Original Article

Adults' Sex Difference in a Dynamic Mental Rotation Task

Validating Infant Results

Martin Heil,[1] Markus Krüger,[2] Horst Krist,[2] Scott P. Johnson,[3] and David S. Moore[4]

[1]Institute of Experimental Psychology, Heinrich-Heine-University, Düsseldorf, Germany
[2]Institute of Psychology, Ernst-Moritz-Arndt-Universität Greifswald, Germany
[3]Department of Psychology, University of California Los Angeles, CA, USA
[4]Pitzer College and Claremont Graduate University, Claremont, CA, USA

Abstract: With the Mental Rotation Test (MRT), large and reliable sex differences are found. Used with children younger than about 9 or 10 years, MRT performance is at chance level. Simpler tasks used with younger children have revealed inconclusive results. Moore and Johnson (2008, 2011) observed sex differences in infants using a habituation task with 3D cube figures rotating back and forth in depth through a 240° angle. Thereafter, female infants treated similarly the original figure and a mirror-image cube figure presented revolving through the previously unseen 120° angle, whereas male infants behaved as if they recognized the familiar object. In the present study, 256 adults participated in the MRT as well as in a modified two-alternative forced-choice dynamic version of the infants' task. Sex differences were present for both tasks. More importantly, there was a positive correlation in performance across both tasks for both women and men. Since the new task turned out to be simpler, it might be suitable also for children. We present the first, although indirect, evidence that the sex effects reported by Moore and Johnson might indeed reflect early sex differences in mental rotation.

Keywords: sex difference, visual-spatial cognition, mental rotation, infant cognition

The cognitive process of imagining an object rotating in space is called mental rotation, first demonstrated in the seminal work of Shepard and Metzler (1971). Participants in that study were presented with pairs of perspective drawings of 3D cube figures and were asked to decide whether these two objects were identical or not. The authors found a linear relationship between response time (RT) and angular disparity, suggesting that the representation of the object was mentally rotated with a constant speed. Similar results were also obtained with much simpler 2D stimuli (e.g., Shepard & Cooper, 1982).

Vandenberg and Kuse (1978; see Peters et al., 1995, for a modern version) used these block drawings to produce the paper-pencil Mental Rotation Test (MRT), a non-chronometric task in which participants were presented 24 items, each consisting of a 3D target block figure and four choice figures. Two of these were identical to the target figure but were rotated in depth, while the other two could not be matched by rotation. In this task, too, participants are able to provide evidence of mental rotation.

Although the cause(s) are still far from being understood (e.g., Levine, Foley, Lourenco, Ehrlich, & Ratliff, 2016), it seems to be widely accepted that men outperform women on mental rotation tasks. However, a closer look at the data reveals that the empirical reality is much more complex. With the MRT, the sex difference indeed amounts to one standard deviation (see the meta-analysis of Voyer, Voyer, & Bryden, 1995). Sex differences turned out to be significantly larger when the test was administered with some time constraints compared to when such constraints were absent (Voyer, 2011). Tests with simpler stimuli like the 2D Card Rotation Test (CRT; Ekstrom, French, & Harman, 1976), for example, yield a substantially smaller effect size of $d = 0.3$, indicating the importance of the stimulus material used. Moreover, RT effects in chronometric approaches usually (e.g., Jansen-Osmann & Heil, 2007. But see Voyer et al., 2006) reveal no sex effect at all. Peters (2005), for example, preselected a sample to establish substantial and reliable sex differences in the MRT that amounted to $d = 1.5$. With these participants, no sex differences in RT or error rates were found for the chronometric,

Shepard-Metzler-like version of mental rotation with 3D cube figures used in the MRT (i.e., when pairs of perspective drawings of 3D cube figures were presented). To make things even more complicated, when the usual MRT is reduced in complexity to a paper-pencil version with pairwise presentation of the cube figures, the size of the sex effect is not affected (Titze, Heil, & Jansen, 2010). Thus, sex differences in mental rotation are unreliable when tested with protocols other than the standard MRT administered with time constraints.

With the MRT, however, the situation is relatively straightforward: The sex difference is large, reliable, probably independent of participants' cultural background (Peters, Lehmann, Takahira, Takeuchi, & Jordan, 2006), and stable over the past several decades (for meta-analyses, see e.g., Voyer, 2011; Voyer et al., 1995). Moreover, the sex difference can be found as long as the participants can complete the task, that is, it is observed with elderly people aged 60–70 years (Jansen & Heil, 2010) and also with children aged 9 or 10 years (Titze, Jansen, & Heil, 2010).

