Mental rotation in three-dimensional jigsaw puzzles and its association with assembling blocks: A pilot study

Yoshioka, Tsuyoshi

Department of Human Health Sciences, Graduate School of Medicine, Kyoto University

A Thesis
Submitted: 13 January 2023 / Resubmitted: 16 February 2023 / Accepted: 24 March 2023

Abstract

The jigsaw puzzle works as one rotates a puzzle piece to fit into its counterpart piece. Thus, mental rotation is expected to occur in solving the puzzle, three-dimensional in the present study. On this point, the process might differ from that of a traditional task that compares, or matches, a target stimulus with a reference stimulus. The rationale is because of how the processes would be internally represented in mind. Experiment 1 was performed through these notions. Besides, Experiment 2 was planned as correlational research to apply the results into vocational rehabilitation. The application is intended to detect an association between the performance of mental rotation and that of building blocks. Although similar studies were performed, those methodologies needed more details. Given this point, a unique data-curating technique is proposed in the present study to combine mean reaction time and error rate by assigning ranks based on priority. The technique was applied to predict those capable of assembling blocks. In addition, whether the strength of the association differs was also checked. The present study was conducted as a pilot study to explore its potential, fine-tune the design of experiments, and estimate sample size through power analysis. Experiment 1 suggested that mental rotation is utilized to fit a puzzle piece into its counterpart and that the task deviates from that of matching objects. Experiment 2 indicated that the capability of solving the puzzle among individuals strongly associates with that of assembling blocks. Besides, it implied the potential of utilizing the priority-based rank-ordering technique to predict assembling capabilities among individuals through computer-based behavioral data. The results of the present study were contrasted with that of similar research to permeably explain behaviors through a proposed unique model.

Introduction

Mental rotation and jigsaw puzzle. Mental rotation is a process in which one mentally rotates an object to have the same orientation with a reference stimulus (Shepard & Metzler, 1971). Its unique feature becomes apparent in a graph, though they vary across studies to some extent, where reaction time linearly increases until the angle of 180° per the angular disparity of two stimuli for their mental matching (Shepard & Metzler, 1971). Since its first discovery, considerable studies have been performed with various stimuli to confirm its versatility, such as hands (Cooper & Shepard, 1975; Parsons, 1987; Sekiyama, 1982).

The stimulus of the hand is distinctive from that of an object. This is because it differs in how they are interpreted as strategies to work on the tasks: egocentric or allocentric (Tomasino & Gremese, 2016). The uniqueness of mentally rotating hands becomes evident in a graph, where trends of left and right hands were indicated to shift symmetrically away from the angle of 180° in reaction time (Parsons, 1987; Sekiyama, 1982). The apparent distinction became lit when studies indicated an association between physical limitation and perception in mental rotation (Cooper & Shepard, 1975; Parsons, 1987; Sekiyama, 1982). Besides, a close link between the perceptual ratings of physical awkwardness or difficulty and the reaction time across angles was implied (Parsons, 1987; Sekiyama, 1983). Thus, mentally rotating hands in adduction seems to be faster than their abduction because less physical constraint exists that one perceives it as such. In addition, the role of physical mobility in mentally rotating hands is further supported by a study, where the stimuli-specific performance was found to deteriorate with hands crossed and put on the back (Ionta et al., 2007). In addition, the performance with the stimuli of the dominant, or preferred, hand was reported to outperform that of its opposite laterality (Ionta et al., 2007; Parsons, 1987). Hence, what piles up through nature and nurture likely matters in mental rotation; so much so that, but research through the stimuli of a jigsaw puzzle piece appears scarce in three-dimension, or 3D, especially in neurological research (Tomasino & Gremese, 2016). In this regard, testing the stimulus could prove helpful to see whether mental rotation occurs.

The jigsaw puzzle is a puzzle game that requires to assemble tile pieces into one meaningful assembly. Each puzzle piece has male and female sides, and the game is mostly enjoyed on a flat plane. One would start mentally rotating a piece rather than physically rotating it to solve the puzzle. In this regard, (a) mental rotation is expected to occur in solving puzzles, 3D in the present study. Besides, (b) the mental process may differ from that of comparing traditional 3D objects. The mental procedure to solve puzzles saves time yet has the disadvantage of mental workloads. Indeed, a recent study through functional near-infrared spectroscopy, or fNIRS, indicated the involvement of the bilateral or unilateral dorsolateral prefrontal cortex in solving a two-dimensional, or 2D, jigsaw puzzle per the degree of difficulty (Mutlu et al., 2020). Thus, one would have the heaviest metal loads to solve a 3D puzzle at the angle of 180°.

Interconnection between motor and cognition. In response to the task difficulty, one remedy as a countermeasure could be achieved through gesture. The gesture is a physical movement often seen when talking. Not only in talks but also the gesture is related to mental rotation, by which the performance was reported to facilitate mentally rotating a 3D object (Chu & Kita, 2011). Moreover, the physical motor manipulation of a joystick congruent with the direction of mental rotation was found to facilitate its mental manipulation (Wexler et al., 1998; Wohlschläger & Wohlschläger, 1998), whereas that incongruent with the direction was found to impede the mental manipulation (Wexler et al., 1998; Wohlschläger & Wohlschläger, 1998). Thus, regardless of egocentric or allocentric, mental rotation is likely interconnected with motor processes to facilitate or delay the cognitive process. The interconnection should become more evident, particularly in hands that are used to manipulate objects. Because of this manual influence, it might be better to control its physical mobility when testing mental rotation.

Task-specific schemas. Besides the discovery of the gesture effect, the study also found the number of gestures to decrease over time (Chu & Kita, 2011). This likely comes from the practice effect, and the cause was proposedly due to the internalization of computing spatial transformations (Chu & Kita, 2011). The finding is reasonable because manipulating hands simply consumes physical efforts on the task that becomes easier in computation after the internalization.

On its methodology, the very study was tested by comparing 3D objects (Chu & Kita, 2011). In this regard, the internalization should also occur in a similar task and might differ in degrees; namely, (c) another task with a 3D puzzle piece might vary, with or without gesture, in how its computation goes mentally represented over trials. This comes from that they require different tasks: one for matching 3D objects and the other for fitting a 3D object into its 3D puzzle piece.

Additionally, (c) it might differ at the beginning of the internalization. This is because the computation could be interfered with by an already-established internal representation in mind, or schema, which was coined in psychology to conceptualize what mentally piles up that changes (Bartlett, 1932); the documented concept likely dates back to the *De Anima* by Aristotle as *grammateion* or *tabula rasa* (Agamben, 1999; Duschinsky, 2012). For example, mentally parking a car between two cars in a parking lot should differ from simply mentally dragging the vehicle into the space. It was expected that the initially-internalized representation of a parking action in a first-person view, inside the vehicle as a driver or outside as a giant operator, would avoid obstacles and make a longer mental walk dependent on distance. The cognitive behavior could be analogically based on image scanning, where reaction time was found to increase linearly per distance in scanning the image map (Kosslyn et al., 1978). In another point of view, it might be applied through the representation of a parking motion as a third-person view. Noted is that mental process apparently differs among them; mental manipulation through a driver is restricted on the capabilities of the driver and the car (e.g., speed); that through a giant is apparently operated with a toy car with less physical weight; that through image scanning is to operate the focus of attention; that through a third person view passively sees a simulated parking action. The processes above are reasonable to be operated uniquely based on the internally-established representation of an action or a motion, regardless of active or passive. In contrast, the process could be otherwise facilitated thanks to the clues as a landmark, two vehicles with space in-between.

Either way, it was expected that (c) the initial representation of solving the puzzle becomes updated over trials to a similar representation, or spatial processing, as in matching 3D objects. This is because both tasks essentially share the same shape that has the exact spatial coordinates: a 3D object or the space of its 3D puzzle piece.

Hypotheses in Experiment 1. With these notions in account, Experiment 1 was conducted to clarify the following hypotheses: (a) whether mental rotation occurs in fitting a 3D object into its 3D puzzle piece; (b) whether the performance differs between the task of matching 3D objects and the task of fitting a 3D object into its 3D puzzle piece; and (c) whether the mutual difference dissipates over trials through internalization if there has a difference between the tasks.

A social issue and vocational rehabilitation. In addition to Experiment 1, the second experiment was performed as correlational research in consideration of applying the results from Experiment 1 into vocational rehabilitation, the purpose of which is to help a disabled person in employment and to promote their integration into society (International Labour Organisation & International Labour Conference, 1983). The necessity of applying this purpose becomes clear by grasping the state of affairs on a state scale, specifically Japan. A recent review on the history of mental health care in the country revealed that it has been slow in turning into the current to normalize inpatients with mental disorders into society (Kanata, 2016). To tackle this issue, the necessity of measures was proposed to downsize or close psychiatric hospitals that are highly represented by private hospitals and also to secure places where they want to live (Kanata, 2016). In this regard, healthcare professions, such as occupational therapists and social workers, can help them normalized into a society and provide them one alternative to enhance the quality of life by securing jobs. More specifically, occupational therapy can be utilized to help them acquire occupational skills.

Besides mental disorders, noted is that the inpatient proportion was comparable to the ratio of those with intellectual disabilities (Kanata, 2016). This area of research seems scarcely investigated, particularly in occupational therapy; a recent scoping review revealed its dim representations in interventional studies to support the youth with intellectual disabilities transition from school to work (Rosner et al., 2020). The very review also reported a need for more solid interventional evidence in this field (Rosner et al., 2020). Thus, there is room for occupational therapists or related practitioners to tap into this area.

Occupational therapy and education. Occupational therapy can be also utilized in supporting those with difficulties in education, which are with many interventional research. However, there seems to have less solid evidence of the effect of mental rotation in interventional studies, where recent research found mixed evidence of mental rotation training to improve arithmetic performance in primary school children (Cheng & Mix, 2014; Cheung et al., 2020; Hawes et al., 2015). It should be noted that those studies were based on different interventional periods and tasks. A far transfer into arithmetic performance was found through on-day training or one-week training with puzzle-like tasks with 2D objects (Cheng & Mix, 2014; Cheung et al., 2020). This is contrary to what was reported in the other study, where no such apparent transfer was reported through training with puzzle plus matching tasks with 2D objects three sessions a week over six-weeks (Hawes et al., 2015). It should be also noted that the effect of the on-day training might have been due to the priming effect (Cheng & Mix, 2014; Hawes et al., 2015). In these regards, results from the present study may provide some clues about task effects and become helpful for its application into education. The present study was designed mainly in consideration of vocational rehabilitation but also of educational rehabilitation for tapping future studies.

A cognitive primer to assembling. Among domains of vocational rehabilitation, one area resides in helping acquire new vocational skills, one of which is assembling. Assembling enables a designed whole object out of part components, and technological development has largely replaced the task with automation. Nevertheless, there is a necessity for human care where a small amount, custom-made products, or meticulous work is required. In this regard, those with disabilities could help fill the gap as one area to further facilitate their assimilation into society. However, commercial companies are to aim for profits. Because of this basic concept, employers could become reluctant to hire those with difficulties because, in one regard, it requires initial investment to train the employees to acquire new skills. Also, employees may leave the company early after the skill acquisition.

With these points considered, low cost-effectiveness could be partly overcome through the priming process, by which those with difficulties are primed to ignite reasonable assembling skills right after employment. In this way, both parties can benefit from the lowered hiring threshold. This priming process considers both the short term and the long term in its beneficial effects. With these accounts, one priming to assembling objects could be applied through cognitive training with mental rotation, the association of which in performance was tested in some studies (Brosnan, 1998; Zhang et al., 2020). These studies, however, used the data of overall accuracy through the aggregation of angles for analysis (Brosnan, 1998; Zhang et al., 2020). The procedure might have, in turn, obscured the unique characteristics of mental rotation. With this point taken into account, angles are separately treated in the present study. Besides, a unique priority-based rank-ordering technique, touched on in its method section, was proposed in the present study to treat behavioral data on separate angles and to check (d) whether the technique has its potential usage in future research.

The speed-accuracy compatibility on task conditions. The uniqueness of mentally rotating a 3D object to compare with two reference stimuli is seen as a mountain-like shape in both reaction time and error rate (Chu & Kita, 2011). Besides, larger standard errors tended to be graphically situated around the angle of 180° (Chu & Kita, 2011). This characteristic would be because the same angle exposed the most challenging condition with the heaviest mental loads, as implied in the fNIRS study (Mutlu et al., 2020), that add to random variation. The rationale that reaction time and error rate shift in tandem is reasonable, given the interconnection between motor and cognition in mental rotation.

Indeed, physically-manipulating actions were implied to make fewer errors with their temporally shorter walk closer to the target position and also on a broader area of the target position (Fitts, 1954). This implies that speed is compatible with accuracy per the degree of difficulty, which for clarity was termed the speed-accuracy compatibility; the concept differs from the speed-accuracy tradeoff, in that the tradeoff works as a seesaw whereas the compatibility walks in tandem. Hence, one is expected to make more errors in a longer walk during a difficult motor task if the one takes a stable and reasonable strategy; the behavior also applies, as indicated, to a difficult psychological task because of the closely-related interconnection between motor and cognition.

In consideration of this point, the shift of strategy in a group as a whole on separate task conditions through the within-participants design could be detected where this speed-accuracy compatibility is violated (e.g., conservative strategy with less errors and longer reaction time on an easy task condition; whereas, liberal strategy with more errors and shorter reaction time on a difficult task condition). In the example above, one is expected to make less errors in a shorter reaction time on an easy task and to make more errors in a longer reaction time on a difficult task if the one takes a stable and reasonable strategy, regardless of conservative or liberal.

The rationale is further supported from psychophysics, which was coined to connect outer physical stimulus with inner sensation (Fechner, 1860/1912); the Weber's law suggests a linearity between the perceived change in stimuli intensity and the initial stimulus intensity (Fechner, 1860/1912; Weber, 1834/1912). The law implies that the increase of perceived stimulus intensity weakens as the stimulus intensity increases, which is specified by how the perceived strength behaves on the logarithm of the stimuli strength (Fechner, 1860/1912). Hence, it becomes more difficult to differentiate two stimuli as the intensity of initial stimulus increases, which eventually comes with more errors in accordance. Hence, both laws in psychophysics are interchangeable, and they play a role to link physical stimulus in physics to sensation and perception in psychology. The speed-accuracy compatibility covers these ideas by utilizing the behavioral data of accuracy and reaction time at the end point of a flow.

The speed-accuracy compatibility on a task condition. The speed-accuracy compatibility above was on a group level, applicable as a whole on separate task conditions in accordance with the degree of difficulty. On this point, the notion could be expanded into an individual level as parts of a whole as wholes on a single task condition. However, the degree of difficulty appears unattainable because of a single task. In this regard, the focus could be shifted from a group level to an individual level so the degree of difficulty is described from behavioral data on an individual level. In line with the speed-accuracy compatibility, one is expected to make less errors in a shorter reaction time; in other words, one who has the highest rank in accuracy is expected to also have the highest rank in reaction time. Furthermore, variation in performance through continuous data is expected to become large at a different task in line with the speed-accuracy compatibility as mentioned on a group level, which makes differentiating individual capabilities in performance easier. In consideration of these points, it is expected that (e) those capable of coordinated actions through motor yet cognitive functions would make better performance at a difficult task.

However, a cautious approach is required because the assumption bases on that the same stable and reasonable strategy was utilized throughout task conditions on an individual level. This becomes apparent when considering that group data has variation in accuracy whereas individual data has not. As a result, the degree of differentiation likely differs between the two; each strategy on task conditions as a group is not interpreted to be different as long as they are not far away from variation; however, the interpretation differs on an individual level because accuracy has not variation on this level. Thus, it is more expected to falsely detect a violation of the speed-accuracy compatibility (i.e., Type I error). In consideration of this point, the speed-accuracy compatibility is likely applicable on a group level but inapplicable on an individual level.

Nevertheless, the disadvantage in a way is an advantage in another way, and accuracy index with no variation could be utilized to order capability among individuals based on the needs of a society (i.e., accuracy over speed within a specified time frame). This notion is incorporated in detecting those capable, which is to be touched on in the method section at Experiment 2 as the priority-based rank-ordering technique.

