Research Article

Psychological Science
2023, Vol. 34(3) 345–357

DOI: 10.1177/09567976221137313
www.psychologicalscience.org/PS
SAGE

The Development of Spatial Cognition and Its Malleability Assessed in Mass Population via a Mobile Game

Shan Xu[1], Yiying Song[1], and Jia Liu[2,3]
[1]Faculty of Psychology, Beijing Normal University; [2]Department of Psychology, Tsinghua University; and [3]Tsinghua Laboratory of Brain and Intelligence, Tsinghua University

Abstract
Spatial cognition is a fundamental aspect of human intelligence, but our understanding of its developmental trajectory across the life span is limited. Here, we applied game-based assessment on mobile devices to engage a large sample from China ($N = 216{,}713$) with a wide age range (from under 10 years old to above 60) in multiple participations of a mental rotation task, a typical measure of spatial cognition. We found that spatial ability developed asynchronously with its malleability. Whereas mental rotation performance peaked at the age of 28, with males performing better than females, the effect of training from repeated participation peaked at 18, probably laying the foundation for the development of spatial ability. In contrast, children showed particularly low malleability, and a follow-up experiment revealed that the underdeveloped ability of mirror-image discrimination likely hindered the malleability of spatial cognition during this period. The intermingled relation of ability and malleability illustrates dynamics in the development of spatial cognition, inviting broad research on the development of other cognitive functions.

Keywords
spatial cognition, cognitive development, malleability, mirror-image discrimination, game-based assessment

Received 3/13/22; Revision accepted 9/23/22

Spatial cognition is a core aspect of human intellectual ability. It contributes to various aspects of human life, from mundane activities such as navigating in a complex urban environment or assembling an IKEA chair, all the way up to advanced technological feats such as charting the complex three-dimensional structures of protein molecules or designing new models of aircraft engine. During development, spatial abilities can predict school performance (Casey et al., 1995; Halpern et al., 2007) and are particularly associated with academic achievements and attainment in science, technology, engineering, and mathematics (STEM; Newcombe & Frick, 2010; Wai et al., 2009). Consequently, improving spatial ability through training is of great significance. However, little is known about the developmental trajectory of spatial ability across the life span and, more importantly, the critical period for its training. For instance, at which stage of development will training on spatial ability be most effective? Will elementary school students benefit from repetitive training on spatial ability?

Though there is consensus that infants in their first several months show evidence of rudimentary mental rotation and 3D perception ability (Christodoulou et al., 2016; Hespos & Rochat, 1997) and that spatial ability improves with age from childhood (Vander Heyden et al., 2016), our understanding of the development of spatial ability across the life span is not yet comprehensive. For example, the maturation of spatial ability from childhood to adulthood and its decline at older age have not been systematically measured and depicted. Filling this gap requires testing all age groups with the same task and collecting a sufficiently large number of participants at each developmental stage.

More importantly, it is critical to know not only how spatial ability changes with age but also when spatial ability is most sensitive to training during development

Corresponding Author:
Yiying Song, Faculty of Psychology, Beijing Normal University
Email: songyiying@bnu.edu.cn

in order to efficiently improve spatial ability with training. Note that the malleability of a cognitive ability usually develops asynchronously with the ability itself. For instance, the malleability of second-language acquisition peaks long before language proficiency reaches its peak (Snow & Hoefnagel-Höhle, 1978). In the field of spatial cognition, previous studies have reported malleability in both childhood and adulthood (Uttal, Meadow, et al., 2013) with a variety of training tasks, such as simple practice on mental rotation (Meneghetti et al., 2016) and training on motor skills related to spatial cognition (Moreau et al., 2012). However, researchers have not investigated how the malleability of spatial ability changes with age, which is also vital for STEM-related education.

To address this issue, in the present study, we depicted the developmental trajectory of spatial ability and further characterized the developmental trajectory of its malleability to locate the developmental stage at which individuals are most sensitive to training. To overcome the drawback that systematic investigation of malleability development requires a large number of participants of different ages to perform the same task multiple times, we adopted a new approach to the investigation of cognitive development through online data collection with mobile devices. By allowing researchers to access large populations with unprecedented ease, online data collection has been fruitful in effectively accumulating a large amount of data, remarkably enlarging the power of psychological studies. The validity of this approach has been demonstrated by replicating various classic laboratory findings (Germine et al., 2012; McGraw et al., 2000) as well as providing novel insights unrevealed by traditional laboratory approaches (Fortenbaugh et al., 2015; Hartshorne & Germine, 2015; Nosek et al., 2009). Further, the introduction of game-based assessment into online data collection has changed often time-consuming psychological studies into motivating recreational activities (Spiers et al., 2016), and it is possible to have participants repeatedly participate in the same task. The change in performance across multiple repetitions reflects the effect of training, which makes this approach perfect for measuring the malleability of cognitive functions.

Specifically, the present study packed a representative spatial task, mental rotation, into a mobile game embedded in a popular Chinese instant messaging (IM) mobile app (WeChat). In the context of a famous fairy tale, the Three Piglets, a series of trials were presented in which the participants were shown several 3D objects, and they needed to choose the one that had a surface matching a target shape (Fig. 1a). As with typical mental rotation tasks, the game requires the ability of mental manipulation on the representation of objects and the ability to predict the appearance of a given object once it has been rotated into a different orientation. The number of trials correctly completed in the game was considered an index of mental rotation ability. In addition, because a substantial number of participants played this game multiple times, the increase rate of scores with repetition of the plays indicated training-induced malleability of mental rotation. In this way, the present study collected data from an unprecedentedly large sample (N = 216,713) with a wide age range (from under 10 years old to above 60) and thus was able to depict the developmental trajectories of mental rotation ability and, more importantly, its training-induced malleability.

Statement of Relevance

Spatial cognition is a well-established predictor of achievement and attainment in science, technology, engineering, and mathematics education. To determine the developmental trajectory of spatial cognition and the best time to improve spatial skills, we collected data in the present study from a large population (N > 200,000) while they played a mobile game involving mental rotation multiple times. We found that 28-year-olds achieved the highest score in the game, whereas 18-year-olds received the greatest benefit from training. The asynchrony in development suggests that malleability may lay the foundation for developing spatial ability. Interestingly, children under the age of 10 showed particularly low malleability of spatial cognition, which was probably hindered by the underdeveloped ability of mirror-image discrimination in this period. The present study showcases the potential of online game-based data collection for unveiling the dynamic interaction of ability and malleability during the development of spatial cognition.

Open Practice Statement

The study was not preregistered. The data and code for analyses have been made publicly available via GitHub and can be accessed at https://github.com/shan0903/ThreePiglet2022.