Identifying the age of the onset of the sex difference is difficult, however, although this knowledge might give clues with respect to its cause(s). Unfortunately, if children are younger than 10 years, performance in the MRT approaches chance level and as a consequence no reliable sex differences can be observed (Titze, Jansen, et al., 2010). When simpler stimuli and 2D tasks are used instead of the MRT, empirical results are ambiguous. Sometimes behavioral sex differences are observed in preschoolers that are absent in other studies (e.g., Hahn, Jansen, & Heil, 2010a, 2010b; Neuburger, Jansen, Heil, & Quaiser-Pohl, 2011). These simpler 2D tasks, however, rarely yield sex differences with adults. Thus, it is unknown whether the fact that sex differences in young children are small (or even absent) is due to participants' age or due to the type of task that was (and had to be) used.

Moore and Johnson (2008; see also Quinn & Liben, 2008, 2014, for converging evidence based upon simpler 2D stimuli) observed sex differences in 5-month-old infants using a habituation task with 3D cube figures (although conflicting results exist; for a review, see Levine et al., 2016). The infants were habituated to a cube figure repeatedly revolving in depth through a 240° angle. In the test trials, the original or a mirror-image cube figure was presented revolving through the previously unseen 120° angle, thus being a new percept in either case. Whereas female 5-month-olds treated the two test forms similarly, male 5-month-olds provided evidence of recognizing the familiar object from the new perspective by looking longer at the mirror-image test object than at the familiar object. The sex difference in this novelty preference amounted to $d = 0.66$. Next, Moore and Johnson (2011) replicated the experiment with 3-month-old infants, and again female infants in the test trials treated the familiar and the mirror-image cube figures similarly. In contrast, 3-month-old male infants provided evidence of recognizing the familiar object from the new perspective, but at this age, they looked longer at the familiar test object than at the mirror-image test object. The sex difference in the younger infants' familiarity preference amounted to $d = 0.81$.

Moore and Johnson (2011) interpreted the familiarity preferences in younger male infants in the same way that they interpreted novelty preferences in older male infants: as evidence of mental rotation. Specifically, they argued that infants who recognized an object when it was viewed from a novel perspective must have been capable of rotating a mental representation of that object. They further argued that compared to older infants, younger infants should be expected to process habituation stimuli more slowly. Therefore, familiarity rather than novelty preferences were predictable in younger infants capable of mental rotation (in addition to seeing Moore & Johnson, 2011, for a detailed explanation of their similar interpretations of novelty and familiarity preferences, see Hunter, Ames, & Koopman, 1983, for a thorough consideration of the effects of age, stimulus complexity, and familiarization times on infants' post-habituation novelty versus familiarity preferences).

Based on these exciting results, many questions emerged, from self-evident ones like whether the sex effect observed by Moore and Johnson (2008, 2011) is reliable and robust, to far-reaching ones regarding the stability of the effect, that is, whether adults' sex difference can be predicted by their looking behavior when they were infants, and the validity of the effect, that is, are male infants actually engaging in mental rotation during this task, questions that are very difficult to answer (see, e.g., Frick, Möhring, & Newcombe, 2014).

In the present paper, we addressed three simpler but related questions: if the infants' looking preference task is modified into an adults' two-alternative forced choice task, (a) do we find a sex difference in behavior and (b) is that behavior related to MRT performance? If we could provide positive evidence for both possibilities, this would validate the new dynamic mental rotation task. It would open new research possibilities regarding the stability of the effect reported by Moore and Johnson (2008, 2011) since it would reduce differences between the tasks used with infants and with adults. Finally, it would also constitute first (although rather indirect) empirical evidence that the Moore and Johnson (2008, 2011) task might have measured the cognitive process of (male) infants' mental rotation.

Methods

Participants

Altogether, 256 adults (128 women and 128 men) participated. Their age ranged from 18 to 35 years (M = 23.2, SD = 3.3). Participants younger than 36 years were recruited on campus and needed to have general qualification for university entrance to be allowed to take part in this experiment. Participants studying Psychology were not allowed to take part because of the huge number of mental rotation experiments carried out in our Department. Furthermore, self-reported right-handedness was required, because handedness is correlated with mental rotation performance (see, e.g., Somers, Shields, Boks, Kahn, & Sommer, 2015). Participants were paid €3 for their participation.