Hypotheses in Experiment 2. In consideration of these notions, it was expected that (e) those capable of mental rotation at angles around 180° would be also capable of assembling blocks. Additionally, (f) the association strength of performance ranks might differ, if at least one association was observed, between the tasks: between the task of assembling objects and the task of matching 3D objects; between the task of assembling objects and the task of fitting a 3D object into its 3D puzzle piece. Experiment 2 was conducted to check these three hypotheses.

The purposes of the present study. The present research was conducted as a pilot study to check the hypotheses, (g) explore its potential, fine-tune the design of experiments, and estimate sample size through power analysis. The research questions from (a) to (g) are to be handled throughout the paper. Part of the tentative results was reported at a conference (Yoshioka & Liang, 2022). The present paper covers and completes it.

Experiment 1

The experiment aimed to clarify (1) whether mental rotation occurs when fitting a 3D object into its 3D puzzle piece. If the occurrence was confirmed, it aimed to clarify (2) whether a difference exists in performance between matching 3D objects and fitting a 3D object into its 3D puzzle piece. Besides, it aimed to clarify (3) whether performance differs in these two tasks over trials and (4) whether the subjective score of internalizing reference stimuli differs in these two tasks. Where no significance was observed in performance, the analysis aimed to see (5) how much sample size is required for a future study in Experiment 1.

Method

Participant. 10 healthy young participants were recruited from Kyoto University and its affiliated institution on a convenience basis: Female = 3, Male = 7; Left-handed = 2, Right-handed = 8; Mean age = 25.3 years, SD age = 4.0 years. Handedness was sifted through the FLANDERS handedness questionnaire (Nicholls et al., 2013; Okubo et al., 2014). All participants had normal to corrected-to-normal vision.

Matching task (MT). This task required to find either a reference 3D object of the left or the right that matches with a target 3D object (See Figure 1). Two reference 3D objects and one target 3D object have protrusions facing away from the onlooker, as in Figure 1, which differs from those stimuli facing towards the onlooker (Chu & Kita, 2011).

Fitting task (FT). This task required to find either reference 3D object of the left or the right that fits into a target 3D puzzle piece (See Figure 1). The same two reference 3D objects in MT were used.

Stimuli. The stimuli were created through Blender (Community, 2021). Each object was created by piling unit cubes, in which an aggregated 3D object had nine of unit cubes, and its 3D puzzle piece had 116 of unit cubes. The 3D puzzle piece was created by subtracting the 3D object from a whole cube that has 125 of unit cubes. The space in the cube corresponds to the 3D object.

Each stimulus was created by rotating a target 3D object or a target 3D puzzle piece along a bisector axis at the geometric center point of the whole cube. The rotation had six angles: 0°, 60°, 120°, 180°, 240°, and 300° (the Angle factor). This process was applied for two tasks and two literalities: MT and FT (the Task factor); Left and Right, respectively. In total, 24 stimuli were prepared.

Apparatus. The experiment was performed on a 15.6-inch portable display through Linux Ubuntu version 20.04.2.0 LTS with Matlab R2021a, along with Psychtoolbox-3 (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997). A foot pedal with two steps was used to respond either left or right.

Design. The experiment consisted of six sessions containing MT and FT in each session, along with a practice period in the beginning. Each session had 24 trials of MT and FT randomly assigned. Participants were to finish six sessions of 144 trials, along with 10 trials for practice. Recess was prepared after each session. At the end of the experiment, participants completed a questionnaire regarding their strategy to solve the tasks throughout sessions.

Questionnaire. The paper-and-pencil questionnaire was administered at the end of the experiment. For simplicity, every two sessions from the first were collapsed into three stages: Early, Middle, and Late (the Stage factor). Among these stages, four check-boxes were prepared. Participants were to choose one of four check-boxes depending on each stage and each task. The first two choices asked whether the task was worked on by looking at the three objects and by comparing them through rotation from above to below or from below to above. The last two choices asked whether the task was worked on by looking solely at the target object and by comparing it with the reference objects in the mental image through rotation from above to below or below to above. For FT, the language of target objects was rephrased with the space of the 3D puzzle piece.

Procedure. Participants were situated in a chair before the display monitor on a table, both of which were pre-defined in position, with their hands covered below their bellies. The participants were instructed to respond as fast and accurately as possible by pressing the left or the right of reference 3D objects through the foot pedal that matches the target stimulus. After the practice period, participants started the first session at their discretion by pressing either step of the foot pedal. Reaction time and accuracy were recorded. Recess was prepared after each session. Participants were to press the foot pedal to continue the session at their discretion. Each trial had a maximum time frame of 30 seconds to respond; the subsequent trial was set to begin after the time frame. In total, the whole procedure took roughly one hour.

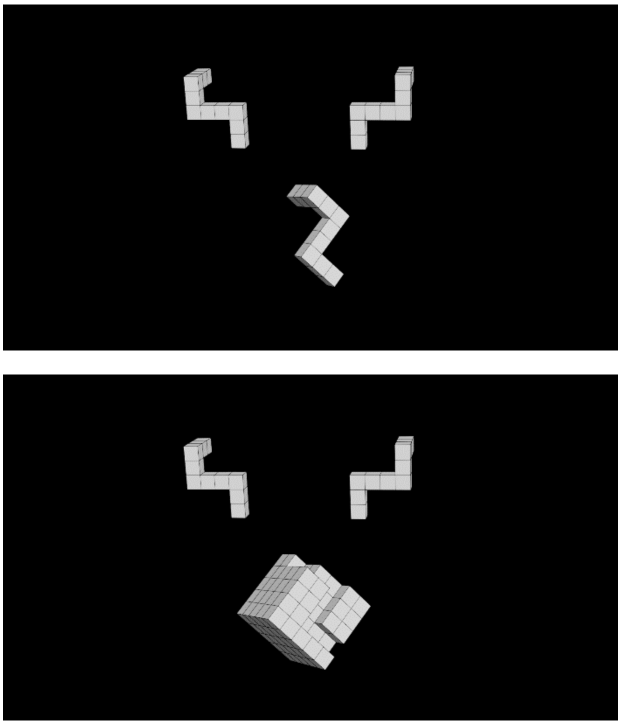

Figure 1. MT and FT: above and below, respectively. In each stimulus, the objects on the top are reference 3D objects. The object on the bottom is a 3D object or a 3D puzzle piece as a target stimulus: above and below. Note. MT (Matching task). FT (Fitting task).

Data curation. For curating the questionnaire, the choices were collapsed into two types for detecting strategical shifts: whether or not to go from looking at the three blocks to the target stimuli alone, the affirmation of which was treated as the initialization of the reference 3D objects.

As for the behavioral data, two data sets were prepared: one with the Task factor and the Angle factor; the other with the Task factor, the Angle factor, and the Stage factor. The first data set was used for analyzing a behavioral trend across angles in FT and task differences in outcome across angles, in addition to analyzing correlation in Experiment 2. The second data set was used for analyzing task differences in outcome across angles over stages. In the same way as the questionnaire, stages classified this data set as the Stage factor (i.e., Early, Middle, and Late). Both data sets contained two response variables: mean reaction time and error rate (mean RT and ER, respectively).

As for outliers, RT measurements on each participant were first sifted through the criteria of 200 msec, which is about the pinnacle of mean simple RT (Woods et al., 2015). Then, the data were further sifted through the median absolute deviation method with a constant b = 1.4826 and a threshold value of

2.5 (Leys et al., 2013). Outliers were detected in 10% of the first data set and in 9% of the second data set, and these outliers were removed from further analysis. The sifted data were calculated for mean RT and ER. Thereafter, mean RT measurements among participants were sifted by the modified boxplot method of three times the interquartile range, through the rstatix package (Kassambara, 2022a). Extreme outliers were detected one in the first data set and two in the second data set; these participants were omitted from the later analysis: 10% and 20%, respectively. The same two participants were also omitted from the questionnaire.

Data analysis. RTs and ERs were first checked for normality through the Shapiro-Wilk test with the Bonferroni-Holm correction before the analysis of variance, or ANOVA. The sum of squares of Type III and Type I were respectively utilized for the parametric test and the non-parametric test. Where the normality assumption was not rejected, the parametric repeated measures ANOVA, or RM-ANOVA, of one-way, two-way, and three-way were performed. The Greenhouse-Geisser correction was applied where the sphericity assumption was rejected through the Mauchly test. Where appropriate, either the paired Student t-test or the RM-ANOVA as additional analyses were performed post hoc with the Bonferroni-Holm correction. Where the normality assumption was rejected or inappropriate for the test, nonparametric analysis was performed. Namely, the Friedman test was applied for one-factor analysis, after which the paired-samples sign test with the Bonferroni-Holm correction was applied where appropriate. For the analysis of multiple factors, the ANOVA through aligned rank transform with repeated measures, or RM-ART-ANOVA, was used (Wobbrock et al., 2011). Where applicable, either the post hoc pairwise contrast test or the RM-ART-ANOVA as additional analyses were applied with the Bonferroni-Holm correction.

For the analysis of the questionnaire, the Cochran-Mantel-Haenszel test was to be processed for a three-factor analysis of repeats (McDonald, 2014). The data was first to be processed through the Breslow-Day test (Breslow & Day, 1980). Where the assumption of the homogeneity of odds ratios was not rejected, the Cochran-Mantel-Haenszel test was to be applied. Where no significance was observed in the very test, the Pearson chi-square test or the Fisher exact test was to be applied at each level of the Stage factor.

Effect sizes were calculated where appropriate for their interpretations (Cohen, 1988; Kay, 2021). All analyses, including statistics, part of outlier detection, and the calculation of sample size, which are extended to Experiment 2, along with the creation of graphs, were processed through the R software (R Core Team, 2022). Graphs were created through the ggpubr package (Kassambara, 2022b).

The Shapiro-Wilk test, the parametric ANOVA, its post hoc pairwise comparisons, the Friedman test, and its post hoc sign test were processed through the rstatix package (Kassambara, 2022a). The nonparametric RM-ART-ANOVA was processed through the ARTool package (Kay et al., 2021; Wobbrock et al., 2011). Its pairwise comparison was to be processed through the ART-C, which utilizes contrast tests within the ART paradigm and is applicable for either a between-participants or a within-participants designs (Elkin et al., 2021). The Breslow-Day test was to be processed through the DescTools package (Signorell, 2022).

The power analysis was processed through the Superpower package (Caldwell & Lakens, 2022; Lakens & Caldwell, 2021). In the very analysis, the average of correlation coefficients through the Fisher z-transformation was utilized. Besides, the alpha level with the Bonferroni correction was applied where appropriate.

Results

RT and ER among angles in FT. The Shapiro-Wilk test with the Bonferroni-Holm correction showed no rejection of the normality assumption on all RTs [$p > .05$]. The one-way RM-ANOVA showed a significant main effect of the Angle factor [$F(1.67,13.38) = 11.75$, $p = .002$, $\eta_p^2 = .595$] (See the top graph of Figure 2). Its post hoc pairwise comparison of the paired Student t-test with the Bonferroni-Holm correction showed significant differences in mean RT among angles between 0° and 60°, 0° and 180°, along with 0° and 300° [$t(8) = −5.01$, $p = .015$, $d = −1.67$; $t(8) = −5.19$, $p = .013$, $d = −1.73$; $t(8) = −4.47$, $p = .027$, $d = −1.49$, respectively].

As for ER, the Shapiro-Wilk test was not applied because the same values of zero were observed across participants. The Friedman test showed a significant inequality of rank sums across angles [$X^2(5) = 21.44$, $p < .001$, W = .48] (See the bottom graph of Figure 2). However, its post hoc paired-samples sign test with the Bonferroni-Holm correction showed no significant inequality of signs among angles, the smallest p-value between 0° and 180° and between 180° and 300° [p = .240].

RT and ER of tasks among angles. The Shapiro-Wilk test with the Bonferroni-Holm correction showed no rejection of the normality assumption on all RTs [$p > .05$]. The two-way RM-ANOVA showed a significant interaction effect between the Task factor and the Angle factor [$F(5,40) = 2.50$, $p = .046$, $\eta_p^2 = .238$] (See the top graph of Figure 3). However, its post hoc one-way RM-ANOVA at each level of the Angle factor with the Bonferroni-Holm correction showed no significant simple main effect of the Task factor at all six angles, the smallest p-value at 300° [$F(1,8) = 6.81$, $p = .186$, $\eta_p^2 = .460$].

As for ER, the Shapiro-Wilk test was not applied because the same values of zero were observed across participants. The two-way RM-ART-ANOVA showed no significant interaction effect between the Task factor and the Angle factor nor a significant main effect of the Task factor [$F(5,88) = .47$, $p = .796$, $\eta_p^2 = .026$; $F(1,88) = .89$, $p = .348$, $\eta_p^2 = .010$, respectively] (See the bottom graph of Figure 3).

RT and ER of tasks among angles over stages. The Shapiro-Wilk test with the Bonferroni-Holm correction showed no rejection of the normality assumption on all RTs [$p > .05$]. The three-way RM-ANOVA showed no significant interaction effect among the Task factor, the Stage factor, and the Angle factor [$F(10,70) = 1.24$, $p = .279$, $\eta_p^2 = .151$] (See the top graph of Figure 4).

As for ER, the Shapiro-Wilk test was not applied because the same values of zero were observed across participants. The three-way RM-ART-ANOVA showed an interaction effect among the Task factor, the Stage factor, and the Angle factor [F

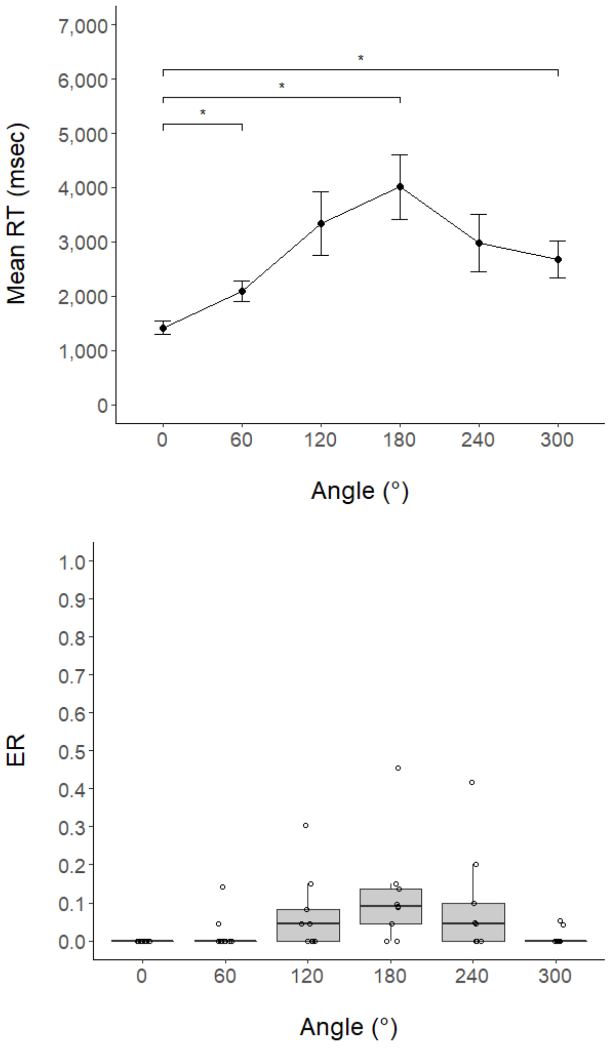

Figure 2. RT vs. Angle and ER vs. Angle in FT: above and below, respectively. Note. FT (Fitting task). * $p < .05$

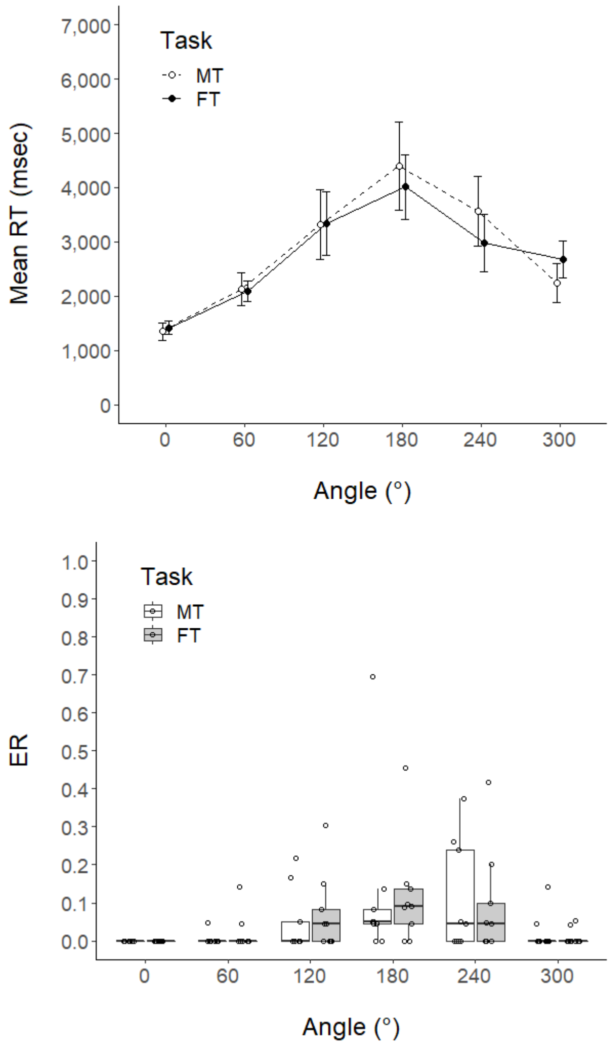

Figure 3. RT vs. Angle and ER vs. Angle, grouped by the Task factor: above and below, respectively. Note. MT (Matching task). FT (Fitting task).

