

Spatial Tasks on a Large, High-Resolution, Tiled Display: A Male Inferiority in Performance with a Mental Rotation Task

Bernt Ivar Olsen¹, Bruno Laeng^{2,3}, Kari-Ann Kristiansen⁴,
and Gunnar Hartvigsen^{1,4}

¹ Department of Computer Science, University of Tromsø, 9037 Tromsø, Norway

² Department of Psychology, University of Oslo, Oslo, Norway

³ Department of Biological & Medical Psychology, University of Bergen, Bergen, Norway

⁴ University Hospital of Northern Norway

{bernt-ivar.olsen,gunnar.hartvigsen}@uit.no,
bruno.laeng@psykologi.uio.no, kari-ann.kristiansen@unn.no

Abstract. In previous research we have investigated the effect of screen size on the perceptual mental rotation task (MRT) by comparing performance on a large 230 inches display with that on a standard 14.1 inches laptop display. The former work indicated that females might gain an advantage over males on a larger display. The current study confirms a significant female advantage over male performance in the MRT. However, our current findings helped to reveal that, instead of improving the females' performance, the screen size had a detrimental effect on male performance, while female performance actually remained unaffected by both the large object size than the standard one.

Keywords: Tiled display, Spatial Tasks, Mental Rotation, Sex differences.

1 Introduction

The main motivation behind the present project is to understand the effects on new users of a very large, high-resolution display called a Display Wall [1] in the medical domain of Radiology [2]. The present study extends our previous work where we tested effects of display size as well as effects of *expectations* among the participants [3]. Specifically, we have investigated the effect of screen size on a well-known cognitive task, called “mental rotation”, by comparing performance on a large 230 inches display with that on a standard 14.1 inches laptop display. In a previous study, we found that, unexpectedly, women had faster response times (RTs) on the large display than men did. However, the main aim of this previous study was to assess the effect of introducing a novel technology artifact, such as the Display Wall and how the “novelty” of such a situation, compared to the more traditional experience with the much smaller screens of a laptop computer, could affect performance. Hence, in that study, we specifically induced expectations by informing half of the subjects, at the outset of the experiment, that either the Display Wall would yield superior performance or, for the other half of the subjects, that we expected worse performance with the new Display Wall than with the traditional display. We found that females

who were told to expect large screen superiority did significantly outperform all other groups in the Display Wall condition. Although suggestive, our previous study was, first of all, based on a rather small sample size and, secondly, it remained unclear whether women would have outperformed men also if they were not given in advance any specific positive expectation. The present study specifically attempted to answer the last question.

1.1 A Couple of Words about the Findings in Our Previous Study

Before we proceed, in order to discuss the present study to full detail, we would like to recap and shed some more light on the results on our first study [3]. The details are available in the previous work, but for ease of reading and understanding we recap some of the results here. The former study had its focus on the effect of introducing a novel technological artifact to subjects who had no prior experience with such equipment. Although we found effects of very large objects deteriorating male performance to the point of a 603ms female advantage in the large screen, there were a couple of concerns that we wanted to address in a new study – in order to “disentangle” the finding of a male disadvantage in mental rotation task with very large objects with that of expectations regarding novel technology.

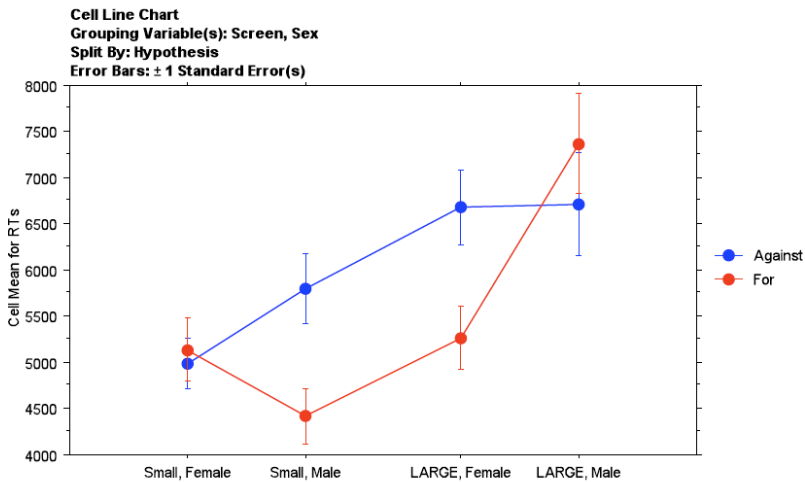


Fig. 1. Mean response times from males/females and large and small display sizes, split by expectation of which condition to perform better in