Material and Procedure

For this study we created a modified two-alternative forced choice adult version (called the 2AFC dynamic task) of the infant habituation task developed by Moore and Johnson (2008, 2011), which used the original videos (see Moore & Johnson, 2008) that presented a simplified block figure constructed of seven cubes. One figure was arbitrarily called the L-object whereas its mirror image was called the R-object. Two familiarization and two test videos were used. The former ones (length = 10.67 s) presented the L- or the R-object rotating back and forth at 45°/s around the vertical axis through a 240° arc, that is, after reaching its maximum extent of rotation, the object rotated back to its starting point. The test videos of the L- and R-object continued the rotation with the same speed through the previously unseen 120° of arc, continuously rotating back and forth until a response was given.

Each trial consisted of two videos, a familiarization video followed after a 10 s interval by a test video. Twenty trials were presented. Ten familiarization videos used the L-object (five followed by the L- and five followed by the R-test video) and 10 used the R-object, that is, in 50% of the cases, the familiarization and test videos presented the same object successively. Participants responded "same" by pressing the left mouse button and "different" by pressing the right one.

Additionally, the MRT (redrawn version of Peters et al., 1995), a paper-and-pencil test of mental rotation ability, was used. The MRT consists of two 12-item sections, each with a 3-min time limit, separated by a 3-min break. Each item consists of a row of one standard cube figure and four comparison cubes. Two comparison cube figures are correct matches rotated in depth; the remaining two are incorrect matches.

The individual test sessions, which lasted about 30 min, took place in a laboratory at the Heinrich-Heine-University of Düsseldorf. Half of the participants (with an equal number of women and men) worked on the MRT first, the other half started with the 2AFC dynamic task. Since sequence of task had no effect (either as a main effect or in an interaction term), results are presented collapsed across this factor.

Statistical Analysis

The design of the study involved "sex" (male, female) as the independent variable and "2AFC dynamic task performance" and "MRT performance," respectively, as dependent variables. The dependent variable for the 2AFC dynamic task was the number of "same" responses to same trials minus the number of "same" responses to different trials. Thus, a score of "10" reflects perfect performance whereas guessing at the chance level is reflected by a score of "0." The dependent variable for the MRT was the number of correct answers. An answer was correct when both correct figures were correctly selected. No point was given if only one line drawing was selected correctly. Thus, a score of "24" reflects perfect performance whereas guessing at the chance level is reflected by a score of "6." Given a total sample size of N = 256 and a desired alpha level of α = .05 (one-tailed), effects of size d = 0.2, that is, even small sex effects as defined by Cohen (1977) could be detected with a probability of $1 - \beta$ = .95. The power calculation reported was conducted using the G*Power program (Faul, Erdfelder, Lang, & Buchner, 2007).

Results

Mental Rotation Test performance replicated the well-known sex effect: Men (M = 12.30, SD = 4.63) outperformed women (M = 8.52, SD = 3.75), $F(1, 254) = 51.50$, $p < .001$. The effect size was large, d = 0.90 (95% CI = 0.64 < d < 1.15). Forty-three women (i.e., 33.6%) scored 6 points or lower (i.e., at chance level) in the MRT with no woman reaching the maximum score of 24 points. Fourteen men (i.e., 10.9%) scored 6 points or lower (i.e., at chance level) in the MRT with two men (i.e., 1.6%) reaching the maximum score of 24 points.

More importantly, in the modified two-alternative forced choice (2AFC) dynamic adult version of the infant habituation task, men (M = 7.94, SD = 2.22) also outperformed women (M = 6.75, SD = 2.95), $F(1, 254) = 13.21$, $p < .001$. The effect size was medium, d = 0.46 (95% CI = 0.21 < d < 0.70). Seven women (i.e., 4.7%) scored 0 points or

lower (i.e., at chance level) in this task with 20 women (15.6%) reaching the maximum score of 10 points. One man (i.e., 0.8%) scored 0 points or lower (i.e., at chance level) in this task with 34 men (i.e., 26.6%) reaching the maximum score of 10 points.

Finally, MRT performance was significantly correlated with 2AFC-adult performance in the full sample, $r = .40$, and also when calculated separately both for men ($r = .34$) and for women ($r = .38$), all ps $< .001$. Based on a gratefully acknowledged suggestion by an anonymous reviewer, we also analyzed the data in an analysis of variance (ANOVA) based on a 2 (MRT vs. 2AFC test) by 2 (sex) by 2 (order of presentation) design after applying a z standardization separately for the two tests. In addition to a main effect of sex, $F(1, 252) = 42.02$, $p < .01$, an interaction of sex by test, $F(1, 252) = 7.79$, $p < .01$, reflected the larger sex effect size for the MRT in comparison to the 2AFC. No other main effect or interaction turned out to be significant.

