic brain injury. *J. Head Trauma Rehabil.* **18**, 219–238 (2003).
56. Bell, M. D., Fiszdon, J. M., Greig, T. C. & Wexler, B. E. Social attribution test—multiple choice (SAT-MC) in schizophrenia: comparison with community sample and relationship to neurocognitive, social cognitive and symptom measures. *Schizophr. Res.* **122**, 164–171 (2010).
57. Pinkham, A. E., Harvey, P. D. & Penn, D. L. Social cognition psychometric evaluation: results of the final validation study. *Schizophr. Bull.* **44**, 737–748 (2017).
58. Tsui, H. K. H. et al. Reliability of theory of mind tasks in Schizophrenia, ASD, and nonclinical populations: a systematic review and reliability generalization meta-analysis. *Neuropsychol. Rev.* 1–17 (2024).
59. DeVellis, R. F. Classical test theory. *Med. Care* **44**, 31–44 (2006).
60. Murray, K. et al. A new test of advanced theory of mind: The “Strange Stories Film Task” captures social processing differences in adults with autism spectrum disorders. *Aut. Res.* **10**, 1120–1132 (2017).
61. Weng, Y., Lin, J., Ahorsu, D. K. & Tsang, H. W. H. Neuropathways of theory of mind in schizophrenia: a systematic review and meta-analysis. *Neurosci. Biobehav. Rev.* **137**, 104625 (2022).
62. Brady, R. O. et al. Cerebellar-cortical connectivity is linked to social cognition trans-diagnostically. *Front. Psychiatr.* **11**, 573002 (2020).
63. Quidé, Y. et al. Childhood trauma-related alterations in brain function during a theory-of-mind task in schizophrenia. *Schizophr. Res.* **189**, 162–168 (2017).
64. Kincaid, D. et al. An investigation of associations between experience of childhood trauma and political violence and theory of mind impairments in schizophrenia. *Psychiatr. Res.* **270**, 293–297 (2018).
65. Rajkumar, R. P. Childhood attachment and schizophrenia: The “attachment-developmental-cognitive” (ADC) hypothesis. *Med. Hypoth.* **83**, 276–281 (2014).
66. Weinreb, S., Li, F. & Kurtz, M. M. A meta-analysis of social cognitive deficits in schizophrenia: does world region matter? *Schizophr. Res.* **243**, 206–213 (2022).
67. Tin, L. N. W. et al. High-functioning autism patients share similar but more severe impairments in verbal theory of mind than schizophrenia patients. *Psychol. Med.* **48**, 1264–1273 (2018).
68. Pelletier-Baldelli, A. & Holt, D. J. Are negative symptoms merely the “Real World” consequences of deficits in social cognition? *Schizophr. Bull.* **46**, 236–241 (2019).
69. Strauss, G. P., Bartolomeo, L. A. & Harvey, P. D. Avolition as the core negative symptom in schizophrenia: relevance to pharmacological treatment development. *NPJ Schizophr.* **7**, 16 (2021).
70. Beck, K. I. et al. Cross-cultural comparison of theory of mind deficits in patients with schizophrenia from China and Denmark: different aspects of ToM show different results. *Nordic J. Psychiatr.* **74**, 366–373 (2020).
71. de Villiers, J. The interface of language and theory of mind. *Lingua* **117**, 1858–1878 (2007).
72. Milligan, K., Astington, J. W. & Dack, L. A. Language and theory of mind: meta-analysis of the relation between language ability and false-belief understanding. *Child Dev.* **78**, 622–646 (2007).
73. Stone, V. E., Baron-Cohen, S. & Knight, R. T. Frontal lobe contributions to theory of mind. *J. Cogn. Neurosci.* **10**, 640–656 (1998).
74. Ekman, P. & Friesen, W. V. *Pictures of Facial Affect* (Consulting Psychologists Press, 1976).
75. Lui, S. S. Y. et al. Clustering of schizotypal features in unaffected first-degree relatives of schizophrenia patients. *Schizophr. Bull.* **44**, S536–S546 (2018).
76. Bora, E. Theory of mind and schizotypy: a meta-analysis. *Schizophr. Res.* **222**, 97–103 (2020).
77. Germine, L. T., Garrido, L., Bruce, L. & Hooker, C. Social anhedonia is associated with neural abnormalities during face emotion processing. *NeuroImage* **58**, 935–945 (2011).
78. Pflum, M. J. & Gooding, D. C. Context matters: social cognition task performance in psychometric schizotypes. *Psychiatr. Res.* **264**, 398–403 (2018).
79. Mundfrom, D. J., Shaw, D. G. & Ke, T. L. Minimum sample size recommendations for conducting factor analyses. *Int. J. Test.* **5**, 159–168 (2005).
80. Dellazizzo, L., Potvin, S., Luigi, M. & Dumais, A. Evidence on virtual reality-based therapies for psychiatric disorders: meta-review of meta-analyses. *J. Med. Internet Res.* **22**, e20889 (2020).
81. Vass, E. et al. Virtual reality-based theory of mind intervention in Schizophrenia: Preliminary efficacy results. *Compr. Psychiatr.* **119**, 152350 (2022).

## ACKNOWLEDGEMENTS

We thank all of the participants for participating in this project. We also thank the School of Design (AV Team) and the Department of Computing (Prof. Peter H.F. Ng) of The Hong Kong Polytechnic University for providing professional technical support in producing the video stimuli. Our work was funded by the National Natural Science Foundation of China/RGC Joint Research Scheme (ref: N\_PolyU540/20; 32061160468) and the General Research Funding (ref: 15607622). David H.K. Shum is supported by the Yeung Tsang Wing Yee and Tsang Wing Hing Endowed Professorship, and Raymond C.K. Chan is supported by the Philip K. H. Wong Foundation.

## AUTHOR CONTRIBUTIONS

Yuan Cao and Winnie W.Y. So are the co-first authors of this paper. Dr. Yuan Cao: Conceptualization, investigation, methodology, project administration, and writing—review and editing. Ms. Winnie W.Y. So: Data analysis, writing—original draft, review and editing, stimuli production, methodology, project administration. Dr. Yi Wang: Software, project administration, and writing—review and editing. Ms. Ding-ding Hu: Data collection, software, and writing—review. Mr. Jie Gu: Data collection and writing—review. Ms. Yi Xie: Data collection and writing—review. Prof. Raymond C. K. Chan:

Funding acquisition, supervision, and writing—review and editing. Prof. David H. K. Shum: Funding acquisition, supervision, and writing—review and editing, visualization.

## 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/s41537-025-00594-1>.

**Correspondence** and requests for materials should be addressed to David H. K. Shum.

**Reprints and permission information** is available at <http://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-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/>.

© The Author(s) 2025