Method

Participants

In a designated period, a total of 216,713 participants played the game via the IM mobile app WeChat, and 194,994 of them obtained nonzero scores in at least one

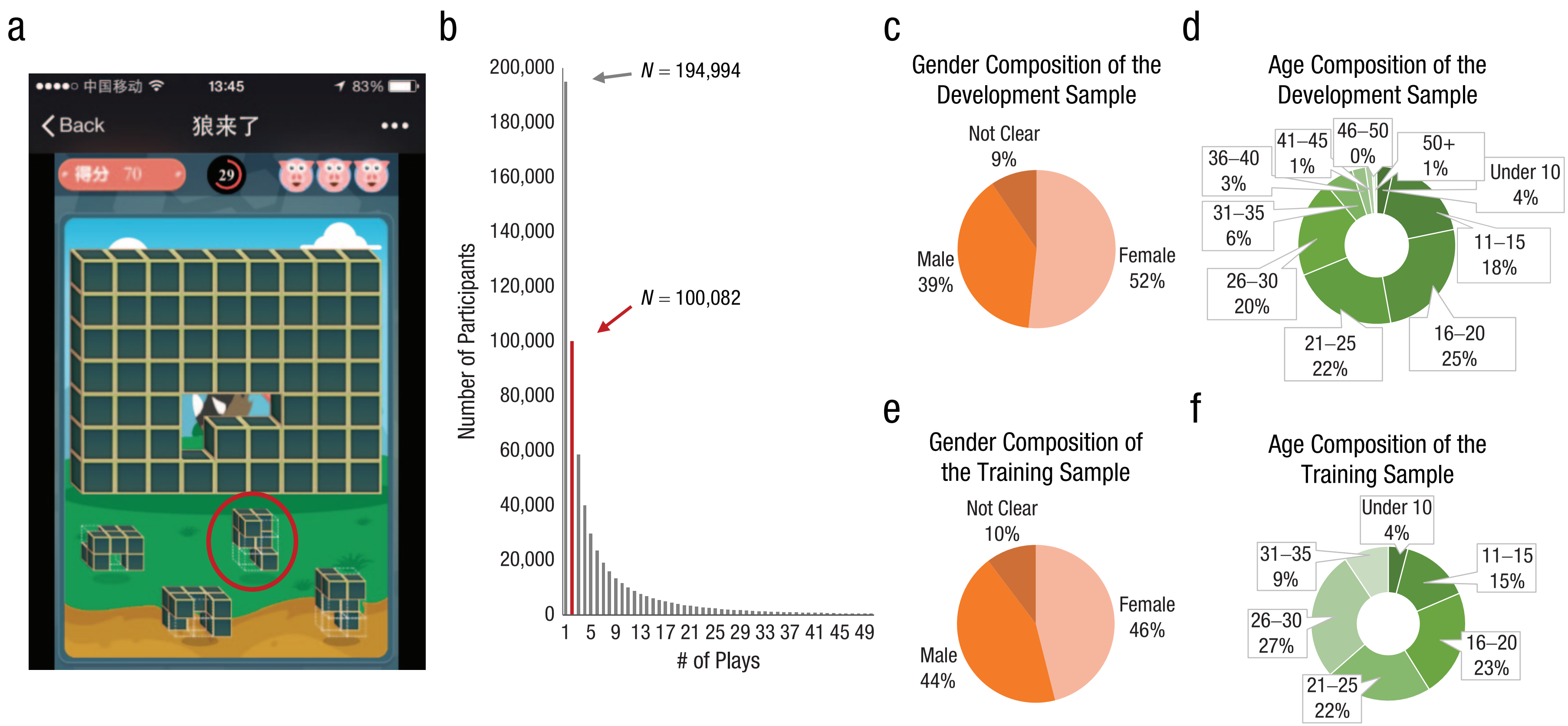

Fig. 1. Task and demographic information of the sample. (a) Screenshot of the game-based assessment. Participants had to choose which one of four 3D objects fitted the opening on the wall. The red circle marks the correct choice, whose right surface fits the opening. (b) In total, 194,994 participants played the game and received valid scores, and 100,082 played at least twice. Among them, 23,347 participants provided age information (i.e., the development sample). (c) Gender composition of the development sample. (d) Age composition of the development sample. (e) Gender composition of the training sample. (f) Age composition of the training sample. The training sample consisted of 1,707 participants (age groups: under 10, 11–15, 16–20, 21–25, 26–30, and 31–35) who played at least 20 times after the practice trial and provided age information.

play. In total, 623,179 nonzero data points were generated, indicating that each participant played the game 3.2 times on average. We considered only the nonzero plays in the following analyses. The number of plays varied among participants: Most participants played more than once, 13 participants played more than 200 times, two played more than 500 times, and one played more than 700 times (Fig. 1b). Because the game did not include any dedicated practice trial, the first nonzero play was considered practice, and the score of the second nonzero play was taken as the index of mental rotation performance. More than 100,000 participants (100,082) in our data set had a valid index of performance (Fig. 1b). Among these participants, 23,347 provided age information (the development sample; see Figs. 1c and 1d for gender and age composition, respectively). Participants reported their age by selecting one of 12 age groups (starting with under 10 years old, they could select age groups in 5-year windows up to 60+ years old). We combined the last three age groups (50–55, 55–60, and 60+) into a 50+ age group because of the small number of participants in these three groups. Informed consent was obtained from all participants before they took part in the assessment.

To examine malleability of mental rotation performance, we focused on participants who had at least 20 nonzero records and provided age information, resulting in a sample size of 1,851 participants. We chose to include all participants who played the game at least 20 times to achieve a balance between feasible sample size and length of training. Notably, very few of these participants fell into the last four age groups. Therefore, we focused on the first six age groups (under 10, 11–15, 16–20, 21–25, 26–30, and 31–35) for analyses of the malleability of spatial ability (the training sample, n = 1,707; see Figs. 1e and 1f for gender and age composition, respectively).

In addition, a separate sample of children (n = 336, ages 6–14) was tested with a pen-and-paper mental rotation task (Vandenberg & Kuse, 1978). These participants received gifts as compensation for their time. Among them, four were excluded from analyses because of the lack of age information, and one was excluded for being the only participant who was 14 years old.

The sample of the game-based assessment was a convenience sample. The experimental protocol was approved by the Beijing Normal University Institutional Review Board.

Game-based assessment of mental rotation

We used a game-based assessment to measure participants' mental rotation ability. The game was embedded in a major Chinese mobile IM application (WeChat) and was freely available to any WeChat user around the world. The model of mobile phones and the screen size used by participants were not specified, but the most popular screen size at the time point was about 5 in. to 6 in. The game contained a sequence of puzzles (trials) of gradually increasing difficulty. In each trial, the participants had 45 s to view an opening in a wall depicted on the screen and to choose the matching 3D object out of four alternatives to fit in the opening. Each of the given choices was embedded in a hypothetical 2 × 2 × 2 (the first four trials), 2 × 2 × 3 (the 5th–16th trials), or 3 × 3 × 3 (from the 17th trials on) block space. In one of its six surfaces, the correct choice had blocks of exactly the same layout as the opening. The surfaces and the block space of each choice were outlined in the game with white dashed lines (Fig. 1a). Besides the change in object complexity, two of the distractors in Trials 15 and 16 had one of their surfaces presenting the mirrored layout of the opening in the wall and thus could not fit in the opening. This challenge was introduced to make the game more engaging and challenging without an a priori hypothesis of the impact on the malleability of mental rotation.

After a response was made, the correct choice was shown to participants by highlighting the surface fitting the opening with a red circle. A trial was counted as failed if participants made a wrong choice or did not make any response in the time limit of 45 s. The game terminated when participants failed three trials in total, and the number of trials correctly completed was recorded as the score in one play of the game, serving as an index of participants' mental rotation performance. After the game, participants received feedback on their own performance relative to other players.

Because of the game-based setting of the assessment, a substantial number of participants played the game multiple times. The change in performance with multiple plays indicated the training-induced malleability of mental rotation. For each participant in the training sample, we modeled his or her learning curve with a natural logarithmic function: $y = a + bln(x)$, where y was the score of play x ($x \in [1, 20]$). In this function, parameter b was a scaling factor denoting the increase rate of the score with repeated plays, which served as an index of the training effect (i.e., malleability).