Discussion

Adults' sex difference in the MRT has proven to be reliable, stable, and substantial with an effect size of about $d = 1.0$ when the test was administered with time constraints. With children, reliable and substantial sex differences based on the MRT have been observed when performance was above chance level, that is, when children were at least 9 or 10 years of age. With younger children, simpler tests have usually revealed ambiguous results that could either be attributed to the simpler tests or to reliable sex differences only emerging at around the age of 10 years (see e.g., Titze, Jansen, et al, 2010). Interestingly, Moore and Johnson (2008, 2011) reported sex differences in infants using a habituation-dishabituation task with dynamic 3D block figures. Obviously, many questions emerged from that (Frick et al., 2014), and the goals of the present study were to find out whether a sex difference in adults' behavior would be found with the modified 2AFC dynamic task, and whether that behavior would be related to MRT performance.

The present study replicated the large effect sizes (see Cohen, 1977) of about $d = 0.90$ previously reported in the MRT, indicating a fundamental sex difference comparable to the effect sizes reported in the meta-analysis by Voyer and colleagues (1995). A medium-sized sex effect was observed for our 2AFC adult version of the infant task, which was slightly smaller than the effect sizes reported by Moore and Johnson (2008, 2011). This might be due to the relatively large number of (male) adult participants with perfect performance.

Moreover, performance in the new task was correlated moderately with MRT performance across the entire sample and for women and men when analyzed separately. One might speculate that the correlation would have been even higher with no ceiling or floor effects in the 2AFC task or in the MRT, respectively. The new task obviously turned out to be substantially easier than the MRT and might even be slightly too easy for adults, probably due to the fact that the rotation is dynamically visible in the new task but has to be completely imagined in the MRT. Thus, it might be that the new task could be well suited even for children younger than 10 years, for whom the MRT is definitely too difficult to be used. Thus, empirical data are needed for the new task with (elementary) school children given that the new task might be suitable for a wider age range than the MRT.

Our data are consistent with the hypothesis that the modified 2AFC dynamic adult version of the infant looking time task indeed involves the cognitive process of mental rotation. That does, of course, not constitute direct evidence that infants indeed used the cognitive process of mental rotation in the Moore and Johnson (2008, 2011) task; instead, the evidence for the validity of the infants' results remains indirect. Nevertheless, the present data constitute the first indirect evidence that (male) infants might use mental rotation when choosing to preferentially fixate a mirror-image test object or a familiar test object seen from a novel perspective. More research is needed to verify this possibility, but our results suggest that it will be worth the effort.

Acknowledgments

We thank Michael Peters for his friendly permission to use the Mental Rotations Test (MRT) in our study. We gratefully acknowledge the helpful suggestions of two anonymous reviewers.

References

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Ekstrom, R., French, J. W., & Harman, H. (1976). *Manual for kit of factor-referenced cognitive tests*. Princeton, NJ: Educational Testing Service.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175–191. https://doi.org/10.3758/BF03193146

Frick, A., Möhring, W., & Newcombe, N. S. (2014). Development of mental transformation abilities. *Trends in Cognitive Sciences, 18*, 536–542. https://doi.org/10.1016/j.tics.2014.05.011

Hahn, N., Jansen, P., & Heil, M. (2010a). Preschoolers' mental rotation: Sex differences in hemispheric asymmetry. *Journal of Cognitive Neuroscience, 22*, 1244–1250. https://doi.org/10.1162/jocn.2009.21236

Hahn, N., Jansen, P., & Heil, M. (2010b). Preschoolers' mental rotation of letters: Sex differences in hemispheric asymmetry. *Cognitive Neuroscience, 1*, 261–267. https://doi.org/10.1080/17588928.2010.485248

Hunter, M. A., Ames, E. W., & Koopman, R. (1983). Effects of stimulus complexity and familiarization time on infant preferences for novel and familiar stimuli. *Developmental Psychology, 19*, 338–352. https://doi.org/10.1037/0012-1649.19.3.338

Jansen, P., & Heil, M. (2010). Gender differences in mental rotation across adulthood. *Experimental Aging Research, 36*, 94–104. https://doi.org/10.1080/03610730903422762