(10,245) = 2.59, p = .005, η_p^2 = .096] (See the bottom graph of Figure 4). However, its post hoc two-way RM-ART-ANOVA at each level of the Angle factor with the Bonferroni-Holm correction showed no significant interaction effect between the Task factor and the Stage factor, the smallest p-value at 180° [F(2,35) = 3.48, p = .210, η_p^2 = .166].

Questionnaire. The statistical analysis was not processed through the Breslow-Day test due to the limitation of data counts, which should be large in each stratum and each of them should have more than five in expected cell counts to be used (Signorell, 2022).

The count data suggested a tendency of the internal representation of reference 3D objects over stages but no apparent discrepancy in proportions between MT and FT [.13 and 0 in the Early stage; .38 and .25 in the Middle stage; .63 and .50 in the Late stage, respectively].

Power analysis. Though a significant main effect in the one-way RM-ANOVA and a significant interaction effect in the two-way RM-ANOVA in mean RT were observed, no significant interaction effect was observed in the three-way RM-ANOVA. In this regard, the power analysis of the interaction effect ind-

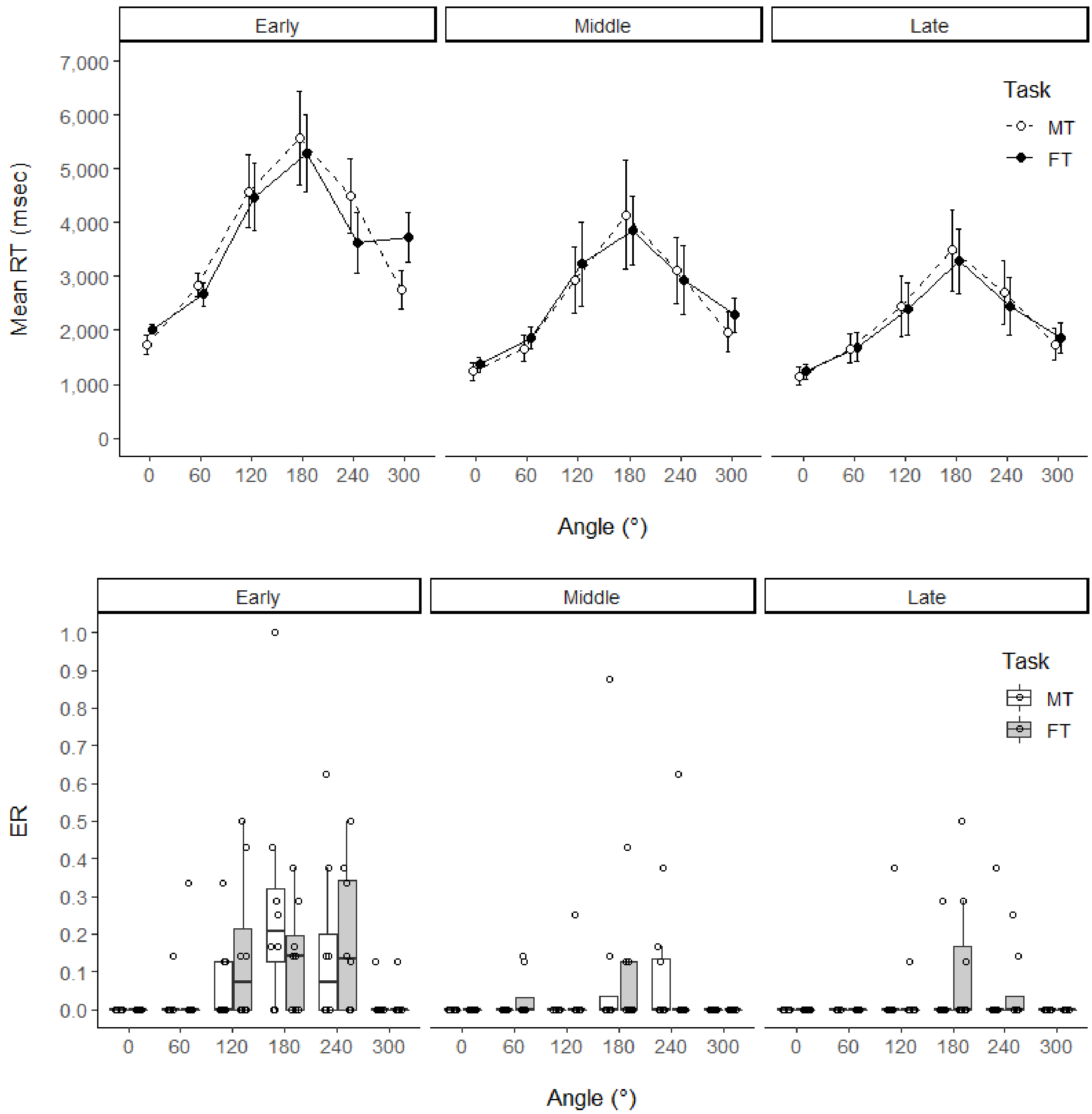

Figure 4. RT vs. Angle and ER vs. Angle, grouped by the Task factor, and separated by the Stage factor: above and below, respectively. Note. MT (Matching task). FT (Fitting task).

icated the powers of .79 and .81 with the sample sizes of 24 and 25, respectively. In addition, the power analysis of its post hoc interaction between the Task factor and the Stage factor at each level of the Angle factor with the Bonferroni correction indicated the powers of .77 and .80 with the sample sizes of 21 and 22, respectively, the smallest size at the angle of 300°.

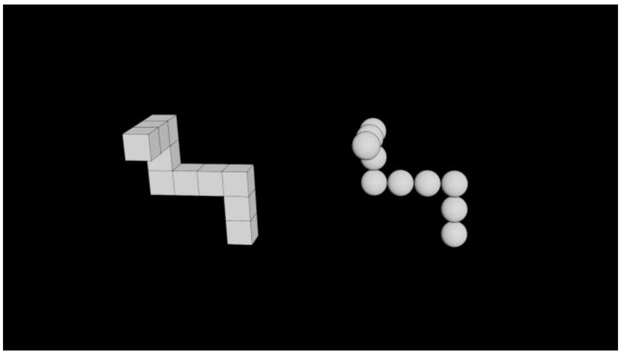

Figure 5. An example of two objects with their units unique in shape is displayed. A 3D object of cubes and a 3D object of spheres: left and right, respectively. The shapes of the two objects are adjusted to align with the present study.

Discussion

Experiment 1 partly aimed to clarify (1) whether mental rotation occurs in FT; and if so, it aimed to clarify (2) whether a difference exists in performance between MT and FT. Besides, it aimed to clarify (3) whether performance differs in these two tasks over stages.

Mental rotation in FT. The first one-factor analysis indicated a similarity of solving the puzzle to that of comparing traditional 3D objects. This is because the distribution shapes mountain-like with the highest mean RT and ER at the angle of 180°. In consideration of this point, the result could be interpreted as the occurrence of mental rotation when fitting a 3D object into its 3D puzzle piece.

Behavioral difference between FT and MT. The second two-factor analysis indicated the effect of the Task factor that was dependent on each level of the Angle factor in mean RT. However, there observed no specific angle level of the effect. In the meantime, one effect size of the latter post hoc analysis is interpreted as large (Cohen, 1988). Thus, there seems to underlie a difference between fitting a 3D object into its 3D puzzle piece and matching 3D objects.

Priority-based p-value correction. The post hoc analysis after the second two-way ANOVA had no significance. In this regard, it might be due to that the smallest p-values were conservatively corrected with the Bonferroni-Holm correction as an exploratory analysis. The procedure inflates Type II errors as the number of comparisons increases. Thus, the correction procedure that is priority-based and pre-planned during the design of experiments could be desirable to test the data in focus for future studies.

Mental transfer. A possible explanation in the difference between FT and MT could be explored by investigating the mental processes step by step. In MT, it could be solved by mentally rotating a reference object, mentally operating the object to move from the original place to a target place, and mentally matching the spatial coordinates of the objects. In FT, the steps could go to mentally rotate a reference object, mentally operate the object to move from the original place to a target place, and to mentally fit it into the spatial coordinates of the space of the counterpart; noted is that the space of the counterpart in FT shares the target object in MT. Both processes include the same mental rotation in the first step, the same mental manipulation in the second step, and a conceptually similar process in the third step. On the second step, there seems to be a shared process in mental manipulation; the process of mentally operating an object to move from one place to another place without rotation is termed mental transfer for simplicity. Hence, mental rotation is a process to mentally rotate an object without mental transfer. By overlapping each step on both tasks, there seems to have no obvious difference on a step level between the two processes (i.e., mental rotation, mental transfer, and matching or fitting processes on the same spatial coordinates). In consideration of this point, there likely has other explanations.

Stereotypes. Previous similar research on college students focused on gender and stimuli factors in mentally rotating a 3D object of cubes and a 3D object of spheres (Rahe & Quaiser-Pohl, 2020). Each unit shape in both wholes shares the same geometric center point but differs in volume (See Figure 5 for an example). The results of the study are to be mainly discussed in the latter discussions.

The very study found no interaction between the stimuli factor and the gender factor but a significant main effect of the stimuli factor with the cubes better in performance than the spheres (Rahe & Quaiser-Pohl, 2020); the effect sizes are respectively interpreted as small and medium (Cohen, 1988). It also found a significant main effect of the gender factor with the male over the female in performance (Rahe & Quaiser-Pohl, 2020), the result of which is aligned with meta-analyses (Linn & Petersen, 1985; Maeda & Yoon, 2013). On the significant main effect of the stimuli factor, however, the cause of the discrepancy was not discussed in detail in the study. Instead, the focus was on a stimuli effect on each gender; a significant simple main effect of the stimuli factor was solely observed in the male (Rahe & Quaiser-Pohl, 2020). Noted is that a 3D object of cubes and that of spheres were proposedly each perceived as male-stereotyped and female-stereotyped (Ruthsatz et al., 2014), which might come from construction toys and necklaces (Rahe & Quaiser-Pohl, 2020). On the gender discrepancy of the simple main effects, the effects of gender-based stereotypes were proposed for its explanation (Rahe & Quaiser-Pohl, 2020).

In consideration of these points, the gender factor or gender-based stimuli factor may play a role in how one interprets the stimuli, which consequently affects the cognitive process. On this point, the present study was performed as a within-participants design, the results of which becomes almost unexplainable from the gender factor.

In regard to a stimuli effect on each gender in the similar research, it should be noted that the interpretation came from an exploratory post hoc analysis after no significant interaction. Thus, it could be rather beneficial to investigate a significant main effect of the stimuli factor. In this regard, there might be another factor to explain this stimuli effect on the previous research: notably, the interconnection between motor and cognition.

The how-schema. <u>Familiarity and performance.</u> Athletes train, and the influence of their immersed motor manipulation can be seen when they repeatedly operate the same action in practice for its replication in a public attempt. In the learning process, the action becomes more coordinated at a later time than what was performed the first time, apparently through the consolidation of the representation of an action or a motion. After replicating the action in a public attempt, the athletes can describe how the action was performed not only from the egocentric state of the athlete but also from the allocentric state of not the athlete (e.g., the mood of spectators); the schema of the state of something, not restricted to an object, is termed the how-schema for simplicity. Hence, the how-schema is expected to have representations of active or static states (e.g., an action, a motion, or atmosphere). In the example above, the athlete practiced an action to consolidate its how-schema; the athlete replicated the action in public based on the how-schema; then, a how-schema of self and a how-schema of not-self were consolidated.

One similar analogy could be applied to handling a 3D object of spheres, in that females perform better thanks to more likely exposure to a necklace of spheres. This familiarity with the necklace to lessen the negative stereotyped effect in mental rotation and to enhance its performance was the rationale why one study on the same two factors was conducted (Rahe et al., 2021); the result was solely seen as a significant main effect of the gender factor (Rahe et al., 2021). Thus, familiarity with stimuli might not always lead to improving performance in mental rotation. In this regard, familiarity through exposure to the female-stereotyped stimuli of a necklace might work the other way around to deteriorate the performance.

The rationale comes from the observations through experience where commercially available necklaces of spherical beads as wholes were vulnerable to physical force (e.g., pulling a necklace of spheres tied with a string). In the very example, the spheres are connected with a stem penetrating the insides, which has to be with plasticity so it can round a neck. Where no stem attended, it requires gluing; the process becomes demanding to connect parts because two spheres are to be glued at one point due to their curvatures, which differs from two cubes glued at squares. Thus, the former would be easily separated or deformed by force. In response, its physical behavior becomes internalized as a how-schema so that the mentally-simulated physical transfer of the spheres becomes delayed, the cognitive influence of which likely emerges as a slow response or a lower score within a specified time frame.

<u>Observational learning.</u> Nevertheless, it fails to explain the following outcome from the previous research: a significant main effect of the stimuli factor regardless of the gender factor (Rahe & Quaiser-Pohl, 2020). In this regard, observation may play a role as touched previously. Specifically, the male may observe behaviors in a third-person view to consolidate a how-schema rather than in a first-person view: namely, through observational learning (Bandura, 2004).

Observational learning enables to acquire knowledge by observing others' behaviors (Bandura, 2004). It proceeds in four steps: "the Attentional Processes, the Retention Processes, the Production Processes, and the Motivational Processes" (Bandura, 2004, p.483). Noted is that the fourth step has the performance of observationally-learned behavior to be enhanced through rewards (Bandura, 2004). Thus, it seems observation itself is not enough to learn a behavior but through the assistance of enhancement. On this point, there should be another kind to be enhanced through observational learning rather than the external behavior itself, as described in the previous step; the third step has concepts to be transformed into a behavior (Bandura, 2004). Hence, it seems an external behavior traces back to its concept; both are virtually a synonym. In consideration of these points, observational learning could be interpreted as consolidating a concept, or a schema, through observation so that the internal representation can be more accurately replicated and externally represented as an explicit or implicit behavior.

With this notion taken into account, males may observe how necklaces of spherical beads are handled by females and then learn how the necklaces are operated according to their physical behaviors, which results in consolidating its how-schema. In this way, the same behavioral shift from the female-stereotyped stimulus on both genders could be explained. This interpretation is independent of the gender factor and permeable to any gender types. Noted is that exposure on this scale needs more than once to consolidate the how-schema. In other words, what piles up is likely to be more consolidated through repeated exposure; the strength of the representation is reasonable to have a continuous range rather than an all-or-none representation (e.g., learning to ride a bicycle for the first time). On this point, the degree could be dependent on how one is exposed to the stimuli to consolidate its fluctuating representation (e.g., frequency or magnitude). Hence, repeated exposure in a society through observational interactions, regardless of explicitly or implicitly, may contribute to facilitate consolidating a how-schema.