We further estimated the peak age of mental rotation performance and that of the malleability, respectively, using a bootstrapped resampling procedure (see also Hartshorne & Germine, 2015). We used the estimated parameter b as the index of malleability and the first nonzero play after the practice of each participant as the index of performance. For each index, we drew 2,500 bootstrapped samples with replacement, each with the same number of participants from each age group as in the original sample, and the age group with the highest mean value of the index was identified as the peak age in each bootstrapped sample. This generated a distribution of the peak age groups for the index. The midpoint of each age group was then used to calculate the median and standard deviation of this distribution, which were considered as the estimate of the peak age and its standard error.

Pen-and-paper assessment

A separate cohort of children was tested with a pen-and-paper mental rotation test adapted from Vandenberg and Kuse (1978). The test consisted of two parts. Mental rotation ability was examined in the first part with 3D cube stimuli and in the second with 2D letter stimuli. Task difficulty was lowered to meet the cognitive ability of elementary school students by changing the original two-out-of-four choice questions into one-out-of-four ones. The participants were given 12 min for the first part and 5 min for the second part. Each part consisted of 24 questions, and each question showed one target, along with one correct choice and three distractors. Half of the questions contained mirrored distractors, as two of the three distractors were mirrored structures of the target, and the third was structurally different; the rest of the questions did not contain mirrored distractors, as all the three distractors were structurally different from the target but not of its mirrored structure. The questions without mirrored distractors were considered to require mental rotation ability, and those with mirrored distractors were considered to require both mental rotation and mirror-image discrimination. Therefore, the mirror-image discrimination scores for the cube and letter tasks were calculated, respectively, as the standardized residuals of the accuracies for questions with mirrored distractors, regressing out the accuracies for the questions without mirrored distractors.

To further examine whether there is any turning point in the developmental trajectory of each task, we assumed that each performance trajectory across age could be presented as a piecewise linear function with one knot. Therefore spline regression was conducted, and the data of each task in the pen-and-paper assessment were fitted using nonlinear least squares based on scipy.optimize.curve_fit (*SciPy* Version 1.5.2; Virtanen et al., 2020) to Equation 1, where x denotes age, y is the performance of the corresponding age, x_0 is the location of the turning point at which performance at this age in the task started to develop at a different rate than at previous ages, and y_0 is performance at the turning point:

$$y = \begin{cases} y_0 + \beta_1 (x - x_0) \text{ if } x < x_0 \\ y_0 + \beta_2 (x - x_0) \text{ if } x \geq x_0. \end{cases} \tag{1}$$

All the inferential statistical tests reported in this study were two-tailed.

Results

The developmental trajectory of mental rotation ability

Figure 2a depicts the developmental trajectory of mental rotation performance based on 23,347 participants. Performance increased rapidly from the under-10 age group to adolescence and adult groups, reached peak level from the 21–25 to 31–35 age groups, and then declined steadily afterward. A one-way analysis of variance (ANOVA) revealed that the main effect of age was significant, $F(9, 23337) = 21.69$, $p < .001$, $\eta_p^2 = .008$. Pairwise comparison between neighboring age groups (Fig. 2a) revealed that performance increased significantly from the under 10 to 11–15 (by 42.26%), 11–15 to 16–20 (by 5.05%), and 16–20 to 21–25 (by 4.85%) age groups and then declined from the 31–35 to 36–40 age group (by 11.87%, ps < .05, false-discovery-rate [FDR] corrected for nine comparisons).

Among the 23,347 participants, 21,131 provided gender information: 9,075 (42.9%) were males, and 12,056 (57.1%) were females. We drew the developmental trajectory for each gender separately to examine whether the trends were similar for males and females (Fig. 2b). A two-way ANOVA with age and gender as between-subject factors revealed significant main effects of age, $F(9, 21111) = 23.32$, $p < .001$, $\eta_p^2 = .010$, and gender, $F(1, 21111) = 26.77$, $p < .001$, $\eta_p^2 = .001$, with males outperforming females in this task. The interaction between these two factors was significant, $F(9, 21111) = 2.62$, $p = .005$, $\eta_p^2 = .001$. Simple effect analyses revealed that the males outperformed the females from the 11–15 age group to the 36–40 age group (ps < .001), but the males and the females showed similar mental rotation performance in childhood (under 10) and after middle adulthoods (from 41–45 to 50+; ps > .17). Notably, the developmental trajectory across genders resembled the developmental trajectory of each gender (Fig. 2a), suggesting that males and females showed similar developmental trajectories. This observation was confirmed after we controlled for the between-genders difference in absolute levels of mental rotation performance (Fig. 2c). That is, we standardized scores of mental rotation performance of each gender into z scores by subtracting the mean scores of the respective gender and then dividing each remainder by the standard deviation. The standardized z scores were submitted to the two-way ANOVA, which no longer revealed a significant interaction between age and gender, $F(9, 21111) = 1.34$, $p = .21$. That is, the developmental trajectories of males and

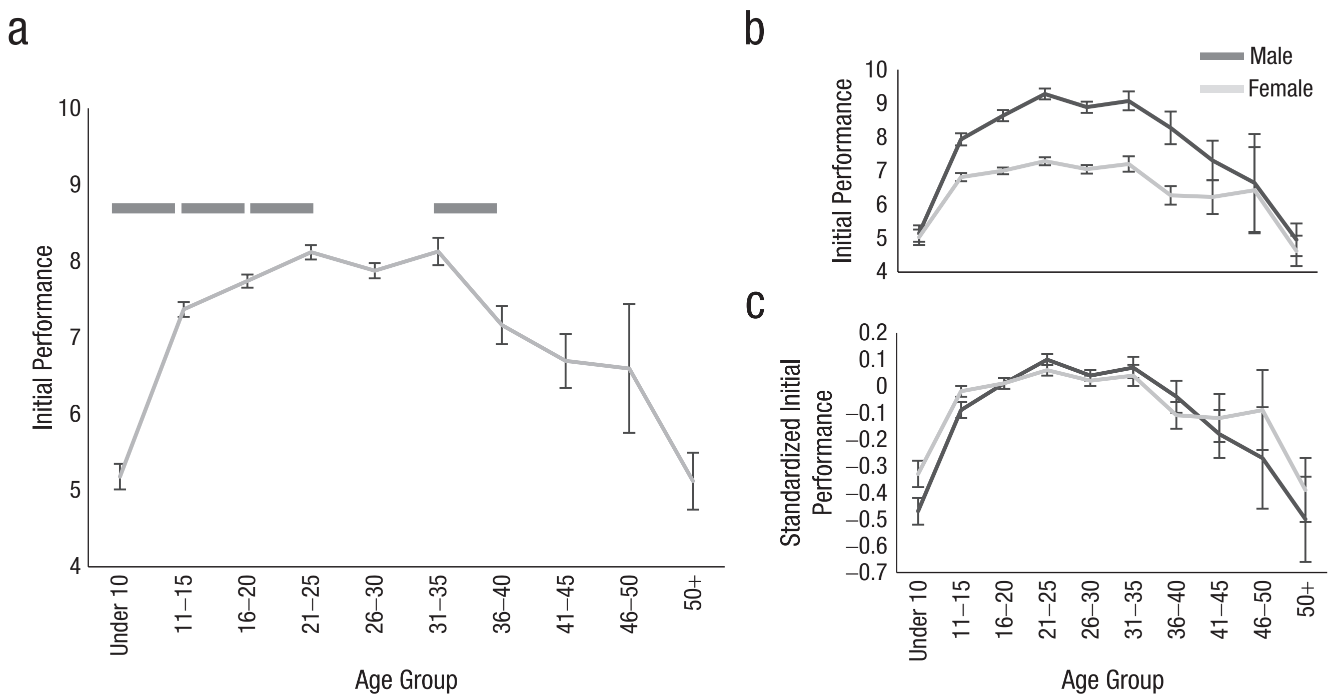

Fig. 2. Developmental trajectory of mental rotation ability. (a) Mean performance of each age group. The horizontal bars indicate significant differences between neighboring age groups (false-discovery-rate corrected, $q = .05$). (b) Development trajectories of performance, separately for males and females. (c) Standardized performance for the male and female participants. In all panels, error bars depict standard errors.

females were similar, although mental rotation performance was higher for males than females during adolescence and early adulthood.