A sample size of N=40, where participants were effectively distributed among 4 groups (Men with positive/negative “induced expectations” towards the novel display technology and equal setup for females), proved to be too small to yield the expected effects. In Figure 1 we illustrate the most relevant results of that study. The mean response times for each of the groups of participants and the respective experiment

conditions revealed that expectation had an influence on male performance only in the small display condition. From these results it seems that males expecting to do better in the large display condition performed better (“for”, “Small, male”) than those believing that the small display was the superior condition. In the Large display condition, females who were expecting the larger display to be the superior condition *did* perform better than females expecting the smaller display to yield superior performance, while males’ performance yielded the opposite result; males expecting better performance in the large display performed slower than those believe the small display to be better. It seems that males performed contrary to “expectations”, while females performed according to “expectations” – at least in the large display condition (no difference in performance between the two groups in the small display condition). In the previous study, hence, “Expectations” (negative/positive) were “forced” upon subjects, while in the current study observations reflect either absence of expectations or what they naturally expect from the novel situation.

Secondly, there was an induced delay in the onset of stimulus in the large display condition, created by network- and processing overhead transferring the screen image from the laptop computer to the computer cluster feeding the projector array. We were not able to measure this delay exactly, which made analysis of within-subjects factors unreliable. This included comparing performance within-subjects on the large-display stimulus to small display-stimulus. To address the issue of delay in the large-display stimulus we have used high-speed camera to measure the mean delay in order to subtract this from the timings of subjects’ performance in the large display condition. Hence, we were able to, this time around, accurately describe how participants generally performed in the mental rotation task on a large display compared to performance on the small display.

2 Method and Experimental Setup

Participants. Thirty-six men and 32 women participated in the study (age range: 18 - 51 years; mean age = 24.0, SD 5.44). All participants participated voluntarily and they were offered two lottery tickets for their participation (this appreciation of their time and effort was introduced after the experiment and, hence, was not used to recruit participants). Nine participants were excluded; four women and five men because they failed to reach a criterion accuracy score of 70% correct in the task; mean accuracy score= 59.6 % and 57.9%, respectively. Two more participants were excluded due to technical failure. The descriptions and analyses shown below consist of responses of the remaining 57 participants. Those who did not understand Norwegian received all the relevant information, instruction and the questionnaire in English. All participants had normal, or corrected-to-normal, eyesight. Participants were recruited from the natural sciences or psychology study programs.

Stimuli and Apparatus. We have used a computerized mental rotation task devised by Peters and Battista [4], which was modeled after the classic mental rotation task introduced by Shepard and Metzler [5]. That is, figures composed of block or cubes are presented as two-dimensional visual images that are constructed of several of such cubes that can be perceived as 3-D figures. As in the Shepard and Metzler paradigm, in the present task participants were shown, in each trial, a pair of cube stimuli where

one figure either matched with respect to the other (by applying, mentally, a rotation to the figures) or did not (i.e., the other figure was a mirror image of the other, so that no amount of rotation could make the figures identical). Thus, the participants' task simply consists in deciding, as quickly but as accurately as possible, whether the two figures are the 'same' or 'different'.

The display device used in this experiment was what has become to be known a "display wall", which consists of many projectors that back-project a seemingly coherent picture across the canvas (see Figure 2). Display wall technology is the result of a kind of natural development of the display from a small, single display to today's trend of larger displays and multiple-monitor configurations. With a display wall, testing for the effect of display size is

equivalent to testing how the size of an object can affect the task of mental rotation. The present large 'display wall' consists of 28 projectors, back-projecting an image onto a screen surface. There are 7×1024 horizontal pixels and 4×768 vertical pixels, seven by four tiles, on the large screen, a total of approximately 22 million pixels. The physical visible screen size of the large screen



Fig. 2. Mental Rotation stimulus examples on the Display Wall

is 230 inches. Within the color spectrum there are 22 million pixels of red, green and blue. The small screen setup featured a 14.1 inches screen on a laptop computer, a Dell D600 with a native resolution of 1400×1050 pixels and a 24-bit color spectrum.