Summary. In consideration of these accounts, the present experimental tasks might have similarly been worked on through a how-schema that had been consolidated through repeated exposure and remained established. Namely, the behavioral difference in the present experiment could be explained by that the puzzle task required a unique mental representation of an action or a motion that differs from the representation of matching 3D objects. A further study may shed light into what exactly makes the difference if it has (e.g., whether obstacles play a role in the process).

The Gestalt principles and the what-schema. The principle of proximity and the what-schema. Contrary to the how-schema, it could be otherwise looked in another point of view onto the result, the worse performance with the spheres over the cubes on the similar research (Rahe & Quaiser-Pohl, 2020): namely, perception. Perception lies in transit to schemas, and what is perceived becomes consolidated into a schema. Thus, it likely plays a role because what is perceived at first may differ from what is perceived at last. One explanation could be applied through the principle of proximity in the Gestalt principles of grouping (Koffka, 1935; Wertheimer, 1923/1938).

According to this principle, ones closer to each other are perceived as a group (Koffka, 1935). By taking this principle into account, a 3D object of cubes likely becomes more of a grouped unit of items than that of spheres. This is because the surface of hemicubes is closer each other than that of hemispheres; simply put, the cubes have less space between neighboring items. This is contrary to a 3D object of spheres that have more space between neighboring items. Thus, a 3D object of spheres might have been perceived as a pseudo-whole, which contrasts with a whole of cubes, rather than a whole of spheres.

In consideration of this point, the mental process might have varied between the two stimuli: mentally manipulating a pseudo-whole in the spheres; whereas mentally manipulating a whole in the cubes. The former mental process would be with more mental workloads. This comes from that a loosely-integrated whole likely has sharper representations of parts, so that these parts are processed as multiple units in parallel. This is contrary to the latter mental process in that a whole of parts is processed as one unit in parallel. This point may play a role in light of a behavioral study, where the number of items to be possessed in the short term as working memory was suggested to have its finite capacity of around seven as chunks (G. A. Miller, 1956; G. A. Miller et al., 1960); more recently, it was suggested to be around four chunks (Cowan, 2001); the worked memory is reasonable to decay to clear the table. Hence, integrating parts into a meaningful whole as a chunk seems to help the cognitive process in the working memory. On this point, the previous research utilized 10 items to each create a 3D object of cubes and a 3D object of spheres (Rahe & Quaiser-Pohl, 2020). Thus, processing a 3D object of spheres might have been burdened with more consuming mental workloads with each sphere roughly treated as each meaningful whole, which likely ended up in worse performance. In this regard, the interpretation could explain, in another point of view, the following result from the previous research: a significant main effect of the stimuli factor in mentally rotating a 3D object of cubes and a 3D object of spheres (Rahe & Quaiser-Pohl, 2020).

Nevertheless, studies on gender and the stereotyped stimuli seem to encounter mixed outcomes on the effect of the stimuli factor in college students and children (Rahe et al., 2021; Rahe & Quaiser-Pohl, 2019, 2020; Ruthsatz et al., 2014); besides, all the studies utilized angle-free scores (Rahe et al., 2021; Rahe & Quaiser-Pohl, 2019, 2020; Ruthsatz et al., 2014). Thus, there should be room for further investigation.

For future studies, it could see whether the behavior with the spheres shifts based on the number of parts. In addition, checking a behavioral shift over time would be also beneficial. This is because the Gestalt principles concern perception, by which its mental representation piles up through repeated exposure to consolidate the schema of an object, or what-schema for simplicity. It becomes beneficial to check whether repeated exposure consolidates the what-schema of a whole of spheres over time.

In terms with its relation to the present experiment, the same notion could be applied to the 3D puzzle because the experiment utilized a 3D puzzle with more cubes than a 3D object. However, both of the stimuli would be treated as a whole rather than a pseudo-whole, with the parts being all cubes. Thus, the principle of proximity might not be the case to explain the behavioral difference. In this regard, another principle could play a role; namely, the principle of figure-ground (Koffka, 1935; Pind, 2012; Rubin, 1915).

<u>The principle of figure-ground and the what-schema.</u> According to this principle, one is perceived upon another (Koffka, 1935). By taking this principle into account, the dark background stands out less frequently as the figure because the shape itself has not enough meaning and has less saliency. Thus, the focus would be on the very objects. On this point, a 3D object is likely perceived as the figure because it has not another apparent object due to the continuation of parts within a whole. On the contrary, it differs in a 3D puzzle piece, where it apparently has two representations within a whole: the very 3D puzzle piece and its space.

The relationship between the figure and the ground is interchangeable and shifts based on how they are perceived. In this way, the 3D puzzle piece and its space are perceived respectively as the figure and the ground, and vice versa. On this point, it would be reasonable for the 3D puzzle piece to be worked on first because it has a physical existence; and later, its space of meaning becomes gradually manageable through the consolidation of its what-schema. This space differs from a blank space, or the state of *tabula rasa*, because its meaning is expected to be created accompanied after the introduction of an object that has its meaning.

<u>The what-schema and the how-schema.</u> On the point of strategy to mentally solve the 3D puzzle, it is inefficient to take a detour to fit the 3D object into its 3D puzzle piece because of more consumption in mental efforts. In response, a countermeasure could be to perceive its space as one representational object and to mentally transfer the 3D object on a straight path to match with the spatially-represented object. It likely requires repeated exposure through perception to facilitate consolidating the what-schema over stages, as touched on in the aforementioned discussion on the principle of figure-ground. This gradual consolidation appears similar to what was discussed for the how-schema to be consolidated through observational leaning as one way. In consideration of these points, it is reasonable to see a shift in the what-schema as well as the how-schema in tandem because what to be worked on shifts over time.

The third three-factor analysis indicated no three-way interaction in mean RT. However, its effect size is interpreted as medium (Cohen, 1988). In addition, the nonparametric test found a significant three-way interaction in aligned-ranks of performance. Besides, it would be beneficial for the stereotyped mental representation of actions to dissipate over time through repeated exposure so that it reaches a reasonable strategy to optimize the behavior adapted to each environment. This is so it aligns with the task-specific rule to match the spatial coordinates of a 3D object with that of the corresponding space of its 3D puzzle piece. The implication is visually described in Figure 4, where the line graph on FT becomes both more linearly plus symmetrical over stages and compares with that of MT.

<u>Summary.</u> In consideration of these aforementioned accounts, the behavior in the present experiment could be explained by the shift of the what-schema and the how-schema from unique to similar representations through the Gestalt principles and observational learning as potential mediums.

The how-schema on gender difference. As described in the discussion, the similar research found a significant main effect of the gender factor with the male over the female in performance (Rahe & Quaiser-Pohl, 2020). The result is, as mentioned, in line with meta-analyses (Linn & Petersen, 1985; Maeda & Yoon, 2013). On this point, a modularly-different approach with the how-schema could be employed to explain the characteristic behavior rather than focusing on one cognitive module of the spatial ability itself: namely, physical strength.

As sometimes seen in sports, a game is held separately between the female and the male, the cause of which might be to ensure fairness to control the gender-based factor. Indeed, part of an anatomical study on biceps brachii indicated no significant gender difference in the average number of its fibers nor the number of its motor units (A. E. J. Miller et al., 1993); however, the male had a significantly wider cross-sectional area and a significantly greater muscle strength in elbow flexion but no significant discrepancy in their ratio (A. E. J. Miller et al., 1993). The study also reported a significantly-positive correlation between the cross-sectional area and the muscle strength (A. E. J. Miller et al., 1993). Thus, the results were primarily indicative of wider fibers that create more physical strength in the male (A. E. J. Miller et al., 1993). Noted is that the female was better in enduring repetitive elbow flexion of the muscle loads at 60% of the maximum isotonic strength (A. E. J. Miller et al., 1993). Hence, motor function seems modular as in cognitive function, and each gender appears to have their respective forte. In the case of the male, they seem physically strong to operate an object with upper limbs, which makes the object physically transferred faster when operated harder in line with laws in physics. However, this notion still needs more to reasonably explain the gender discrepancy of the psychological behavior in mental rotation from the meta-analyses.

In this regard, the how-schema could be used for the explanation. The physical behavior bases on a how-schema that is utilized for its actual motor behavior; the mentally-simulated behavior would also occur through a retrieval from the same or a similar how-schema. The rationale comes from the likely interconnection between motor and cognition, as touched on in the Introduction. Thus, a gender discrepancy in physical behavior characterized by physical strength is to be reflected in a cognitive behavior. In this way, it could explain why the male consistently outperforms the female in mental rotation. Noted is that the interpretation is not specific to the gender factor but the factor of physical characteristics. Thus, it might be beneficial to test for future studies the factor of physical characteristics as in the followings: a between-participants design (e.g., athletes or bodybuilders and those-not among females).

Experiment 2

This experiment aimed to describe (1) whether a priority-based rank-ordering technique has the potential of its usage. Besides, it aimed to clarify (2) whether the performance ranks of mental rotation tasks from Experiment 1 are correlated with that of assembling blocks. In addition, if at least one significant association was confirmed, it aimed to clarify (3) whether a difference in the strength of association exists between MT and FT towards assembling blocks. Where no significance was observed, the analysis sought to clarify (4) how much sample size is required for a future study in Experiment 2 and across the two experiments.

Method

Sample size. Design. The whole sample size in the present study was estimated in consideration of Experiment 2. The Pearson product-moment correlation coefficient of $r = .75$, which is the midpoint of the large effect size range (Cohen, 1988), was adopted. Besides the correlation coefficient, alpha level = .05, approximate power = .80, and two-tails were used to estimate the sample size of 10; the calculation was processed through the pwr2ppl package (Aberson, 2019, 2022).

Accuracy index. Similar research was performed for preschool or primary school children to see an association between building blocks and mental rotation (Brosnan, 1998; Zhang et al., 2020). As noted in the Introduction, the very studies used the data of aggregated angle-free accuracy for behavioral analysis (Brosnan, 1998; Zhang et al., 2020). One possible reason might be to ensure focused concentration through small trials. However, the sensitivity of detecting an association might have been obscured at the cost of losing the unique feature of mental rotation, which would become detrimental to differentiating individual capabilities.

Among the very studies, one study found no significant association between a score based on accuracy in mental rotation and the complexity score of building blocks (Zhang et al., 2020). About its methodology, however, it should be noted that 70 unit blocks in unique shapes were utilized at their discretion to construct a purpose-designed house. (Zhang et al., 2020). This differs from a traditional 3D object of multiple unit cubes, more complex in shape, to be mentally rotated dynamically. Besides, 2D house stimuli with nine of black or white windows were utilized in the mental rotation task (Zhang et al., 2020). This also differs from that of mental rotation in 3D. Furthermore, the very study did not describe a specified time frame for the mental rotation task (Zhang et al., 2020). This might have missed individual differences in capability due to the speed-accuracy tradeoff. On these points, however, closely matching factors between building blocks and mental rotation might have been unrealizable because the tasks would otherwise become much more difficult for preschool children.

Meanwhile, another study found a significant Pearson product-moment correlation coefficient of $r = .38$ between the error rate in mental rotation and the test pages reached in constructing blocks (Brosnan, 1998). This effect size is interpreted as medium (Cohen, 1988). However, it should be noted that a score based on accuracy, away from reaction time, was used solely for primary analysis within a specified time frame (Brosnan, 1998).

The priority-based rank-ordering technique. This accuracy index has a disadvantage dependent on task difficulties: one might encounter the ceiling effect or non-normality data. It should be reasonable that among those who have the maximum score, or broadly the same score, those who complete the task in a shorter time are more capable. For example, those students who are not best at math get scores below the maximum out of 100, where each of them is primarily identifiable by score, whereas those capable in math score 100. The capability is evident in that the latter group is more skilled in math than the former group. Under this notion, identifiable tags can be assigned to the former group but not to the latter group due to the same equal values.

This procedure encounters a drawback in both traditional parametric and nonparametric correlations because one might either proceed into analysis, discard the data of the latter group, or conduct a separate analysis for the two groups. The drawback should become more apparent in plenty of equal values. On this point, ordering ranks further could be applied to the latter group by assigning ranks based on how fast students completed the math test. In this way, all students can be incorporated into the same analysis.

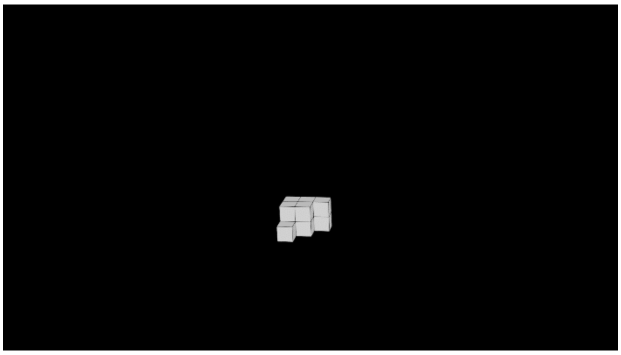

Figure 6. AT (Assembling task). A target object is displayed.

The flow from ER as a primary priority then to RT as a secondary priority is because the accuracy scale goes as discrete with two ceilings, whereas that of RT goes as continuous with one ceiling. The ceiling of the latter is virtually impossible to be reached. This is because it takes around 200 msec on average at the pinnacle to simply respond to a stimulus (Woods et al., 2015). Besides, accuracy likely matters more at the end point of a flow in many fields than speed in society (e.g., assembling goods in the manufacturing industry, entrance exams in education, and marathons in sports). In these regards, a technique of ordering ranks based on priority should be helpful to differentiate and predict those capable among a group of individuals.

Summary. With these notations considered, the strength of association would be stronger through the Spearman rank-order correlation if angles were separately treated and both mean RT plus ER were incorporated into a data set of ranks.

Participants. The same participants of Experiment 1 joined the experiment.

Stimuli. Target objects were created through Blender (Community, 2021).

Assembling task (AT). This task required to assemble with blocks a target object that appears on the display monitor (See Figure 6).

Apparatus. The stimuli appeared on the same equipment used at Experiment 1. The same foot pedal with two steps was used to record the completion time. 3D plastic block parts, flat or zigzag in 3D shape, with embedded magnets were used to construct a target object.

Design. The experiment included three sessions, each having a three-minute time frame, along with a practice period in the beginning. Each session had one trial. The practice period utilized two block components, and the task was worked on in a pre-defined order.

Procedure. Participants were situated in the same way as in Experiment 1. Participants were instructed to assemble a target object as fast and accurately as possible by using three block components. Participants were to press either step of the foot pedal after completing the assembly. RT and accuracy were recorded. The block components were arranged in a predefined alignment onto a wooden board that was covered with a towel. Participants were to uncover the towel to start the assembly after a target stimulus appeared on the display monitor.

Data curation. Mean RTs and ERs were consolidated into one data set as an ordinal scale. First, ERs were given ranks: smaller values with higher ranks. Second, if ERs had the same rank, mean RTs were used to rearrange the rank order among them: smaller values with higher ranks. The same procedure was applied for those of MT and FT from Experiment 1.

Data analysis. The Pearson product-moment correlation was applied to detect an association in performance ranks between AT and either MT or FT from Experiment 1, through the rstatix package (Kassambara, 2022a). Then, where at least one significant difference in the Spearman rank-order correlation coefficient from zero was observed in either correlation, a comparison of these coefficients at the very angle was applied through the Hotelling-Williams t-test (Weaver & Wuensch, 2013). The calculation was processed through the cocor package (Diedenhofen, 2022; Diedenhofen & Musch, 2015). The procedure is because the Fisher z-transformation of the Spearman rank-order correlation coefficients, especially in data under non-normality, was indicated to be tolerant (Myers & Sirois, 2006).

Power analysis was processed through the pwr2ppl package (Aberson, 2019, 2022).

Results

Correlation in each angle. The Pearson product-moment correlation showed a significant association of performance ranks between AT and FT solely at the angle of 180° with the Bonferroni-Holm correction [$r_s(7) = .883$, $p = .019$] (See Table 1). At this angle, however, the Hotelling-Williams t-test showed no significant difference in the correlation coefficients between FT and MT [$t(6) = 1.97$, $p = .097$, $q = .93$].

Table 1. The Spearman rank-order correlation coefficients of MT and FT towards AT in performance at each level of the Angle factor.