The development trajectory of the malleability of mental rotation ability

To improve a person's spatial ability, we need to know not only when he or she would achieve high performance during development but also when he or she is most ready to learn, that is, most sensitive to training. In our study, a substantial number of participants played the game more than once. Consequently, each of these participants generated a series of scores, which revealed how their performance changed as a function of the number of times they played (i.e., malleability). The average learning curve across all participants showed that the growth of performance tended to be stable after 20 repetitions (see Fig. S1 in the Supplemental Material available online). To achieve a balance between length of training and feasible sample size, we chose all participants who played the game at least 20 times ($n = 1,851$), including those without age information, and described the learning curve of these participants. The learning curve showed a quick initial climb followed by gradual growth, which resembles a natural logarithmic function (Fig. 3a).

Then, we examined how the malleability of mental rotation changed with age. We selected participants who played the game at least 20 times and provided age information. Because the participant number was small in the age groups after 35 years old ($n = 144$), we limited our analysis to participants younger than 35 years old (i.e., the training sample, $n = 1,707$). As shown in Figure 3b, the learning curves of each age group in the training sample all showed a logarithmic-like shape, and thus the scaling factor b of the fitted natural logarithmic function served as the index of malleability. The malleability of mental rotation showed a rapid increase from childhood to early adulthood (Fig. 3c), reaching peak level from the 16–20 age group and remaining high until the 31–35 age group. A one-way ANOVA revealed a significant main effect of age on malleability, $F(5, 1701) = 5.73$, $p < .001$, $\eta_p^2 = .02$. Pairwise comparison between neighboring age groups (Fig. 3c) revealed that malleability increased significantly from under-10 to 11–15 and from 11–15 to 16–20 age groups (ps < .05, FDR-corrected for five comparisons) but not between later neighboring groups (ps > .3, FDR-corrected). We also performed these analyses for learning curves of 10 and 30 repetitions, respectively, and observed similar development trajectories of malleability (see Fig. S2 in the Supplemental Material).

In addition, to confirm the validity of the scaling factor as an index of malleability, we calculated the

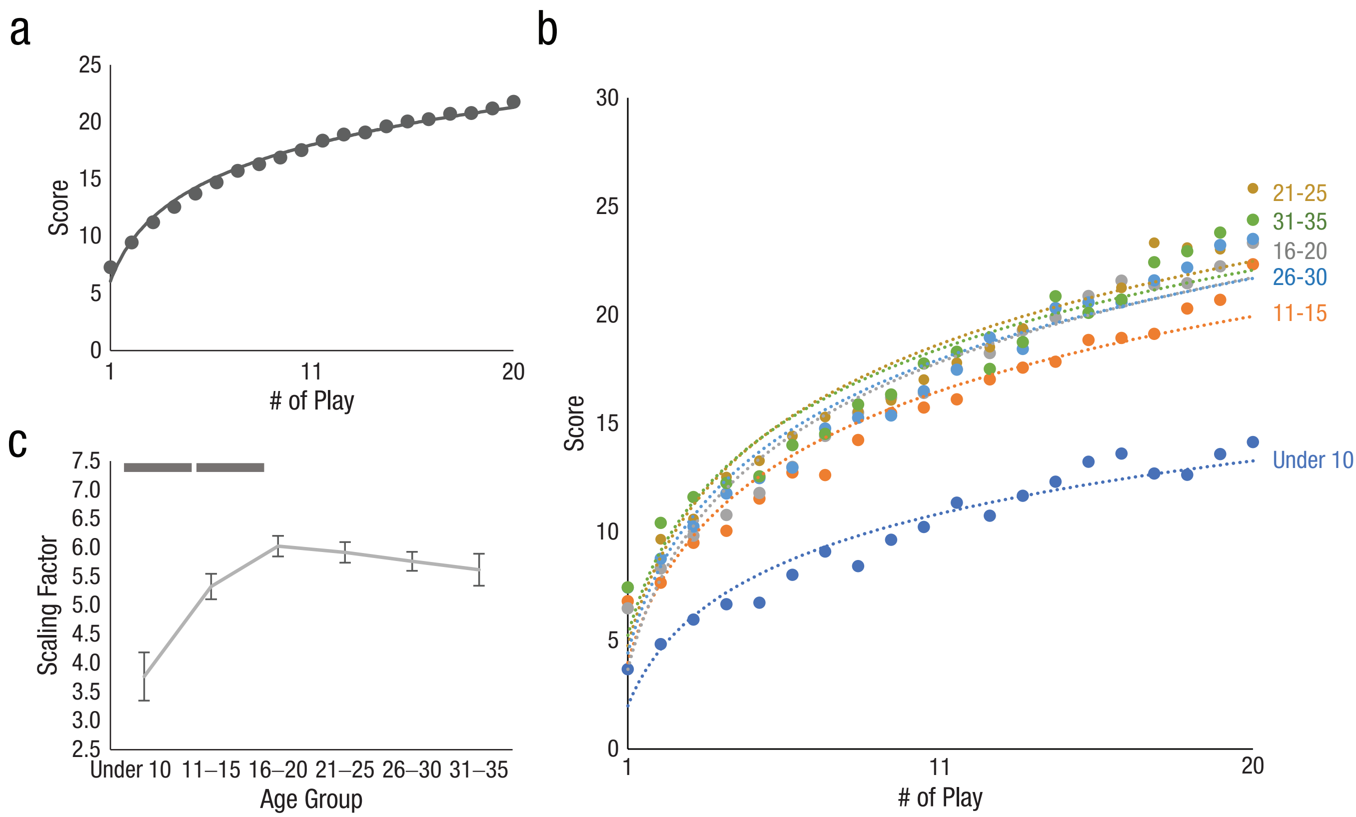

Fig. 3. Malleability of mental rotation. (a) Learning curve of all participants across age groups. (b) Learning curves of each age group. The dashed lines depict the natural logarithmic function fitted to each age group. (c) Age-related changes in malleability indexed by the scaling factor of the fitted natural logarithmic function. Error bars denote standard errors. Horizontal bars denote significant differences between neighboring age groups (false-discovery-rate corrected, q = .05).

difference between the maximum score that a participant achieved in his or her first 20 plays and his or her performance of the first play after the practice as another index of malleability (see Fig. S3 in the Supplemental Material). Exactly the same pattern of development was obtained, confirming the validity of the scaling factor as an index of malleability. For brevity, we used the scaling factor in the following analyses.