Jansen-Osmann, P., & Heil, M. (2007). Suitable stimuli to obtain (no) gender differences in the speed of cognitive processes involved in mental rotation. *Brain and Cognition, 64*, 217–227. https://doi.org/10.1016/j.bandc.2007.03.002

Levine, S. C., Foley, A., Lourenco, S., Ehrlich, S., & Ratliff, K. (2016). Sex differences in spatial cognition: Advancing the conversation. *Wiley Interdisciplinary Reviews – Cognitive Science, 7*, 127–155. https://doi.org/10.1002/wcs.1380

Moore, D. S., & Johnson, S. P. (2008). Mental rotation in human infants: A sex difference. *Psychological Science, 19*, 1063–1066. https://doi.org/10.1111/j.1467-9280.2008.02200.x

Moore, D. S., & Johnson, S. P. (2011). Mental rotation of dynamic, three-dimensional stimuli by 3-month-old infants. *Infancy, 16*, 435–445. https://doi.org/10.1111/j.1532-7078.2010.00058.x

Neuburger, S., Jansen, P., Heil, M., & Quaiser-Pohl, C. (2011). Gender differences in pre-adolescent's mental rotation performance: Do they depend on grade und stimuli? *Personality and Individual Differences, 50*, 1238–1242. https://doi.org/10.1016/j.paid.2011.02.017

Peters, M. (2005). Sex differences and the factor of time in solving Vandenberg and Kuse mental rotation problems. *Brain and Cognition, 57*, 176–184. https://doi.org/10.1016/j.bandc.2004.08.052

Peters, M., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R., & Richardson, C. (1995). A redrawn Vandenberg and Kuse mental rotations test: Different versions and factors that affect performance. *Brain and Cognition, 28*, 39–58. https://doi.org/10.1006/brcg.1995.1032

Peters, M., Lehmann, W., Takahira, S., Takeuchi, Y., & Jordan, K. (2006). Mental rotation test performance in four crosscultural samples (*n* = 3367): Overall sex differences and the role of academic program in performance. *Cortex, 42*, 1005–1014. https://doi.org/10.1016/S0010-9452(08)70206-5

Quinn, P. C., & Liben, L. S. (2008). A sex difference in mental rotation in young infants. *Psychological Science, 19*, 1067–1070. https://doi.org/10.1111/j.1467-9280.2008.02201.x

Quinn, P. C., & Liben, L. S. (2014). A sex difference in mental rotation in infants: Convergent evidence. *Infancy, 19*, 103–116. https://doi.org/10.1111/infa.12033

Shepard, R. N., & Cooper, L. A. (1982). *Mental images and their transformations*. Cambridge, MA: MIT Press.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171*, 701–703. https://doi.org/10.1126/science.171.3972.701

Somers, M., Shields, L. S., Boks, M. P., Kahn, R. S., & Sommer, I. E. (2015). Cognitive benefits of right-handedness: A meta-analysis. *Neuroscience and Biobehavioral Reviews, 51*, 48–63. https://doi.org/10.1016/j.neubiorev.2015.01.003

Titze, C., Heil, M., & Jansen, P. (2010). Pairwise presentation of cube figures does not reduce gender differences in mental rotation performance. *Journal of Individual Differences, 31*, 101–105. https://doi.org/10.1027/1614-0001/a000018

Titze, C., Jansen, P., & Heil, M. (2010). Mental rotations performance and the effect of gender in fourth graders and adults. *European Journal of Developmental Psychology, 7*, 432–444. https://doi.org/10.1080/17405620802548214

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills, 47*, 599–604. https://doi.org/10.2466/pms.1978.47.2.599

Voyer, D. (2011). Time limits and gender differences on paper-and-pencil tests of mental rotation: A meta-analysis. *Psychonomic Bulletin & Review, 18*, 267–277. https://doi.org/10.3758/s13423-010-0042-0

Voyer, D., Butler, T., Cordero, J., Brake, B., Silbersweig, D., Stern, E., & Imperato-McGinley, J. (2006). The relation between computerized and paper-and-pencil mental rotation tasks: A validation study. *Journal of Clinical and Experimental Neuropsychology, 28*, 928–939. https://doi.org/10.1080/13803390591004310

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin, 117*, 250–270. https://doi.org/10.1037/0033-2909.117.2.250

Received February 28, 2017
Revision received May 29, 2017
Accepted June 15, 2017
Published online January 12, 2018

Martin Heil
Institute of Experimental Psychology
Heinrich-Heine-University
Universitätsstr. 1
40225 Düsseldorf
Germany
martin.heil@hhu.de