Angle	MT	FT
0°	−.300 (1)	−.100 (.798)
60°	−.300 (1)	.483 (1)
120°	.233 (1)	.383 (1)
180°	.433 (1)	.883* (.019)
240°	.317 (1)	.367 (1)
300°	−.450 (1)	.367 (1)

Note. MT (Matching task). FT (Fitting task). AT (Assembling task). * $p < .05$

Power analysis. Though one significant correlation was observed at the angle of 180°, no significant difference in the Spearman rank-order correlation coefficients was observed at the very angle in the Hotelling-Williams t-test. In this regard, the power analysis at the very angle indicated the powers of .79 and .82 with the sample sizes of 17 and 18, respectively.

Discussion

Experiment 2 aimed to describe (1) whether a priority-based rank-ordering technique has the potential of its usage. Besides, it aimed to clarify (2) whether the performance ranks of MT and FT from Experiment 1 are correlated with that of AT. In addition, if at least one significant association was confirmed, it sought to describe (3) whether a difference in the strength of association exists between MT and FT towards AT. Where no significance was observed, the analysis sought to clarify (4) how much sample size is required for a future study in Experiment 2 and across the two experiments.

Sample size. The correlation analysis indicated a significant association in performance ranks solely between FT and AT at the angle of 180°; the effect size is interpreted as large (Cohen, 1988). Thus, the sample size of nine still seemed manageable to detect a significant correlation. However, one sample was omitted from the analysis, which was short of the designed sample size of 10. On this point, it would be better to have an extra sample on top of the designed size; it also applies to Experiment 1.

From the two experiments, the desirable sample size for the main study based on the present design of experiments, also in consideration of potential outlier data, would be around 30.

The priority-based rank-ordering technique. The correlation outcome implies that the method of priority-based rank-ordering through the ER-then-RT order might effectively predict the capability of the performance among individuals. The technique would be unique supposedly in this area of research in mental rotation: many seem to prefer ER (Brosnan, 1998; Cheung et al., 2020; Hawes et al., 2015; Zhang et al., 2020); some utilize mean RT (Lyu et al., 2016; Sekiyama, 1983); few utilize a combined continuous data (Dahm et al., 2022).

Besides its usage in predicting an association, integrating mean RT and ER as one data set of ranks might be also useful in interventional research. The implication comes from a recent interventional study of puzzle video game, where a potential usage to adopt a combined continuous data of RT and ER through a formula was indicated (Safaei et al., 2022). In these regards, the priority-based rank-ordering technique could be utilized for practical purposes as one alternative or additional analysis, especially when one encounters the ceiling effect or non-normality. In doing so, the procedure of ordering ranks based on priority should be also stated beforehand.

Priority-based paired p-value correction. The analysis might indicate that the capability of assembling among individuals is predicted through the performance ranks of FT at the angle with the highest mental workloads. However, it requires a cautious approach because the same outcome should have also occurred between MT, which behaves similarly in mental rotation at the angle of 180°, and AT. Thus, further research is desirable to see what makes this discrepancy if there indeed underlies a difference.

In doing so, it would be better to use the priority-based correction of p-values, as discussed in the discussion of Experiment 1. This is due to the implication that the p-values in focus might have been conservatively over-compensated; so much so that, but an additional procedure would be necessary for multiple correlations. This is because the same priority should be applied to both correlations in interests to control Type II errors, rather than its correction by one-by-one with step weights. For example, if two correlations are involved, its correction by pairs with even weights would be better (i.e., six pairs with weights of two steps until 12 in calculating the coefficients of Experiment 2). In consideration of this point, the correction procedure of p-values that are paired, priority-based, and pre-planned during the design of experiments for multiple correlations should facilitate the interpretation of future studies.

Computer-based prediction of assembling skills. In terms of the application into society, it indicated its potential for vocational rehabilitation. This is because behavioral performance through a computer screen could be used for predicting the capability of assembling blocks among individuals. This advantage is to save cost and time with merits for both employers and employees, compared with dexterity tests that necessitate analog materials (Yancosek & Howell, 2009). Also, the benefit would be massive in that the task can be applied in a large group without its corresponding well-versed examiners.

Besides, its application as one psychological test could be expanded to optimize, before deciding which skill to hone, future careers of the potential workforce (e.g., students at vocational school). This notion is plastic and could be expanded into academic education, which is to be discussed in a later section.

A short-term cognitive primer. Another application could be to use the psychological test to warm up fundamental cognitive functions right before assembling designed wholes to facilitate the behavior.

The first part of warming up basic functions would be helpful to facilitate performance on general tasks. This might be slightly indicative from an interventional study, where mental rotation training with puzzle plus matching tasks with 2D objects in children had a marginally significant interaction effect as an intermediate transfer into one of similar tasks (i.e., 2D puzzle-like tasks) (Hawes et al., 2015); the effect size is interpreted as medium on the lower boundary (Cohen, 1988). Noted is that this could be interpreted as a near transfer because of a close similarity to one of the three training games.

For the second part of facilitating the assembly, a cognitive test with the specific stimuli that is utilized on field likely brings a stronger effect onto productivity, thanks to its close similarity. This becomes somewhat indicative in the same study, where the mental rotation training with 2D objects had a significant interaction effect as a near transfer effect into closely-related tasks (i.e., 2D mental rotation tasks with stimuli of animals or letters) (Hawes et al., 2015); the effect sizes are respectively interpreted as large and medium (Cohen, 1988). Thus, it may be more productive to set the stimuli to on-site conditions as a short-term cognitive primer during employment to prime the subsequent assembling process. In consideration of this point, differentiating a match and a mismatch on two consecutive tasks would be preferential.

Indeed, meta-analyses on the effect of mental practice before its cognitive or physical attempt excluded studies that had a mismatch between what was mentally practiced and what was performed (Driskell et al., 1994; Toth et al., 2020). On the effect of the mental practice, the studies reported a significantly beneficial effect of mental practice before its attempt onto the subsequent cognitive and physical performances (Driskell et al., 1994; Toth et al., 2020); the magnitude of mental practice effect seemed to be stronger onto the subsequent cognitive task than onto the subsequent physical task (Driskell et al., 1994). Additionally, the study indicated that the positive effect of the mental practice gradually weakened as an interval to its immediately-subsequent congruent attempt widened (Driskell et al., 1994). In consideration of these points, it is reasonable that the mental rehearsal of solving puzzles could become a short-term positive primer to the subsequent assembling if the mental practice was performed right before the subsequent assembling.

A long-term cognitive primer. The short-term cognitive primer is not applicable before employment; thus, cognitive training with general stimuli as a cognitive primer for the long-term benefit is preferred. The result of the present experiment is indicative of an association but no causal relationship; in this regard, future studies could check whether the training through solving 3D puzzles facilitates the acquisition of general assembling skills.

On this point, the close relationship between motor and cognition can be utilized. As indicated on the meta-analyses, the mental practice before its attempt led to a significant positive effect onto the subsequent cognitive and physical performances (Driskell et al., 1994; Toth et al., 2020); the positive effect seemed to be stronger onto the cognitive task than the physical task (Driskell et al., 1994). However, the later positive effect onto the physical task differed among task types between internally cued movement and externally cued movement (Toth et al., 2020); the magnitude of mental practice effect through externally cued movement was indicatively comparable to that of cognitive task (Toth et al., 2020). In consideration of these points, it is expected that the cognitive training through solving 3D puzzles bring a more positive effect onto the physical task of assembling in the long-term especially if the task was set externally cued. This could be seen on field where parts are assembled on a conveyer belt step by step.

Education on stereotypes and the who-schema. Stereotypes and the who-schema. One might encounter during cognitive training, however, a psychological obstacle to sharpening the skill due to another factor: stereotypes. A recent meta-analysis reported that a gender difference where males perform better on mental rotation were modulated by the time frame in the test methodology (Maeda & Yoon, 2013). Additionally, the review proposed at least five factors to explain the discrepancy in mental rotation: biology, strategy, experience, affect, and test methodology (Maeda & Yoon, 2013). Regarding the affect factor, stereotypes of an individual or a group of individuals would be powerful to have an influence onto one's abilities because of its relevance to a schema; its repeated exposure further consolidates the schema of self or others, or who-schema for simplicity.

Indeed, a study on stereotypes reported that primary school children around eight or 10 years old already hold gender-stereotyped beliefs where better performance primarily lies on mental rotation plus math for the male and on the first language for the female (Moè, 2018). However, the stereotypes was not always aligned among the first-two tested tasks where a significantly better performance of the male was solely observed on mental rotation (Moè, 2018).

Stereotype effects. In this regard to the discrepancy between genders in mental rotation performance and its self-awareness, the stereotype threat effect and the stereotype lift effect may play a role (Rahe & Quaiser-Pohl, 2020), not in terms with a gender factor as discussed in the Discussion at Experiment 1 but an affect factor. The former effect was coined to conceptualize how stereotypes work as a negative primer in performance so that one becomes characteristic of a negatively-stereotyped group where the one belongs (Steele & Aronson, 1995). On the contrary, the latter effect was coined to conceptualize how stereotypes work as a positive primer in performance so that one becomes outperforming towards a negatively-stereotyped outgroup or a negatively-stereotyped task (Walton & Cohen, 2003). Thus, being primed either explicitly or implicitly through stereotypes seems influential to facilitate or impede performance, dependent on what to be referenced as norms (e.g., ethnicity or task).

On this point to the stereotype effects, stereotypes in vocational training may become a promotor in one way to prime better performance and to encourage skill acquisition (e.g., the stereotype lift due to an outgroup that has the lower score in solving puzzles). In contrast, it may become a suppressor in another way to prime worse performance and to discourage the acquisition (e.g., the stereotype threat due to the gender bias in mental rotation). Noted is that the stereotype lift effect may become detrimental to a whole group, because of its corresponding friction between groups that are created through a partition within the whole group.

Summary. With these accounts taken, it might be effective for instructors to become aware of the stereotype effects and to work on preventing adverse effects onto consolidating the who-schema (e.g., holding workshops on gender stereotypes). Besides, the association between mental rotation and assembling to differentiate capabilities would be better for its usage as a reference rather than a determiner.

The cognitive aptitude test on elemental abilities. Elasticity and plasticity. Another application to a society could be expanded into academic education, one of which is a cognitive aptitude test at entrance exams for detecting potential talents based on elemental abilities. In this way, a more fair chance could be given to those with socioeconomic disadvantages, who are likely not exposed to enough preparatory education (e.g., private tutor or cram school).

Indeed, a meta-analytic study indicated that the gap in academic performance has widened over decades in the US between the children from low-income families and those from high-income families (Reardon, 2011). The implication is crucial, and the trend was suggested to be led by increased parental investment into child education due to shifting demands on societal and individual levels over the years (Reardon, 2011). Noted is that the factor of parental education was indicated to be a stronger predictor to academic achievement than that of family income (Reardon, 2011). Nevertheless, both socioeconomic factors seem to have a crucial role in child education. In these regards, the psychological test on fundamental skills may be suited to weigh more on nature rather than nurture, or more broadly elasticity over plasticity, regardless of parental socioeconomic status.

The test battery on plasticity. However, it might have a drawback because elemental abilities could be still trained through nurture. On this point, there has to be an upper line on the ground level with less plasticity where the performance reaches a plateau even after intensive efforts. This becomes apparent when a small set of abilities is singled out for competition rather than a combined set (e.g., sprint and baseball). In addition, these abilities are commonly utilized daily and accessible to all citizens to be enhanced (e.g., spatial ability to throw a squashed ball of paper into a garbage bin).

In terms of elemental abilities, the psychological test of the present study requires neither numerical nor verbal comprehension, which should be permeable across ages. Besides, because the present psychological experiment analyzes the data in detail, it could be used as a supplementary or an alternative to the visuo-spatial task of cognitive tests that measures their abilities broadly, such as the Wechsler Adult Intelligence Scale (Wechsler, 2008).

As the name indicates, the test kit measures and quantifies intelligence; the latest inventory specifies scores on the following indexes: “Verbal Comprehension, Perceptual Reasoning, Working Memory, and Processing Speed” (Wechsler, 2008, p.8). These indexes further consist of core sets and supplementary subsets (Wechsler, 2008). The general intelligence can be calculated solely from the former sets (Wechsler, 2008). Noted is that the verbal and memory tasks require to be well aware of the concepts, or schemas, of both words and the methodology for arithmetic calculation. The procedure differs from the other tasks: the visual reasoning task requires to mentally or physically rotate objects; the speed task requires to find objects and to write the responses. These two tasks are permeable, or elemental, across ages regardless of how much rich education one has received to solidify schemas on the higher cognitive level. Noted is that the visual reasoning task uses an overall score that is free of task difficulties. Thus, it might be one way to differentiate individuals per the degree of difficulty by detecting with sensitivity those capable, as utilized in the present experiment.

Furthermore, but not strongly, excessive reliance onto the test battery may lead to a commercial monopoly as a hindrance to the benefits of citizens. This might generally occur when a citizen pays a high price, to receive a benefit from academic studies, to a mediator that is equipped for any what is priced by the sole provider. This situation would be detrimental to those with socioeconomic backgrounds, as touched in the discussion. However, it is reasonable for the developer to be given with credits because the development of intellectual products come with efforts and dedication. In these regards, it might be beneficial to have another standardized inventory to focus on elemental abilities regardless of socioeconomic status, mainly in consideration of the common good.

The test battery on elasticity. Even with these considerations, what to be measured for elemental abilities remains open, in consideration of the modularity of cognition. On this point, the interconnection between motor and cognition could be utilized to categorize main sets and subsets. Indeed, physical skills can be divided into the followings in one way: “agility, balance, coordination, speed, power, and reaction time” (Caspersen et al., 1985, p.128). In a similar way but more broadly, cognitive abilities should be also divided into main sets. Then, these main sets should be further divided into subsets. In the case of spatial abilities as the main set, it could be categorized as the followings: “spatial perception, mental rotation, and spatial visualization” (Linn & Petersen, 1985, p.1482). In this regard, the results from the present study can be utilized to measure one’s strength in mental rotation among the spatial skills, which is then to be used to estimate potential applied skills (e.g., assembling or math). Further blushing of the elemental cognition test is desirable. Future studies may summarize each set or dig into correlations between elemental abilities and also applied skills (e.g., designing architecture).

Arithmetic, mental rotation, and schemas. <u>Symbolic and non-symbolic arithmetic.</u> Another usage in education is to teach arithmetic. As touched on in the Introduction, recent studies found mixed evidence of mental rotation training to improve arithmetic performance in primary school children through training with mental rotation tasks (Cheng & Mix, 2014; Cheung et al., 2020; Hawes et al., 2015). In this regard, the puzzle task may be associated with arithmetic addition. This is because solving puzzles, the task of which is similar to building blocks, requires to mentally construct a meaningful whole by adding parts; the concept can be stated through the Gestalt theory by “The whole is something else than the sum of its parts” (Koffka, 1935, p.176). In an apparently similar way, the arithmetic addition might require to construct a meaningful whole by adding parts of Arabic numerals, which should be unique from that of mentally comparing two wholes.

There is likely, however, some underlying conceptual discrepancy between the puzzle task and the arithmetic addition in Arabic numerals. This is because each Arabic numeral is unique as an abstract symbol that points to a count. Thus, literally adding, or combining, these symbols is meaningless (e.g., 1 + 5 = 15 or 1 + 5 = 51). This is different from that of some Roman numerals (e.g., not I + V = IV but I + V = VI). In this regard, the calculation of Arabic numerals itself is not arithmetic but rather mnemonic.

Operands in the examples above are symbols, which are created artificial and to be conceptually operated; whereas, physical objects are existant to be physically or mentally operated. Thus, it makes no sense to those who are not familiar with the symbols when facing them, which is contrary to the non-symbols. Non-symbols and symbols are not separate but connected in that language abstractly integrate non-symbols as parts into a symbol as a whole; one symbol can stand for affluent meaning. These operands are then to be translated into another whole through symbolic or non-symbolic arithmetic.

<u>Exact non-symbolic arithmetic.</u> Arithmetic has four basic operators: addition, subtraction, multiplication, and division. The last two operators can be respectively compensated by addition and subtraction. Furthermore, arithmetic addition can compensate for the other three operators if negative operands are incorporated; however, minus counts are artificial, and it becomes a burden to grasp its practical purpose in early learners. Thus, the following discussion bases on arithmetic addition and subtraction with plus counts.