The developmental asynchrony between behavioral performance and malleability of spatial cognition

Visual inspection of Figures 2a and 3c suggests that the developmental trajectory of malleability peaked earlier (~16–20 years old) than did behavioral performance (after 21–25 years old). To quantify the developmental asynchrony between malleability and performance of mental rotation, we generated bootstrapped estimates to locate the age at which behavioral performance and malleability peaked, respectively. As shown in Figure 4a, the estimates from 2,500 bootstrapped samples showed a clear dissociation between the distributions of the peak age of performance and malleability, where the malleability peaked at around 18 years old (*Mdn* = 18, *M* = 21.04, 95% confidence interval [CI] = [20.87, 21.22], Fig. 4b), about 10 years earlier than the peak age of behavioral performance (*Mdn* = 28, *M* = 28.13, 95% CI = [27.97, 28.33]). Similar results were observed for learning curves based on 10 and 30 repetitions (see Fig. S2 in the Supplemental Material).

The finding that malleability peaked much earlier than behavioral performance in the mental rotation task is consistent with the intuitive speculation that malleability is a driving force for the development of behavioral performance. To test the link between malleability and performance in the mental rotation task, we selected the participants whose performance (i.e., the performance of the first play after practice) was among the top or bottom 20% of each age group in the training sample. Figure 4c shows the learning curves of the high and low performers across all age groups. Consistent with our speculation, results revealed that the high

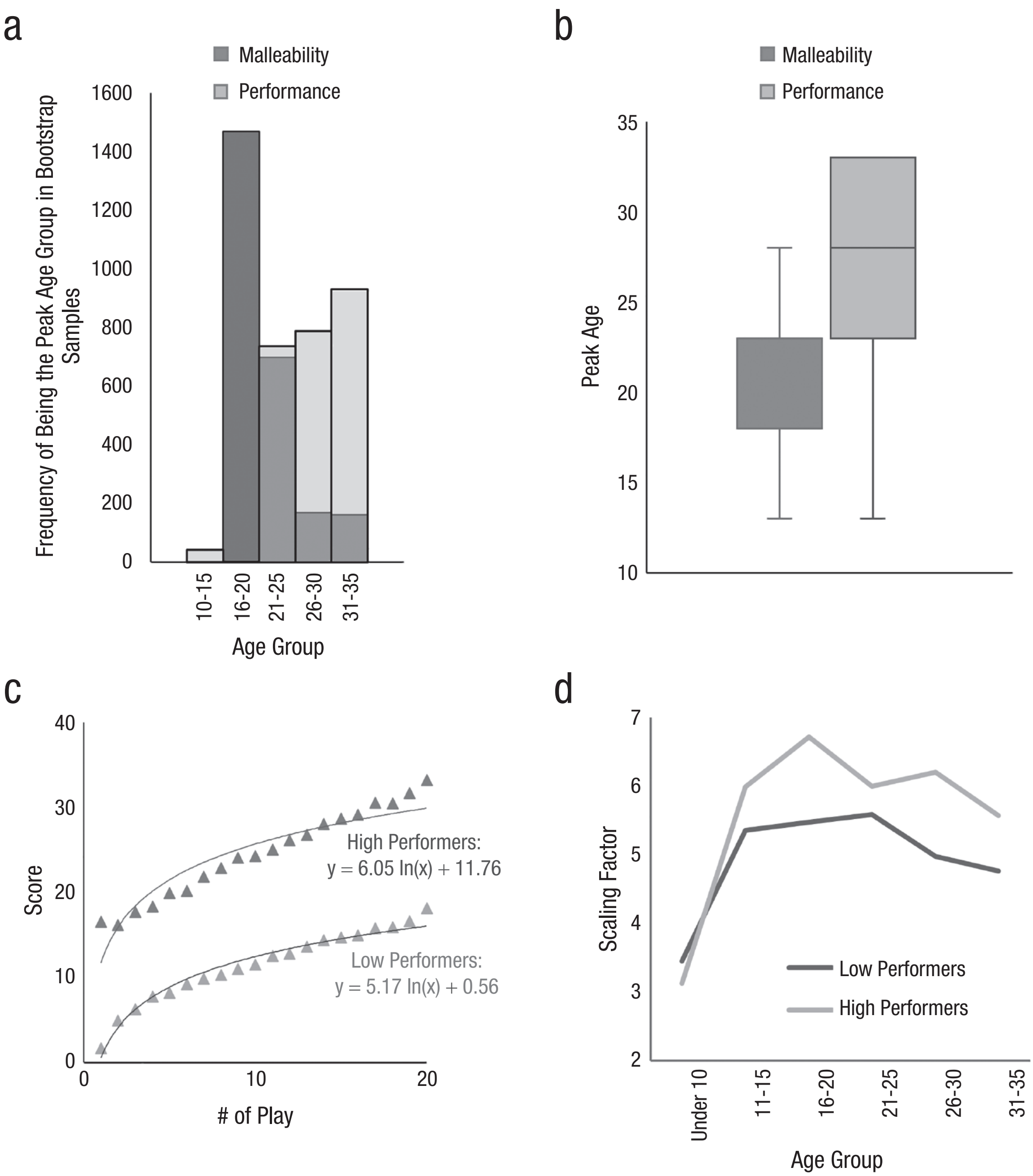

Fig. 4. Relation between performance and malleability. (a) Peak age distributions across bootstrapped samples for malleability and behavioral performance in the training sample. (b) Bootstrapped estimates of the peak ages of malleability and behavioral performance. The central line of each box denotes the median, the bottom line denotes the first quartile, the top line denotes the third quartile, and the whiskers denote the minimum and the maximum values of the distribution. (c) Learning curves of the high and low performers across all age groups. (d) Malleability of the high and the low performers in each age group.

performers across all groups showed a larger scaling factor (b = 6.05, Fig. 4c) than the low performers (b = 5.17), suggesting that high performers had higher malleability. Figure 4d charts the malleability of the high and low performers of each age group. In all age groups except the under-10 age group, the high performers showed higher malleability than the low performers.

Difficulty in mirror-image discrimination hindered the malleability of mental rotation

Surprisingly, for the under-10 age group, the high and low performers showed no difference in malleability, and their malleability was even smaller than that of the

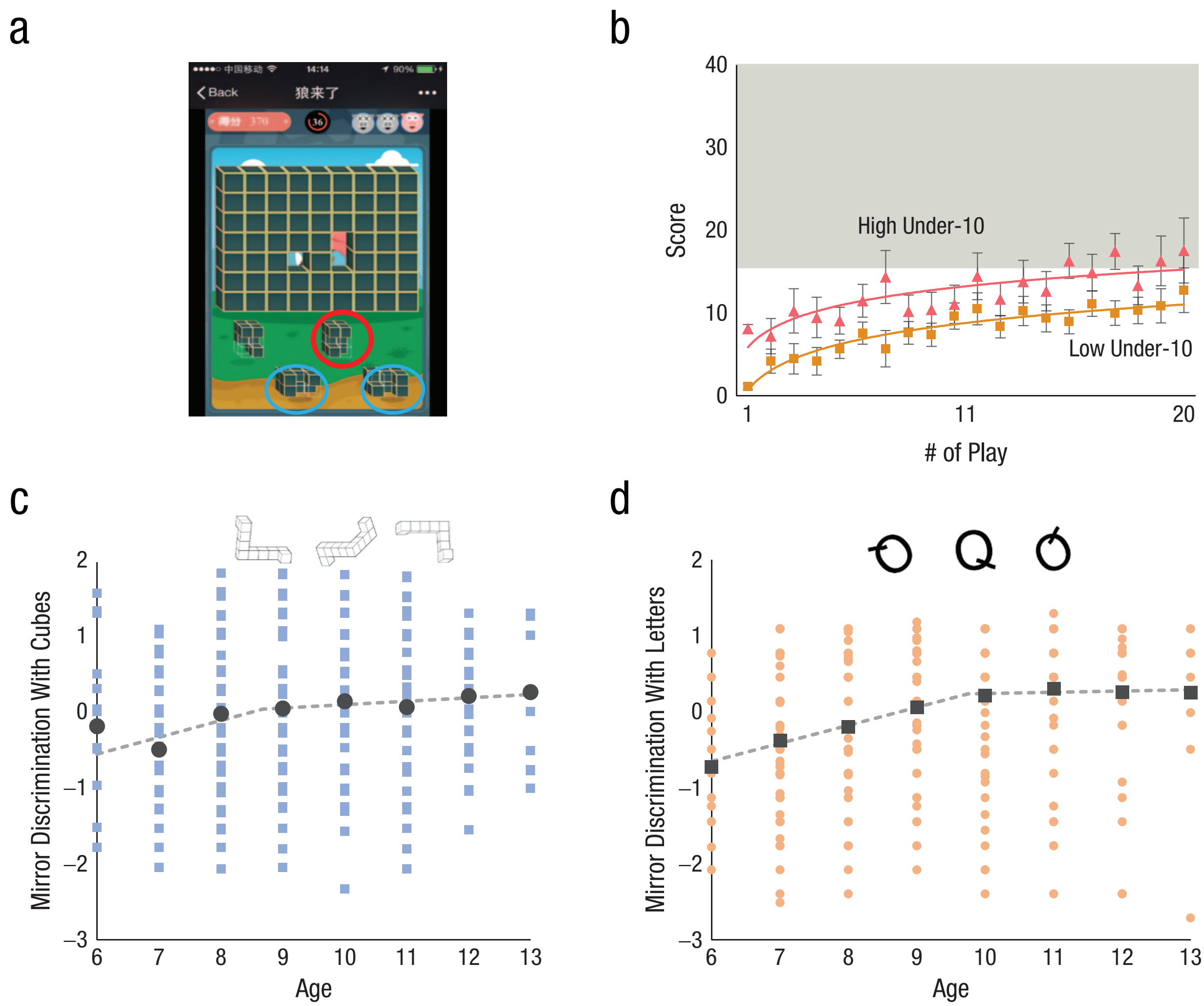

Fig. 5. Relation between mirror-image discrimination and malleability of mental rotation. (a) Screenshot illustrating a mirror-image discrimination trial in which the correct choice (red circle, front surface) was accompanied by two alternatives with the mirrored layout of the opening in one of the surfaces (blue circles, front surfaces). Such trials were scheduled as Trials 15 and 16. (b) Participants in the under-10 age group had difficulty passing Trial 15 and beyond, marked by the gray region. Error bars denote standard errors. (c) Developmental trajectory of mirror-image discrimination in the 3D mental rotation task. (d) Developmental trajectory of mirror-image discrimination in the 2D mental rotation task. Spline regression trendlines are depicted for each task by dashed lines. The turning points were estimated to be at 8.7 and 9.7 years old, respectively. Each dot represents a score from one participant.

low performers of other age groups (Fig. 4d). Careful examination of the learning curves of the under-10 age group suggests that both the high and low performers encountered unusual difficulty in passing the 15th trial, as the high performers needed more than 15 attempts on average to pass it, and the low performers were less likely to pass this trial even after more than 20 attempts (Fig. 5b). What is special about the 15th trial is that this trial and the 16th trial introduced the mirrored layout of the target (Fig. 5a). That is, in these trials, participants needed to discriminate the original from the mirrored layouts. Therefore, the ability of mirror-image discrimination is likely the prerequisite for malleability.

To test this conjecture, we tested a separate sample of children ($n = 331$, ages 6–13 years old) with pen-and-paper mental rotation tasks, including a 3D cube task (see Fig. 5c for example stimuli) and a 2D letter task (Fig. 5d). Half of the questions required mirror-image discrimination (i.e., the mirror condition), whereas the other half did not (i.e., the nonmirror condition). We first calculated accuracy for the questions with and without mirrored distractors in each task and observed a trend of performance increasing with age between 6 to 13 years old across tasks, especially for the items with mirrored distractors (Fig. S4 in the Supplemental Material). Figure 5c shows the scores of 3D mirror-image discrimination, the standardized residuals of the accuracy for the mirror condition regressing out that for the nonmirror condition, which increased with age from 6 to 13 years old. Spline regression analysis revealed a turning point at the age of 8.7 years old, suggesting that the ability of 3D mirror-image discrimination improved rapidly between 6 and around 9 years old, followed by gradual growth after 9 years old. In parallel, results from the 2D mirror-image discrimination task showed a similar development trend, with a turning point at the age of 9.7 years old (Fig. 5d). Together, these results confirmed the speculation drawn from the game-based assessment that mirror-image discrimination ability is not yet fully developed before 10 years old, which likely hindered the malleability of mental rotation.

Discussion

Taking advantage of online data collection and game-based assessment, we characterized comprehensive developmental trajectories of spatial cognition and its malleability in the present study. We found an asynchrony between the development of spatial cognitive performance and that of its malleability, as the learning rate peaked about 10 years earlier than performance on mental rotation. On the other hand, malleability also depended on the maturation of spatial ability, as children under the age of 10 showed particularly low malleability, which likely resulted from their underdeveloped ability of mirror-image discrimination. In short, by combining both coarse but massive online data and fine-tuned experimental data, our study illustrates the interaction of spatial cognition and its malleability during development, which also sheds light on translational implications for STEM-related educational practice.

The developmental trajectory of mental rotation performance observed here is consistent with previous reports on development in adulthood of spatial abilities based on laboratory studies (Wilson et al., 1975). Also, the widespread malleability of mental rotation across the life span is in line with the reported malleability of spatial abilities in various developmental stages (Bavelier et al., 2012; Feng et al., 2007; Uttal, Miller, et al., 2013). The convergence with the findings of traditional laboratory studies (Germine et al., 2012; Hartshorne & Germine, 2015) endorses the reliability and validity of the online game-based assessment approach in the present study. Importantly, our study extended previous studies in two aspects. First, our results filled the gap between the known developmental trends of spatial abilities in childhood (Frick et al., 2013; Hawes et al., 2015; Marmor, 1975) and adulthood (Wilson et al., 1975) by providing information on the transition between childhood and adulthood and by depicting a continuous development trajectory across the life span with the same task. Second, the present study comprehensively examined the malleability of spatial ability from childhood to adulthood, identifying late adolescence (16–20 years old) as a critical period when individuals are most sensitive to training on mental rotation in particular and possibly spatial cognition in general.

A particularly intriguing finding of the present study is the asynchrony between the development of performance and its malleability. The asynchronous developmental paces of malleability and performance have been documented in other cognitive domains, such as language acquisition (Snow & Hoefnagel-Höhle, 1978), implying a general principle for the development of cognitive functions. For instance, the training effect from daily practice in spatial cognition may accumulate through development and thus lays the foundation for performance change. Supporting this idea, our results showed that in adolescent and adult age groups, individuals with higher performance also showed higher malleability.