In the case of exact non-symbolic addition, the calculation would be processed by integrating two operands (e.g., • + ••••• = ••••••). In the very example, rows can help grasp the count by mentally transferring and fitting a dot (e.g., • + ••••• = • + ∷. = ⠿); (A) the process of mentally transferring and fitting a piece into the space is conceptually similar in part to that of solving puzzles.

The explanation of the very mental manipulation in the background is slightly different from a visual example, where two sets of tokens on a flat plane were approached in the arithmetic addition by overlapping visual patterns through mental imagery (e.g., ˙. + .˙ = ∷) (Hawes et al., 2015). The basis of this example would come from a field study, where indigenous Australian children whose language has not enough concept of counting words were reliant on the pattern strategy to make their mats by overlapping spatial layouts of tokens hidden under a cover (Butterworth et al., 2011). This is contrary to the outcome where their counterparts of urban children whose language has the concepts relied on the enumeration strategy to make their mats by aligning tokens randomly or linearly (Butterworth et al., 2011); noted is that audible counting was frequently accompanied during the very process (Butterworth et al., 2011).

In the case of exact non-symbolic subtraction, however, an alternative strategy could be taken. As an example to calculate dots through mental manipulation, each dot in the subtrahend can be mentally transferred to match with each dot in the minuend (e.g., •• is spatially matched with •••); the remaining unmatched dots equal to a difference (i.e., •). In this regard, (B) the matching task in mental rotation might be related in part with the exact non-symbolic arithmetic subtraction, in that two whole objects of parts are compared to see whether both match.

Non-symbols and mental rotation. As described in (A) and (B), mental rotation may be conceptually similar in part to non-symbolic arithmetic in methodology: specifically, FT and the addition; MT and the subtraction. However, similarity in degrees likely differs.

In the case of FT and the addition, the arithmetic addition requires to mentally transfer and fit one elemental unit into the puzzle space; whereas, FT requires to mentally rotate, mentally transfer and fit a part of elemental units into the space of the counterpart puzzle piece to create a whole of elemental units. Thus, the similarity in methodology seems to solely lie in the processes of mental transfer and part of fitting.

In the case of MT and the subtraction, the arithmetic subtraction requires to mentally transfer and match each elemental unit of two operands to find unmatched dots; whereas, MT requires to mentally rotate, mentally transfer, and match two wholes. Thus, the purpose of comparison might be different between them in their reaches of range. The distinction becomes more apparent through the Venn diagram (i.e., A covers B becomes A > B; A fails to cover B becomes A ≠ B; A covers B and vice versa, or A matches B, becomes A = B); MT only covers equality and inequality. Hence, the similarity seems to lie on the process of mental transfer and part of matching process.

In consideration of these points, mental rotation might be slightly connected to non-symbolic arithmetic; nevertheless, there has to be a better way to learn arithmetic, the point of which is to be taken into account. The calculation methodology in the examples discussed relied on mental manipulation in small counts. In this regard, there should be a maximum point where the spatial strategy works when processing large counts.

The speed-accuracy compatibility in searches. The mental operation to process information is conducted in either serial or parallel (Townsend, 1990). The serial processing is a sequential process where each object or subtask is worked on exclusively one at a time in the same average amount of time (Townsend, 1990). This differs from the parallel processing, which is a simultaneous process where multiple objects or subtasks are worked on at the same time (Townsend, 1990).

One example is seen graphically in the conjunction search, where reaction time was indicated to increase linearly per the spatial range of items (Treisman & Gelade, 1980); the order was suggested to be processed attentively in serial (Treisman & Gelade, 1980). The attentive process requires a foveal focus onto a target to exactly process a part of the whole information that is approximated at the start. The serial process differs from how features were indicated to be processed pre-attentively in parallel (Treisman & Gelade, 1980). These conjunction and feature searches are distinctive in how they are processed; however, the two processing could be explained through the feature-integration theory of attention to automatically process features in parallel and attentively integrate these features in serial into a complex whole (Treisman & Gelade, 1980).

Noted is that the serial processing model seems more general when the index in focus is reaction time (Townsend, 1990); whereas, the parallel processing model seems more general when the index in focus is accuracy (Townsend, 1990). In these regards, the two processes might be supplementary each other (Townsend, 1990). These notions could be applied into the feature-integration theory of attention: the parallel processing to process features as accurate as possible, whereas the serial processing to integrate features as fast as possible. This is reasonable because integration inevitably fails in its accuracy if what to be integrated are still obscure in its accuracy. In other words, accuracy in a static parallel process likely matters more than speed in a dynamic serial process within human cognition; the term static is meant for not motionless in the absolute domain but in the relative domain: namely, the law of inertia, or the first law of motion (Newton, 1687/1846). The interpretation well aligns with the ER-then-RT order of the priority-based rank-ordering technique.

Another example of processing information is also similarly seen graphically in the memory search, where reaction time was found to increase linearly per the number of items to mentally scan one-to-six items memorized in the short term to detect a target item (Sternberg, 1966). Though the serial processing was proposed for its explanation (Sternberg, 1966), the parallel processing was also proposed as another interpretation to mimic its linearity (Townsend, 1990).

The memory search shares with the conjunction search the same linearity between the reaction time and the degree of difficulty but might differ in how error rates proceed. In the conjunction search, higher error rates were indicated on a wider range of items (Treisman & Gelade, 1980). The result could be interpreted as the speed-accuracy compatibility that the very task condition on the conjunction search required a longer foveal walk to reach and to identify a target; subsequently, it came with errors per the task difficulty. The trend, however, seems different in the other task; low error rates were observed over all the trials with less than 1.5% in the memory search (Sternberg, 1966). On this point, however, the result might be because the intensity of the task was lenient in remembering the maximum of six items in the short term. This is indicative from a study, as touched on the Discussion at Experiment 1, where the working memory was suggested to have a finite capacity of around four chunks (Cowan, 2001). In another way, the participants might have utilized chunks to handle more items than their capacities. On the other hand, the participants in the memory search might have been with a larger capacity in working memory, because the study was participated by well-learned college students (Sternberg, 1966).

With these points taken, exact arithmetic in large non-symbolic counts would be also processed in a longer foveal walk and with a small working bag of memory; the consequence comes with more errors in accordance with the speed-accuracy compatibility.

The size of working memory and non-symbols. Indeed, a study on Amazonian indigenes whose language has apparent numeric concepts until five indicated that their performance was significantly lower than that of their French counterparts in exact non-symbolic arithmetic subtraction in dots (Pica et al., 2004); this was so even though the indigenes managed the approximate non-symbolic addition to compare two wholes in degrees comparable to that of their French counterparts (Pica et al., 2004). Noted is that the indigenes managed the minuend until three in the very subtraction above a chance level through a verbal response (Pica et al., 2004). This may imply that they could handle the exact subtraction not where operands are conceptualized until five through language but where the minuend and the subtrahend sum to around four (i.e., ••• − •• is manageable but more difficult in •••• − ••). This calculating process further implies that the exact subtraction itself in a small count was likely applied through the mental manipulation rather than the language manipulation (i.e., •••• − •• as it is rather than •••• – •• to 4 – 2).

The interpretation above partly came from that the calculation is restricted by the capacity of working memory, which was suggested to be around four chunks (Cowan, 2001); the language is simply a medium to convey the calculated outcome, regardless of its capability to have concepts until or beyond five. In other words, exact non-symbolic arithmetic subtraction in small counts seems to not necessarily require exact numerical language to calculate small non-symbolic counts. Thus, one is expected to behave in a similar way when responding non-verbally. Indeed, the interpretation is supported on a different task, where the same result in the very subtraction was replicated through a visual response (Pica et al., 2004). In consideration of these points, it is likely that the capacity of the working memory plays a role in exact non-symbolic arithmetic subtraction in small counts, regardless of numerical concepts through language.

A similar analogy could be applied to exact arithmetic addition, in line with the interpretation that the capacity of the working memory binds arithmetic calculation in small counts (i.e., ••• + •• is manageable but more difficult in •••• + ••). The operands in the former sums to five, which is around the capacity of four; whereas, those in the latter sums to six, more far from four. Thus, one is expected to make more mistakes in the latter arithmetic, regardless of whether the one has a language concept of six.

The interpretation states that exact non-symbolic arithmetic in small counts can be processed without the help of language; however, the calculation can be also applied through language manipulation.

Exact symbolic arithmetic in small counts. On the contrary to the non-symbolic arithmetic, the symbolic calculation goes that wholes of dots are translated into symbols, calculated through language manipulation, and then translated back into non-symbols (e.g., ••• and ••, to 3 and 2, to 3 + 2, to 5, and to •••••; ••• and ••, to 3 and 2, to 3 – 2, to 1, and to •). Noted is that it requires the methodological language concept to calculate two operands of symbols (i.e., 3 + 2 = 5 and 3 – 2 = 1).

Dynamic and static methodologies. The interpretations discussed above suggest that arithmetic calculation might be operated through a hierarchy of two systems: dynamic methodology through mental manipulation in a serial process and static methodology through language manipulation in a parallel process. The distinction is clear in that the former manipulation requires operating the spatial information of objects (i.e., physical stimuli on the cognitive ground level); whereas, the latter manipulation requires operating non-spatial information of objects (i.e., psychological concepts on the cognitive higher level).

However, the interpretation is restricted in small counts; thus, an additional operator is necessary to handle large counts: specifically, enumeration as a dynamic methodology. This operator seems crucial in arithmetic because both arithmetic addition and subtraction can benefit from the process; the operator makes it possible to grasp exact counts of operands through a serial process. In addition to enumeration, handling non-symbols in an operand as a whole, or a chunk, rather than parts as wholes should be also crucial because of the capacity restriction of the working memory (i.e., spatial chunking through layers or non-spatial chunking through language).

Exact arithmetic in large counts. The term chunk is defined as "a collection of concepts that have strong associations to one another and much weaker associations to other chunks concurrently in use" (Cowan, 2001, p.89). This notion could be helpful in exact non-symbolic arithmetic in large counts. For example, large dots can be separated for later calculation (e.g., •••••, to ••• and ••, and to a chunk of ••• and a chunk of ••). In the example, the original chunks of five was translated into the two separate chunks, which becomes manageable. Larger counts can be attained by utilizing the maximum concept in counts and conceptualizing the added-one as part of enumeration (e.g., •••••, to five, to five + •, and to a new concept of six)

Conceptually, arithmetic in large counts could be separated into chunks: the concept chunk, the receiver chunk, the transit chunk, and the sender chunk; then, they operate in parallel whereas transferring operands and enumeration operate in serial. Each dot in the transit chunk from the sender chunk is mentally transferred into the receiver chunk to be cumulatively enumerated to update the concept chunk (e.g., ••••••, to •••• + ••, to the receiver chunk of •••• connected to the concept chunk of four + the transit chunk of • + the sender chunk of •, to the receiver chunk of ••••• connected to the concept chunk of five + the transit chunk of •, and to the receiver chunk of •••••• connected to the concept chunk of six). The last chunks can be virtually treated as one because the receiver chunk and the concept chunk have a strong association to one another.

With these accounts taken into account, another strategy other than the enumeration strategy might have been utilized by the Amazonian indigenes: specifically, the methodology to see the non-symbols as a whole, as in the pattern strategy by the indigenous Australian children (Butterworth et al., 2011). Indeed, the Amazonian indigenes seems to rarely use fingers to perceive counts (Pica et al., 2004). Noted is that the indigenous Australian children also did not utilize fingers to perceive counts (Butterworth et al., 2011). Folding fingers represent enumeration process, and the folded fingers can represent concepts (e.g., two folded fingers represent the number of two). Hence, both enumeration in exact non-symbolic arithmetic and linking the non-symbolic counts as a chunk to a numerical concept through language likely play a role in arithmetic addition in large counts. Noted is that the pattern strategy could be interpreted as overlapping or counting layers in serial. In this regard, the pattern strategy could be considered as spatial enumeration, which is unique from arithmetic enumeration. The spatial enumeration seems analogically similar to the attentive serial process to integrate features into a complex whole in the feature-integration theory of attention.

In the case of the subtraction, the strategy with chunks and enumeration in the addition can be utilized to create two sets; these two sets are then to be mentally matched to find the unmatched, as in the Venn diagram in small counts. The process likely takes efforts, and language manipulation should be a better strategy.

Mental and language manipulations. With these accounts considered, both language and mental manipulations would be effective, but supposedly more influential through language in consideration of the advantage in conceptualization on the cognitive higher level. However, the language manipulation implicitly bases on the mental manipulation on the cognitive ground level; thus, both systems working in tandem to logically understand the concepts to link non-symbolic counts to symbols and also to understand the concepts of methodological calculations should facilitate their arithmetic learning (i.e., the what-schema to link non-symbols to symbols; the how-schema for dynamic and static methodologies). This differs from simply memorizing the numerical language.

School children are to learn arithmetic early in education. The whole procedure requires steps and consumes mental efforts through the transformation between numeral concepts and mental images. In response, one might want to simply, in line with the decimal system, memorize semantically and procedurally the followings: basic Arabic numbers and the calculation in the first digit; the column method in large numbers (e.g., addition with carry-over). These memorizations could be taken as consolidating the what-schemas and the how-schemas as mentioned. In the memorizing process, children would prefer the logic, or the order of cause and effect, why they are so conceptually: namely, the why-schema for simplicity. This process is operated by counting and ordering operands of facts (i.e., logic enumeration), and it should be more preferred than being urged for no reason to memorize.

Arithmetic enumeration and logic enumeration apparently differ in whether operands are identifiable (i.e., combination and permutation). If unidentifiable in a whole, it becomes arithmetic enumeration (i.e., counting operands); whereas, if identifiable in a whole, it becomes logic enumeration (i.e., counting and ordering operands). In consideration of these accounts, non-symbolic support as a supplementary in a logical explanation could help consolidate the why-schema and assist the learning process as mentioned.

One downside is that mental and language manipulations may be tough to early starters in arithmetic, especially in that they have to count from one to unknown in large counts. On this point, it could be an effective strategy to guess approximately how many counts are existant there. In this way, they can grasp an approximate sense of confidence with an estimated goal in mind, as in efficiently scheduling physical efforts in a marathon with the solid goal line yet vaguely away from the start line; it should be chaotic to start an unknown distance without grasping approximately where the goal line stands.

Approximate arithmetic in large counts. Indeed, the effectiveness in approximate arithmetic is indicatively supported by an interventional study, where preschool children with a small knowledge of concepts in arithmetic significantly improved more through non-symbolic approximate arithmetic training with homogeneous items in 10 daily sessions over two-to-three weeks (Park et al., 2016); the effect size was interpreted as small (Cohen, 1988). Thus, approximate arithmetic seems related to exact arithmetic in general. In light of this, approximation may play a role in arithmetic as a chunk. It should be noted that part of the training might have utilized exact arithmetic in the study, in that the children were instructed to imagine the total counts after arithmetic addition or subtraction, as the bumper task, right before all the four tasks of comparing or matching two sets (i.e., to imagine total items dropped into or taken from a bucket through two arrays right before comparing or matching the count in the bucket with the count in another bucket) (Park et al., 2016); each operand in the two arrays ranged from 4 to 16 items in the addition or from 12 to 40 in the subtraction in approximately 87% of the trials (Park et al., 2016). Thus, there might be some room for further investigation to see whether approximate arithmetic in large counts help better arithmetic and also to see whether approximation is indeed utilized for a goal-directed approach in exact arithmetic.

Summary. In consideration of these accounts, the computational process of mental rotation might be indirectly, not directly, related in part to arithmetic to help consolidate the schemas. On this point, the intervention to educate arithmetic through mental rotation in solving puzzles for a whole and in comparing wholes might be respectively effective in arithmetic addition and subtraction to a small degree slightly in effect sizes. However, exact arithmetic training through dynamic and static methodologies, in addition to approximate non-symbolic arithmetic training, is likely more effective in consolidating the schemas, which eventually comes with better performance.