On the other hand, both the exploratory online test and the confirmatory experiment suggest that mirror-image discrimination might be a gatekeeping ability that constrained the malleability of young children. This result is consistent with previous findings that children at the age of 7 to 8 or even older cannot reliably perform above chance in mental rotation tasks involving mirror-image discrimination (Aaron & Malatesha, 1974; Hawes et al., 2015). The particular difficulty of mirror-image discrimination, which persists across the life span (Bornstein et al., 1978; Gibson et al., 1962; Gregory & McCloskey, 2010), suggests that it may involve unique cognitive mechanisms not implicated in the discrimination of rotated stimuli. Indeed, a previous study shows that children at the age of 6 show a steady decrease in error rate in discriminating 3D objects of different orientations, but the rate of mirror-image confusion stays high at least until 7 years old (Hawes et al., 2015). Various cognitive mechanisms may contribute to the particular difficulty of mirror-image discrimination of children. For instance, Paschke et al. (2012) suggest that mirrored images fit less to the top-down expectation; Hamm et al. (2004) propose that mirror-image discrimination requires an additional “flip” of the stimuli out of the picture plane, which requires more complex transformation of mental representation; McCloskey et al. (2006) suggest that mirror-image discrimination requires additionally representing polarity correspondence between the object and the extrinsic axes, which likely poses challenges to young children (see also Gregory & McCloskey, 2010). Our study revealed that the ages between 8 and 9 were critical for developing the ability of mirror-image discrimination. Therefore, future studies can test these three hypotheses with children at this age by examining whether mirror-image discrimination coexists with the contributing cognitive functions proposed by these hypotheses. Indeed, the present study suggests that mirror-image discrimination might hinder mental rotation performance and its malleability in a wider range of the life span. Visual inspection of Figure 4c indicates that the high performers across age groups passed the mirror-image trials without much difficulty and performance improves continually, resulting in a more linear-shaped learning curve compared with the low performers. In contrast, the low performers needed a substantial amount of training to succeed in mirror-image trials; therefore, the learning curve was flattened as an asymptotic line at the point of the mirror-image trials.

The finding that high malleability of spatial cognition occurred in late adolescence is of potentially great translational implication because of the transferability of mental rotation training effects into other spatial skills (Meneghetti et al., 2016; Terlecki & Newcombe, 2005; Uttal, Meadow, et al., 2013) as well as into STEM-related performance (Stransky et al., 2010). That is, training on spatial abilities may be most effective for adolescents and young adults, whereas elementary school children may not benefit as much from it. However, the training should be merit-based because there were significant individual differences in malleability as revealed in this study. Finally, the training may not be fully effective until the gatekeeping ability is mastered.

In addition, the present study replicated gender difference in mental rotation in the literature with a large sample. Although the male advantage in mental rotation has been well established in the literature, the source of this advantage has been a subject of debate, with arguments emphasizing either biological or experience/cultural factors (Miller & Halpern, 2014). The late emergence of gender difference observed in our study (i.e., its presence in adolescence and early adulthood but not in childhood) implies the influence of experience-related and/or environmental factors, consistent with findings of cross-national research (e.g., Levine et al., 2005) and studies on gender stereotype threat (e.g., Schmader, 2010). For instance, males might be encouraged more to participate in spatial-related tasks since childhood than females, and sufficient practice leads to performance advantage in adolescence, which might encourage further participation in spatial-related tasks and a persistent male advantage in the following developmental stages. However, it is also possible that the gradual unfolding of male advantage could be the result of late-onset hormonal influence or sex-dependent brain polymorphism, such as those taking place during and after puberty, though evidence of such influence is mixed in the literature (e.g., Herlitz et al., 2013). Further, the experience-related and/or environmental factors may interact with biological factors (Eagly & Wood, 2013), as gender-dependent educational experience and cultural expectation might facilitate the expression of phenotypes prescribed by biological factors.

In sum, by combining both a coarse online test and a fine-tuned experiment, we depicted comprehensive developmental trajectories of spatial cognition and its malleability, which invites research on the development of other cognitive domains. However, our results should be interpreted with caution given several limitations with online studies. First, the participants were self-selected, and the sample composition was certainly biased by that of the user population of the host app. Also, the present study did not verify the accuracy of age reports made by the participants because we were unable to devise an effective safeguard/check for large-scale online game-based data collection without making participants uncomfortable about their privacy. Therefore, some age reports might be inaccurate. However, this was unlikely to have distorted the pattern of the reported developmental trajectory because, given the large sample size of the present study, the noise caused by potential inaccurate age reports would have affected our estimation of all the age groups unselectively. Admittedly, the potential presence of inaccurate age reports may have led to an underestimation of developmental difference between age groups, and follow-up small-scale laboratory-based experiments are necessary to verify findings from the online experiment. Second, in the game-based assessment, participants reported their age information by selecting one of the age bins with the resolution of 5 years old (except the under-10 age group). Though this is a common practice in online game applications, it admittedly prevented us from depicting the developmental trajectory with finer resolution. This is particularly regrettable for the under-10 age group and the 11–15 age group, in which rapid cognitive development in spatial cognition takes place. To some extent, the pen-and-paper test provided the missing piece by allowing us to analyze mental rotation ability between the ages of 6 and 13 to depict the change in mirror-image discrimination ability during this period. More dedicated investigations are desirable to depict the developmental trajectory at this age range with finer resolution. Third, the online game-based assessment allowed us no control over the time span between participations or between-participation activities. This might have led to additional noise in estimating the malleability of each age group. However, the variation in time span was unlikely to affect the relation between behavioral performance and its malleability of each age group differently. Fourth, online studies provide only cross-sectional data. Therefore, follow-up longitudinal studies on the critical developmental periods identified by online studies are also necessary to rule out confounding factors such as the cohort effect. Fifth, in the present study, the malleability of spatial cognition was measured by the learning rate among multiple plays on mental rotation; therefore, malleability at the behavioral level may not fully capture plasticity in the neural substrates of mental rotation that may additionally reflect contributions from many factors other than training, such as lesion and sociocultural context. Future neuroimaging studies on the development of neural plasticity will likely provide an interesting comparison with the malleability solely from behavioral training. Finally, the findings of an asynchrony between performance and malleability and that of the coincidence between mirror-image discrimination failure and low malleability of spatial cognition did not present causal links, and future work is needed to examine the causality among these factors.

Transparency

Action Editor: Vladimir Sloutsky
Editor: Patricia J. Bauer
Author Contributions
Shan Xu: Formal analysis; Investigation; Writing – original draft; Writing – review & editing.
Yiying Song: Conceptualization; Formal analysis; Writing – original draft; Writing – review & editing.
Jia Liu: Conceptualization; Investigation; Writing – review & editing.
Declaration of Conflicting Interests
The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.
Funding
This study was funded by the National Natural Science Foundation of China (31872786, 31600925, and 31861143039), Tsinghua University Guoqiang Institute (2020GQG1016), Tsinghua University Qiyuan Laboratory, and Beijing Academy of Artificial Intelligence (BAAI).

Acknowledgments

We thank Miao Xu, Xiangzhen Kong, and Jialin Huang for their help and advice in task preparation and data analysis.

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/09567976221137313

References

Aaron, P., & Malatesha, R. (1974). Discrimination of mirror image stimuli in children. *Neuropsychologia*, *12*(4), 549–551.

Bavelier, D., Green, C. S., Pouget, A., & Schrater, P. (2012). Brain plasticity through the life span: Learning to learn and action video games. *Annual Review of Neuroscience*, *35*, 391–416.

Bornstein, M. H., Gross, C. G., & Wolf, J. Z. (1978). Perceptual similarity of mirror images in infancy. *Cognition*, *6*(2), 89–116.

Casey, M. B., Nuttall, R., Pezaris, E., & Benbow, C. P. (1995). The influence of spatial ability on gender differences in mathematics college entrance test scores across diverse samples. *Developmental Psychology*, *31*(4), 697–705.