The schemas of the seven circumstances. Rationale. The aforementioned discussion implies a close link between language and schemas. Even though the language apparently fails to explain some behaviors in arithmetic, it still holds a crucial role to sum parts into a meaningful whole. Nevertheless, a question arises about from where these parts derive; signals might directly come through sensation, indirectly after the integration of sensation, or other systematic routes. Either way, there would be reasonable to have ponds of parts to create a meaningful whole at the end point of a flow. The sources might be this or that based on each interpretation; however, the answer from each interpretation to a question should end up in the same state regardless of how to be solved. With this point into account, the close interconnection between language and concept could be utilized.

Language conceptualizes parts into a meaningful whole as seen defined in a dictionary book. When one wants to see the definition of a word, the one can bring it and flip its pages to find the definition. However, the process could be also performed by giving questions about the same definition of the word to a dictionary with artificial intelligence: namely, seven questions: who, what, where, when, with what or who, why, and how. These conceptual elements derive from the *Nicomachean Ethics* by Aristotle (Sloan, 2010). More specifically, the circumstances of an act is divided into the followings: "(1) the who, (2) the what, (3) around what place or (4) in which time something happens, and sometimes (5) with what, such as an instrument, (6) for the sake of what, such as saving a life, and (7) the how, such as gently or violently" (Sloan, 2010, p.239); any ignorance of these kinds appears to have one act involuntarily (Sloan, 2010). These circumstances are to describe what makes an involuntary action, but it could be used to question and then describe what makes the one act involuntarily; or more broadly, it could be used to question from the point of psychology through philosophy and then describe physical plus psychological behaviors or psychological concepts that both behaviors derive from. Because of the apparently close connection between psychology and philosophy, a proper term might be necessary for clarity, as in psychophysics, which is termed psychophilosophy in the present paper. Furthermore, these elemental psychophilosophical kinds are termed the schemas of the seven circumstances for clarity. Hence, every concept, or schema, is to be described from the elemental seven schemas.

Limitations. Noted is that the number has not to be specific to this seven and might fluctuate. This is because the concept is likely more widely used, as in five W's (and one H) for journalism in one view (Sloan, 2010), or more of 5W1H through the observations through experience. Nevertheless, the seven questions is sustainable to look into the circumstances as long as every circumstance can be described by the kinds. The interpretation here goes with the seven questions.

The questions are useful to describe behaviors or concepts, but there should be more limitations. One additional limitation is that the kinds of the seven circumstances could be all grounded onto the what after rephrased: namely, at what place, at what time, with what or who, on what basis, and in what way. Additionally, the who could be taken as the what because the physical vessel as a person is basically an object. In consideration of these points, the what could be interpreted as the primary schema in the schemas of the seven circumstances. To add more, a new schema of the questions could be added with prepositions and, conveniently, postpositions, when there arises a circumstance indescribable by the seven schemas. However, as mentioned before, the seven questions would be sustainable as long as each circumstance can be described by the kinds.

Another limitation resides in a range of its application restricted within human cognition which may yet encounter another additional circumstance. On this point, however, it may be still beneficial on a practical basis to describe the circumstances of most beings; the present likely-superiority of human cognition as the operator of language might work as a whole to cover parts of the other currently known beings. This is simply to indicate one part of the superior modularities restricted in the animal kingdom. All livings are superior in that they are existant, but their superiority resides in other modules or a mixture of modules, some far exceeding that of human beings. The concept is clear on a diagram, where each whole of organisms resides on leaves of the phylogenetic tree (Darwin, 1859).

One more limitation is that the sources might be better to be defined in another way: namely, through physics. Because our parts are in a physical domain, it is reasonable for every entity to be expressed through natural elemental units, or the seven units through the International System of Units in physics (International Bureau of Weights and Measures, 2019): specifically, "second, metre, kilogram, ampere, kelvin, mole, and candela" (International Bureau of Weights and Measures, 2019, p.130). These units are respectively applied in degrees to express physical concepts: "time, length, mass, electric current, thermodynamic temperature, amount of substance, and luminous intensity" (International Bureau of Weights and Measures, 2019, p.130). It is a nice coincidence that they share the same number of seven in the present physics with that in ancient philosophy to describe behaviors of physical beings. It might also be similar in that the what is fundamental to construct the seven physical dimensions (e.g., the concept of length by connecting two points).

The niche segregation. However, there likely lies a difference between the two systems. This is because one psychophilosophical kind is unique to have what the physical kinds have not: namely, the why. Indeed, the why stands out because of its subjectivity among the domain of other circumstances that bases on objectivity (Sloan, 2010); the objection to its listing was anticipated by Aristotle (Sloan, 2010). The distinction could be seen in a role-playing game, where a story proceeds in a hypothetical world based on the settings that transcend the bounds of the laws in physics; behaviors become unexplainable by the kinds in physics, whereas the psychophilosophical kinds can explain the behavior. To elaborate more in detail, the kinds of physics can describe the behaviors of the hardware but fails to describe the very behavior of the software; on the contrary, the kinds of psychophilosophy can describe both behaviors but abstractly. Hence, it could be that the psychophilosophical kinds operate on plasticity, whereas the physical kinds operate on elasticity.

With the two systems separated, parts that come from the seven physical concepts is to be received as a meaningful whole at the endpoint of a flow by the seven psychophilosophical schemas, such as through the Gestalt principles of grouping after the feature-integration theory of attention. Because the two systems apparently treat unique domains, it may become practically beneficial to have the schemas of the seven circumstances to describe circumstances in the plastic psychophilosophical level whereas the physical seven units describe circumstances below the schemas from the elastic physical level.

Summary. To summarize, it would be beneficial to have a similar standardized system in the domain of psychology in consideration of the kind why, unique from the domain of physics. The schemas of the seven circumstances are deliberate and artificial; however, it could become one way to describe our cognitive dimensions from elemental and constructive standard units, as in philosophy, mathematics, and physics. By its usage, every schema is to be described through a mixture of the seven elemental schemas: namely, who-schema, what-schema, where-schema, when-schema, with-schema, why-schema, and how-schema. The interpretation could be useful for practical purposes as its rough application into a society (e.g., vocational rehabilitation, educational rehabilitation, social rehabilitation, and medical rehabilitation). For future studies, further blushing of the interpretation is desirable (e.g., homeostatic default state to have positive or negative schemas, representative schemas, and representative probability).

The assembling model of schemas. The discussions above mentioned multiple interpretations, which likely makes their comprehension more difficult; subsequently, it likely comes with higher error rates in a longer reaction time in line with the speed-accuracy compatibility. In consideration of this point, general ideas are integrated into a proposed model for parsimony to permeable explain the behaviors: the assembling model of schemas (See Figure 7).

The assembling line proceeds that a black box as a whole from the cognitive underground level is brought into the Sender on the cognitive ground level, above which lies the cognitive higher level. The information of the whole is transmitted through the Approximation to be representatively approximated to the Enumeration. Operands of black parts of the whole is mentally transferred into the Transit to be differentiated. The information of each part as a whole is transmitted through the Exaction to be representatively exacted to the Enumeration. The sifted parts as wholes are mentally transferred into the Receiver. The information of the parts as a whole is transmitted through the Exaction to be representatively exacted to the Enumeration. The gathered

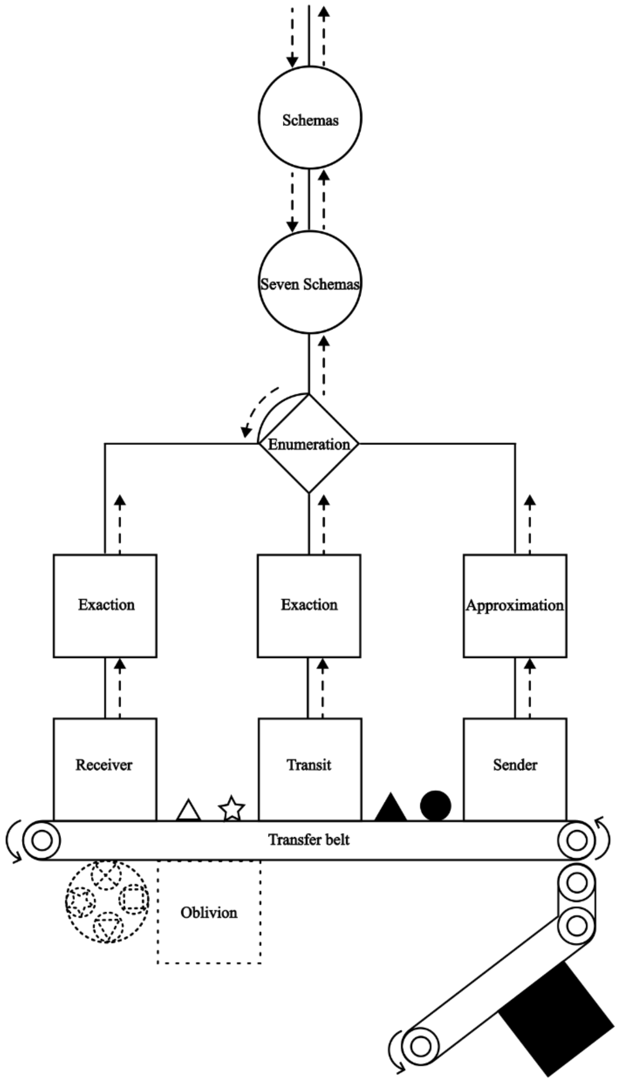

Figure 7. The schematic diagram of the assembling model of schemas. The dashed arrows and the solid circular arrows respectively indicate the direction of signals and the direction of rotation. Dotted shapes and the Oblivion are described for emphasis. Note. Operators, as described from left to right and from bottom to top. Oblivion (the state of oblivion). Transfer belt (Mental transfer). Receiver (the Receiver chunk). Transit (the Transit chunk). Sender (the Sender chunk). Exaction (the Receiver-exaction chunk). Exaction (the Transit-exaction chunk). Approximation (the Sender-approximation chunk). Enumeration (spatial, arithmetic, or logic enumeration). Severn Schemas (the schemas of the seven circumstances). Schemas (schemas).

parts of information in the Enumeration is to be enumerated if conditions are met. The funneled whole of information in the Enumeration is to be transmitted based on conditions to itself for further enumeration, for feedback, or to the Seven Schemas for its consolidation. A schema on the Schemas is consolidated based on the elemental schemas on the Seven Schemas. Physical or cognitive behaviors derive from their schemas on the Schemas. The worked integrated whole, unintegrated parts, and their representations each returns to the Oblivion within a time frame. Not all operators, but the Sender and the Approximation as static methodologies besides the Transfer belt and the Enumeration as dynamic methodologies, require activation to consolidate schemas. The Exaction chunks can cover the Receiver chunk and the Transit chunk, which virtually reduces the number to four chunks in full operation.

Spatial enumeration bases on operands of spatial layers through approximation as intuition and exaction as focus attention. Meanwhile, arithmetic enumeration bases on operands of non-differentiable non-symbols through approximation as intuition and exaction as language. Furthermore, logic enumeration bases on operands of differential facts through approximation as intuition and exaction as verification.

More broadly, learning bases on enumeration with operands through the approximation of a whole and the exaction of the whole as parts. The disparity between the approximation and the exaction is to become narrower through feedback.

Limitation

Questionnaire. The data was planned to detect a tactical difference in how the two tasks were worked on. However, the questions might have been different from what was intended. This is because the internalization of reference 3D objects was reasonable to be shared between the two tasks. Besides, more count data were needed for analyzing three factors. In consideration of these points, acquiring either continuous data through new behavioral or physiological methodology, in addition to another data type through the Likert scale with rank data, would be more desirable.

P-value correction. The Bonferroni-Holm or the Bonferroni corrections were utilized throughout data analysis regardless of the data in interests, and this is because of being exploratory as a post hoc analysis. Future studies could use a pre-planned procedure for correcting p-values on a priority basis.

AT. The task required to assemble three target stimuli in total. However, it might have been short of assigning each participant to a unique rank slot. Preparing more target stimuli would further differentiate individual capabilities in assembling blocks.

Conclusion

The present study indicated that mental rotation is utilized to fit a 3D object into its 3D puzzle piece and that the task deviates from that of matching 3D objects. Additionally, it implied that the capability of fitting a 3D object into its 3D puzzle piece among individuals strongly associates with that of assembling blocks. Besides, it indicated the potential of a priority-based rank-ordering technique to predict individual capabilities. A more rigorous and fine-tuned design of experiments with an estimated sample size would shed more light onto the results acquired. Furthermore, the assembling model of schemas could help interpret our cognition and also support practitioners on field.

Compliance with ethical standards

Ethics approval. The ethics approval was provided by the ethics committee of the Kyoto University Graduate School of Medicine. The present study was conducted in line with the Declaration of Helsinki.

Funding. The present study was part of research supported by JSPS KAKENHI Grant Number 19K19832.

Conflict of interests. The present study is to be reviewed partly by a group member of the aforementioned research where the author is registered. Furthermore, all the reviewers are to come from the Division of Occupational Therapy, to which the author belongs, at the Course of Advanced Rehabilitation Sciences.

Informed consent. Informed consent was written and provided by each participant.

Physical and mental integrity. There was no adverse effect reported.

Remuneration. After the experiment, each participant was compensated with a voucher card worth above the local minimum hourly wage.

Acknowledgements

The author thanks Dr. Hiroyasu Abe, Dr. Ryuta Iseki, and Dr. David Croydon for their assistance in inquiries about statistics. The author thanks Dr. Hiroyuki Muto and Dr. Jun Saiki for their advice on brief results and discussion. The author thanks an anonymous contributor for their assistance in an inquiry about editing the manuscript. The author thanks an anonymous contributor for the ethics application of the present study. The author thanks Hiroyuki Inadomi, Ph.D., and Nang Liang, Ph.D., in general.

References

Aberson, C. L. (2019). *Applied Power Analysis for the Behavioral Sciences* (2nd ed.). Routledge. https://doi.org/10.4324/9781315171500

Aberson, C. L. (2022). *pwr2ppl: Power Analyses for Common Designs (Power to the People)* (0.5.0) [R package]. https://CRAN.R-project.org/package=pwr2ppl

Agamben, G. (1999). 15. Bartleby, or On Contingency. In D. Heller-Roazen (Ed.), *Potentialities: Collected Essays in Philosophy* (pp. 241–274). Stanford University Press. https://doi.org/10.1515/9780804764070-017

Bandura, A. (2004). Observational Learning. In J. H. Byrne (Ed.), *Learning and Memory* (2nd ed., pp. 482–484). Macmillan Reference USA.

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge University Press.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436. https://doi.org/10.1163/156856897X00357

Breslow, N. E., & Day, N. E. (1980). *Statistical methods in cancer research. Volume I - The analysis of case-control studies*. IARC Scientific Publication.

Brosnan, M. J. (1998). Spatial Ability in Children's Play with Lego Blocks. *Perceptual and Motor Skills*, *87*(1), 19–28. https://doi.org/10.2466/pms.1998.87.1.19

Butterworth, B., Reeve, R., & Reynolds, F. (2011). Using mental representations of space when words are unavailable: Studies of enumeration and arithmetic in indigenous Australia. *Journal of Cross-Cultural Psychology*, *42*, 630–638. https://doi.org/10.1177/0022022111406020

Caldwell, A., & Lakens, D. (2022). *Superpower: Simulation-Based Power Analysis for Factorial Designs* (0.2.0) [R package]. https://CRAN.R-project.org/package=Superpower

Caspersen, C. J., Powell, K. E., & Christenson, G. M. (1985). Physical activity, exercise, and physical fitness: Definitions and distinctions for health-related research. *Public Health Reports*, *100*(2), 126–131.

Cheng, Y.-L., & Mix, K. S. (2014). Spatial Training Improves Children's Mathematics Ability. *Journal of Cognition and Development*, *15*(1), 2–11. https://doi.org/10.1080/15248372.2012.725186

Cheung, C.-N., Sung, J. Y., & Lourenco, S. F. (2020). Does training mental rotation transfer to gains in mathematical competence? Assessment of an at-home visuospatial intervention. *Psychological Research*, *84*(7), 2000–2017. https://doi.org/10.1007/s00426-019-01202-5

Chu, M., & Kita, S. (2011). The nature of gestures' beneficial role in spatial problem solving. *Journal of Experimental Psychology: General*, *140*, 102–116. https://doi.org/10.1037/a0021790

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Routledge. https://doi.org/10.4324/9780203771587

Community. (2021). *Blender* (2.93.0) [Computer software]. Blender Foundation.

Cooper, L. A., & Shepard, R. N. (1975). Mental transformation in the identification of left and right hands. *Journal of Experimental Psychology: Human Perception and Performance*, *1*, 48–56. https://doi.org/10.1037/0096-1523.1.1.48

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114. https://doi.org/10.1017/S0140525X01003922

Dahm, S. F., Muraki, E. J., & Pexman, P. M. (2022). Hand and Foot Selection in Mental Body Rotations Involves Motor-Cognitive Interactions. *Brain Sciences*, *12*(11), Article 11. https://doi.org/10.3390/brainsci12111500

Darwin, C. R. (1859). *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life* (1st ed.). John Murray.

Diedenhofen, B. (2022). *cocor: Comparing Correlations* (1.1-4) [R package]. https://CRAN.R-project.org/package=cocor

Diedenhofen, B., & Musch, J. (2015). cocor: A Comprehensive Solution for the Statistical Comparison of Correlations. *PLOS ONE*, *10*(4), e0121945. https://doi.org/10.1371/journal.pone.0121945

Driskell, J. E., Copper, C., & Moran, A. (1994). Does mental practice enhance performance? *Journal of Applied Psychology*, *79*, 481–492. https://doi.org/10.1037/0021-9010.79.4.481

Duschinsky, R. (2012). "Tabula Rasa" and Human Nature. *Philosophy*, *87*(342), 509–529. https://doi.org/10.1017/S0031819112000393

Elkin, L. A., Kay, M., Higgins, J. J., & Wobbrock, J. O. (2021). An Aligned Rank Transform Procedure for Multifactor Contrast Tests. *The 34th Annual ACM Symposium on User Interface Software and Technology*, 754–768. https://doi.org/10.1145/3472749.3474784

Fechner, G. T. (1860/1912). ELEMENTS OF PSYCHOPHYSICS. In B. Rand (Ed.), & H. S. Langfeld (Trans.), *The classical psychologists: Selections illustrating psychology from Anaxagoras to Wundt* (pp. 562–572). Houghton, Mifflin and Company. https://doi.org/10.1037/10885-000

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, *47*, 381–391. https://doi.org/10.1037/h0055392

Hawes, Z., Moss, J., Caswell, B., & Poliszczuk, D. (2015). Effects of mental rotation training on children's spatial and mathematics performance: A randomized controlled study. *Trends in Neuroscience and Education*, *4*(3), 60–68. https://doi.org/10.1016/j.tine.2015.05.001

International Bureau of Weights and Measures. (2019). *Le Système international d'unités [The International System of Units]* (9th ed.). BIPM.

International Labour Organisation & International Labour Conference. (1983). *Convention 159: Convention concerning vocational rehabilitation and employment (disabled persons).* 1–4.

Ionta, S., Fourkas, A. D., Fiorio, M., & Aglioti, S. M. (2007). The influence of hands posture on mental rotation of hands and feet. *Experimental Brain Research*, *183*(1), 1–7. https://doi.org/10.1007/s00221-007-1020-2

Kanata, T. (2016). JAPANESE MENTAL HEALTH CARE IN HISTORICAL CONTEXT: WHY DID JAPAN BECOME A COUNTRY WITH SO MANY PSYCHIATRIC CARE BEDS? *Social Work/Maatskaplike Werk*, *52*(4), Article 4. https://doi.org/10.15270/52-4-526

Kassambara, A. (2022a). *rstatix: Pipe-Friendly Framework for Basic Statistical Tests* (0.7.1). https://CRAN.R-project.org/package=rstatix

Kassambara, A. (2022b). *ggpubr: "ggplot2" Based Publication Ready Plots* (0.5.0) [R package]. https://CRAN.R-project.org/package=ggpubr

Kay, M. (2021, October 12). *Effect Sizes with ART*. The Comprehensive R Archive Network. https://CRAN.R-project.org/web/packages/ARTool/vignettes/art-effect-size.html

Kay, M., Elkin, L. A., Higgins, J. J., & Wobbrock, J. O. (2021). *ARTool: Aligned Rank Transform* (0.11.1) [R package]. https://CRAN.R-project.org/package=ARTool

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception*, *36*(1_suppl), 14. https://doi.org/10.1177/03010066070360S101

Koffka, K. (1935). *Principles of Gestalt psychology*. Harcourt, Brace and Company.

Kosslyn, S. M., Ball, T. M., & Reiser, B. J. (1978). Visual images preserve metric spatial information: Evidence from studies of image scanning. *Journal of Experimental Psychology: Human Perception and Performance*, *4*, 47–60. https://doi.org/10.1037/0096-1523.4.1.47

Lakens, D., & Caldwell, A. R. (2021). Simulation-Based Power Analysis for Factorial Analysis of Variance Designs. *Advances in Methods and Practices in Psychological Science*, *4*(1), 2515245920951503. https://doi.org/10.1177/2515245920951503

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764–766. https://doi.org/10.1016/j.jesp.2013.03.013

Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development*, *56*, 1479–1498. https://doi.org/10.2307/1130467

Lyu, Y., Guo, X., Bekrater-Bodmann, R., Flor, H., & Tong, S. (2016). Phantom limb perception interferes with motor imagery after unilateral upper-limb amputation. *Scientific Reports*, *6*, 21100. https://doi.org/10.1038/srep21100

Maeda, Y., & Yoon, S. Y. (2013). A Meta-Analysis on Gender Differences in Mental Rotation Ability Measured by the Purdue Spatial Visualization Tests: Visualization of Rotations (PSVT:R). *Educational Psychology Review*, *25*(1), 69–94. https://doi.org/10.1007/s10648-012-9215-x

McDonald, J. H. (2014). *Handbook of Biological Statistics* (3rd ed.). Sparky House Publishing.

Miller, A. E. J., MacDougall, J. D., Tarnopolsky, M. A., & Sale, D. G. (1993). Gender differences in strength and muscle fiber characteristics. *European Journal of Applied Physiology and Occupational Physiology*, *66*(3), 254–262. https://doi.org/10.1007/BF00235103

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81–97. https://doi.org/10.1037/h0043158

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behavior*. Henry Holt and Co. https://doi.org/10.1037/10039-000

Moè, A. (2018). Mental rotation and mathematics: Gender-stereotyped beliefs and relationships in primary school children. *Learning and Individual Differences*, *61*, 172–180. https://doi.org/10.1016/j.lindif.2017.12.002

Mutlu, M. C., Erdoğan, S. B., Öztürk, O. C., Canbeyli, R., & Saybaşılı, H. (2020). Functional Near-Infrared Spectroscopy Indicates That Asymmetric Right Hemispheric Activation in Mental Rotation of a Jigsaw Puzzle Decreases With Task Difficulty. *Frontiers in Human Neuroscience*, *14*. https://doi.org/10.3389/fnhum.2020.00252

Myers, L., & Sirois, M. J. (2006). Spearman Correlation Coefficients, Differences Between. In S. Kotz, C. B. Read, N. Balakrishnan, & B. Vidakovic (Eds.), *Encyclopedia of Statistical Sciences* (2nd ed., Vol. 12, pp. 7901–7903). Wiley-Interscience.

Newton, I. (1687/1846). *Newton's Principia: The mathematical principles of natural philosophy* (A. Motte, Trans.). Daniel Adee.

Nicholls, M. E. R., Thomas, N. A., Loetscher, T., & Grimshaw, G. M. (2013). The Flinders Handedness survey (FLANDERS): A brief measure of skilled hand preference. *Cortex*, *49*(10), 2914–2926. https://doi.org/10.1016/j.cortex.2013.02.002

Okubo, M., Suzuki, H., & Nicholls, M. E. R. (2014). 日本語版 FLANDERS 利き手テスト――信頼性と妥当性の検討―― [A Japanese version of the FLANDERS handedness questionnaire]. *Shinrigaku Kenkyu: The Japanese Journal of Psychology*, *85*(5), 474–481. https://doi.org/10.4992/jjpsy.85.13235

Park, J., Bermudez, V., Roberts, R. C., & Brannon, E. M. (2016). Non-symbolic approximate arithmetic training improves math performance in preschoolers. *Journal of Experimental Child Psychology*, *152*, 278–293. https://doi.org/10.1016/j.jecp.2016.07.011

Parsons, L. M. (1987). Imagined spatial transformations of one's hands and feet. *Cognitive Psychology*, *19*(2), 178–241. https://doi.org/10.1016/0010-0285(87)90011-9

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442. https://doi.org/10.1163/156856897x00366

Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and Approximate Arithmetic in an Amazonian Indigene Group. *Science*, *306*(5695), 499–503. https://doi.org/10.1126/science.1102085

Pind, J. L. (2012). Looking back: Figure and ground at 100. *The Psychologist*, *25*(1), 90–91. University of Copenhagen.

R Core Team. (2022). *R: A language and environment for statistical computing* (4.2.1) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Rahe, M., & Quaiser-Pohl, C. (2019). Mental-rotation performance in middle and high-school age: Influence of stimulus material, gender stereotype beliefs, and perceived ability of gendered activities. *Journal of Cognitive Psychology*, *31*(5–6), 594–604. https://doi.org/10.1080/20445911.2019.1649265

Rahe, M., & Quaiser-Pohl, C. (2020). Cubes or Pellets in Mental-Rotation Tests: Effects on Gender Differences and on the Performance in a Subsequent Math Test. *Behavioral Sciences*, *10*(1), Article 1. https://doi.org/10.3390/bs10010012

Rahe, M., Ruthsatz, V., & Quaiser-Pohl, C. (2021). Influence of the stimulus material on gender differences in a mental-rotation test. *Psychological Research*, *85*(8), 2892–2899. https://doi.org/10.1007/s00426-020-01450-w

Reardon, S. F. (2011). The Widening Academic Achievement Gap Between the Rich and the Poor: New Evidence and Possible Explanations. In G. J. Duncan & R. J. Murnane (Eds.), *Whither Opportunity?: Rising Inequality, Schools, and Children's Life Chances* (pp. 91–116). Russell Sage Foundation.

Rosner, T., Grasso, A., Scott-Cole, L., Villalobos, A., & Mulcahey, M. (2020). Scoping Review of School-to-Work Transition for Youth With Intellectual Disabilities: A Practice Gap. *The American Journal of Occupational Therapy*, *74*(2), 7402205020p1-7402205020p23. https://doi.org/10.5014/ajot.2019.035220

Rubin, E. (1915). *Synsoplevede Figurer: Studier i psykologisk Analyse. Første Del [Visually experienced figures: Studies in psychological analysis. Part one]*. Gyldendalske Boghandel, Nordisk Forlag.

Ruthsatz, V., Neuburger, S., Jansen, P., & Quaiser-Pohl, C. (2014). Pellet Figures, the Feminine Answer to Cube Figures? Influence of Stimulus Features and Rotational Axis on the Mental-Rotation Performance of Fourth-Grade Boys and Girls. *Spatial Cognition IX*, 370–382. https://doi.org/10.1007/978-3-319-11215-2_26

Safaei, A., Rahmanian, M., Oraki, M., & Zinchenko, A. (2022). Video Game Play Does Not Improve Spatial Skills When Controlling for Speed-Accuracy Trade-Off: Evidence From Mental-Rotation and Mental-Folding Tasks. *Perceptual and Motor Skills*, *129*(3), 488–512. https://doi.org/10.1177/00315125221078982

Sekiyama, K. (1982). Kinesthetic aspects of mental representations in the identification of left and right hands. *Perception & Psychophysics*, *32*(2), 89–95. https://doi.org/10.3758/BF03204268

Sekiyama, K. (1983). Mental and physical movements of hands: Kinesthetic information preserved in representational systems. *Japanese Psychological Research*, *25*, 95–102. https://doi.org/10.4992/psycholres1954.25.95

Shepard, R. N., & Metzler, J. (1971). Mental Rotation of Three-Dimensional Objects. *Science*, *171*(3972), 701–703. https://doi.org/10.1126/science.171.3972.701

Signorell, A. (2022). *DescTools: Tools for Descriptive Statistics* (0.99.47) [R package]. https://CRAN.R-project.org/package=DescTools

Sloan, M. C. (2010). Aristotle's Nicomachean Ethics as the Original Locus for the Septem Circumstantiae. *Classical Philology*, *105*(3), 236–251. https://doi.org/10.1086/656196

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, *69*, 797–811. https://doi.org/10.1037/0022-3514.69.5.797

Sternberg, S. (1966). High-Speed Scanning in Human Memory. *Science*, *153*(3736), 652–654. https://doi.org/10.1126/science.153.3736.652

Tomasino, B., & Gremese, M. (2016). Effects of Stimulus Type and Strategy on Mental Rotation Network: An Activation Likelihood Estimation Meta-Analysis. *Frontiers in Human Neuroscience*, *9*. https://doi.org/10.3389/fnhum.2015.00693

Toth, A. J., McNeill, E., Hayes, K., Moran, A. P., & Campbell, M. (2020). Does mental practice still enhance performance? A 24 Year follow-up and meta-analytic replication and extension. *Psychology of Sport and Exercise*, *48*, 101672. https://doi.org/10.1016/j.psychsport.2020.101672

Townsend, J. T. (1990). Serial vs. Parallel Processing: Sometimes They Look like Tweedledum and Tweedledee but they can (and Should) be Distinguished. *Psychological Science*, *1*(1), 46–54. https://doi.org/10.1111/j.1467-9280.1990.tb00067.x

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136. https://doi.org/10.1016/0010-0285(80)90005-5

Walton, G. M., & Cohen, G. L. (2003). Stereotype lift. *Journal of Experimental Social Psychology*, *39*, 456–467. https://doi.org/10.1016/S0022-1031(03)00019-2

Weaver, B., & Wuensch, K. L. (2013). SPSS and SAS programs for comparing Pearson correlations and OLS regression coefficients. *Behavior Research Methods*, *45*(3), 880–895. https://doi.org/10.3758/s13428-012-0289-7

Weber, E. H. (1834/1912). THE SENSE OF TOUCH AND THE COMMON FEELING. In B. Rand (Ed. & Trans.), *The classical psychologists: Selections illustrating psychology from Anaxagoras to Wundt* (pp. 557–561). Houghton, Mifflin and Company. https://doi.org/10.1037/10885-000

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale—Fourth Edition* [Database record]. APA PsycTests. https://doi.org/10.1037/t15169-000

Wertheimer, M. (1923/1938). Laws of organization in perceptual forms. In W. D. Ellis (Ed.), *A source book of Gestalt psychology* (pp. 71–88). Kegan Paul, Trench, Trubner & Company. https://doi.org/10.1037/11496-005

Wexler, M., Kosslyn, S. M., & Berthoz, A. (1998). Motor processes in mental rotation. *Cognition*, *68*(1), 77–94. https://doi.org/10.1016/S0010-0277(98)00032-8

Wobbrock, J. O., Findlater, L., Gergle, D., & Higgins, J. J. (2011). The aligned rank transform for nonparametric factorial analyses using only anova procedures. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 143–146. https://doi.org/10.1145/1978942.1978963

Wohlschläger, A., & Wohlschläger, A. (1998). Mental and manual rotation. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 397–412. https://doi.org/10.1037/0096-1523.24.2.397

Woods, D. L., Wyma, J. M., Yund, E. W., Herron, T. J., & Reed, B. (2015). Factors influencing the latency of simple reaction time. *Frontiers in Human Neuroscience*, *9*. https://doi.org/10.3389/fnhum.2015.00131

Yancosek, K. E., & Howell, D. (2009). A Narrative Review of Dexterity Assessments. *Journal of Hand Therapy*, *22*(3), 258–270. https://doi.org/10.1016/j.jht.2008.11.004

Yoshioka, T., & Liang, N. (2022). Mental rotation of three-dimensional object and its cubic jigsaw puzzle piece: Underlining difference in performance through a short-term learning. *The 39th Congress of the International Union of Physiological Sciences*, 729–730.

Zhang, X., Chen, C., Yang, T., & Xu, X. (2020). Spatial Skills Associated With Block-Building Complexity in Preschoolers. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.563493