Christodoulou, J., Johnson, S. P., Moore, D. M., & Moore, D. S. (2016). Seeing double: 5-month-olds' mental rotation of dynamic, 3D block stimuli presented on dual monitors. *Infant Behavior and Development*, *45*, 64–70.

Eagly, A. H., & Wood, W. (2013). The nature-nurture debates: 25 years of challenges in understanding the psychology of gender. *Perspectives on Psychological Science*, *8*(3), 340–357.

Feng, J., Spence, I., & Pratt, J. (2007). Playing an action video game reduces gender differences in spatial cognition. *Psychological Science*, *18*(10), 850–855.

Fortenbaugh, F. C., DeGutis, J., Germine, L., Wilmer, J. B., Grosso, M., Russo, K., & Esterman, M. (2015). Sustained attention across the life span in a sample of 10,000 dissociating ability and strategy. *Psychological Science*, *26*(9), 1497–1510.

Frick, A., Hansen, M. A., & Newcombe, N. S. (2013). Development of mental rotation in 3- to 5-year-old children. *Cognitive Development*, *28*(4), 386–399.

Germine, L., Nakayama, K., Duchaine, B. C., Chabris, C. F., Chatterjee, G., & Wilmer, J. B. (2012). Is the Web as good as the lab? Comparable performance from Web and lab in cognitive/perceptual experiments. *Psychonomic Bulletin & Review*, *19*(5), 847–857.

Gibson, E. P., Gibson, J. J., Pick, A. D., & Osser, H. (1962). A developmental study of the discrimination of letter-like forms. *Journal of Comparative and Physiological Psychology*, *55*(6), 897–906.

Gregory, E., & McCloskey, M. (2010). Mirror-image confusions: Implications for representation and processing of object orientation. *Cognition*, *116*(1), 110–129.

Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbacher, M. A. (2007). The science of sex differences in science and mathematics. *Psychological Science in the Public Interest*, *8*(1), 1–51.

Hamm, J. P., Johnson, B. W., & Corballis, M. C. (2004). One good turn deserves another: An event-related brain potential study of rotated mirror–normal letter discriminations. *Neuropsychologia*, *42*(6), 810–820. https://doi.org/10.1016/j.neuropsychologia.2003.11.009

Hartshorne, J. K., & Germine, L. T. (2015). When does cognitive functioning peak? The asynchronous rise and fall of different cognitive abilities across the life span. *Psychological Science*, *26*(4), 433–443.

Hawes, Z., LeFevre, J.-A., Xu, C., & Bruce, C. D. (2015). Mental rotation with tangible three-dimensional objects: A new measure sensitive to developmental differences in 4- to 8-year-old children. *Mind, Brain, and Education*, *9*(1), 10–18.

Herlitz, A., Reuterskiöld, L., Lovén, J., Thilers, P. P., & Rehnman, J. (2013). Cognitive sex differences are not magnified as a function of age, sex hormones, or puberty development during early adolescence. *Developmental Neuropsychology*, *38*(3), 167–179.

Hespos, S. J., & Rochat, P. (1997). Dynamic mental representation in infancy. *Cognition*, *64*(2), 153–188.

Levine, S. C., Vasilyeva, M., Lourenco, S. F., Newcombe, N. S., & Huttenlocher, J. (2005). Socioeconomic status modifies the sex difference in spatial skill. *Psychological Science*, *16*(11), 841–845.

Marmor, G. S. (1975). Development of kinetic images: When does the child first represent movement in mental images? *Cognitive Psychology*, *7*(4), 548–559.

McCloskey, M., Valtonen, J., & Sherman, J. C. (2006). Representing orientation: A coordinate-system hypothesis and evidence from developmental deficits. *Cognitive Neuropsychology*, *23*(5), 680–713. https://doi.org/10.1080/02643290500538356

McGraw, K. O., Tew, M. D., & Williams, J. E. (2000). The integrity of Web-delivered experiments: Can you trust the data? *Psychological Science*, *11*(6), 502–506.

Meneghetti, C., Borella, E., & Pazzaglia, F. (2016). Mental rotation training: Transfer and maintenance effects on spatial abilities. *Psychological Research*, *80*(1), 113–127.

Miller, D. I., & Halpern, D. F. (2014). The new science of cognitive sex differences. *Trends in Cognitive Sciences*, *18*(1), 37–45.

Moreau, D., Clerc, J., Mansy-Dannay, A., & Guerrien, A. (2012). Enhancing spatial ability through sport practice. *Journal of Individual Differences*, *33*(2), 83–88.

Newcombe, N. S., & Frick, A. (2010). Early education for spatial intelligence: Why, what, and how. *Mind, Brain, and Education*, *4*(3), 102–111.

Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., Bar-Anan, Y., Bergh, R., Cai, H., Gonsalkorale, K., Kesebir, S., Maliszewski, N., Neto, F., Olli, E., Park, J., Schnabel, K., Shiomura, K., Tulbure, B. T., Wiers, R. W., . . . Greenwald, A. G. (2009). National differences in gender–science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences, USA*, *106*(26), 10593–10597.

Paschke, K., Jordan, K., Wüstenberg, T., Baudewig, J., & Müller, J. L. (2012). Mirrored or identical—Is the role of visual perception underestimated in the mental rotation process of 3D-objects? A combined fMRI-eye tracking-study. *Neuropsychologia*, *50*(8), 1844–1851.

Schmader, T. (2010). Stereotype threat deconstructed. *Current Directions in Psychological Science*, *19*(1), 14–18.

Snow, C. E., & Hoefnagel-Höhle, M. (1978). The critical period for language acquisition: Evidence from second language learning. *Child Development*, *49*(4), 1114–1128. https://doi.org/10.2307/1128751

Spiers, H., Manley, E., Silva, R., Dalton, R. C., Wiener, J. M., Hölscher, C., Bohbot, V., & Hornberger, M. (2016, November 16). *Spatial navigation ability assessed in over 1 million people globally* [Paper presentation]. Society for Neuroscience 2016 Meeting, San Diego, CA, United States.

Stransky, D., Wilcox, L. M., & Dubrowski, A. (2010). Mental rotation: Cross-task training and generalization. *Journal of Experimental Psychology: Applied*, *16*(4), 349–360.

Terlecki, M. S., & Newcombe, N. S. (2005). How important is the digital divide? The relation of computer and videogame usage to gender differences in mental rotation ability. *Sex Roles*, *53*(5–6), 433–441.

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013). The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin*, *139*(2), 352–402.

Uttal, D. H., Miller, D. I., & Newcombe, N. S. (2013). Exploring and enhancing spatial thinking: Links to achievement in science, technology, engineering, and mathematics? *Current Directions in Psychological Science*, *22*(5), 367–373.

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, *47*(2), 599–604.

Vander Heyden, K. M., Huizinga, M., Kan, K.-J., & Jolles, J. (2016). A developmental perspective on spatial reasoning: Dissociating object transformation from viewer transformation ability. *Cognitive Development*, *38*, 63–74.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Jarrod Millman, K., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., . . . SciPy 1.0 Contributors (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272.

Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, *101*(4), 817–835.

Wilson, J. R., De Fries, J., Mc Clearn, G., Vandenberg, S., Johnson, R., & Rashad, M. (1975). Cognitive abilities: Use of family data as a control to assess sex and age differences in two ethnic groups. *The International Journal of Aging and Human Development*, *6*(3), 261–276.