The mental rotation stimuli were taken from the large stimulus library provided by Peters and Battista [4], which uses wire frame stimuli of the kind introduced by Shepard and Metzler [5]. Stimuli were presented by use of SuperLab® software that was running on the laptop computer, while the image was transferred to the display wall using a 100MB Ethernet interface and a Java implemented display-server running on the Virtual Network Computing (VNC) server on a Dell PowerEdge 2800, with 2 Xeon 3.8GHz/2MB 800FSB, 8GB Dual Rank DDR2 Memory (4x2GB), 146GB SCSI Ultra320 (15,000rpm) 1in 80 pin Hard Drive x 2 with the RedHat Linux operating system. The computer cluster feeding the projectors is comprised of 28+1 Dell 370 PCs with P4 Prescott, 3.2GHz, 2GB RAM, 1Gbit Ethernet and a 48 port HP switch. The SuperLab interface (stimuli) was transferred to the display and enlarged to fit the larger display area of the wall. As a consequence, the number of (perceived) pixels was held constant between the displays, along with the aspect-ratio (4:3). As

for the screen width and consequential retinal size of the images projected (visual angle of screen), the projected screen (display area covered by Superlab) was measured using a laser-meter to 404cm and 28,5cm for the small screen.

Viewing distance was 370 cm from the large screen, and with the small screen ca. 65 cm. With the laptop computer the participants were instructed to keep a “comfortable viewing distance”. Subsequent to testing, a random sample of the participants was instructed to demonstrate how they solved the task, and measurements were done of the viewing distance. This was recorded and a mean distance of 65 cm was computed. This setup constituted a total visual angle of 57 degrees and 24.7 degrees in the large and small display setup, respectively. This, in turn, constituted between-objects angles of 27.3 and 11.8 degrees, respectively. Total visual angle means the visual angle provided by the display in question, while angle between objects refers to the approximate angle from the person to the midpoints of the objects. During the phase with the large screen, participants used a wireless keyboard in order not to be disturbed by the small screen as they might have been had they used the laptop-keyboard for this task. In the trials with the small screen they used the laptop’s keyboard. The “.” And “Z” keys, which corresponded to a same or different response, were marked with either a green label (for “same”) or a red label (for “different”).

SuperLab©, version 2.02, not only presented all stimuli in a completely randomized order but also kept a record of the key presses and their time occurrence from the onset of the picture. From the Peters and Battista’s stimulus set [4], 19 pairs of images were selected. The second stimulus could differ from the first by 30, 60, 90 120 or 150 degrees of rotation around the vertical axis, for a total 95 pairs. In addition, the second stimulus was either identical to, or a mirror image of the first, for a total set of 190 stimulus pairs. The number of angles was reduced from seven to five from the first experiment to the current. In the de-briefing phase of the previous experiment we had many participants complain that the task was strenuous due to the many trials. This was a consequence of within-subjects design (participants had to complete the experiment in both display size conditions). Excluded angles from the previous experiment were 0 and 180.

In the Large Display condition, the stimuli were transferred from the laptop computer via a 100 Mbit LAN interface (TCP/IP) to a virtual network computer (VNC) server, which would scale the stimuli up to the correct size in order to fill the entire Display Wall canvas. In our previous work [3], we were not able to measure this delay exactly, and given that the variable delay from the previous experiment made analysis within-subjects regarding display size difficult, we attempted to exactly measure the delay in the present study. This was achieved using a high-speed camera, where we videotaped one complete experiment with both displays simultaneously on camera, counted the number of frames that would pass between update of the small screen and subsequent update of the Display Wall screen, recorded the delay between update of the small display and the large display and finally computed the mean delay from a total of 182 samples (trials that were videotaped). What we did in order to measure the delay was to place the laptop display visible in the camera-view so that both displays would be videotaped with 300 frames per second using a Casio Exilim EX-F1 camera for this purpose. Note that during experiments, the laptop computer that was running the experiment software would be set physically aside with the lid

closed so that it would not be visible for the participants (i.e., they would only be experiencing stimuli from the large display – even though the laptop computer would still be producing output on the hidden laptop display).

Procedure. The experiment took place in a room containing a “Display Wall” at the Department of Computer Science, University of Tromsø, Norway. Temperature was set at 20° C and light setting to dark. All participants were tested in the same room with the same equipment. Each participant was pseudo-randomly assigned to groups that began testing with the small screen versus the large conditions. Pseudo-random assignment consisted in alternating the conditions to balance the set with sex and first trial-run condition (large or small screen). This was done to counter-balance for the practice effect in within-subjects design [6]. The participants were given 4 training samples before the start of the experiment to ensure that the participant understood the task. The first two training trials included feedback whether the objects were similar/not similar. The task itself was self-paced and each object remained on the screen until the participant made a decision by pressing one of the two keys “.” or “Z” to indicate that the shapes were either the same or different. The computer recorded the result for each key press by use of SuperLab© 2.0 software. There were a total of 190 trials for each subject in each of the two screen conditions.

Both small-screen and large-screen conditions took place in the same room, each participant sat at the same table, in the same position, in order to try to keep the environmental variables constant. When both conditions were completed, the participants were given the questionnaire that collected some biographical information, like Sex, age, years of education and type of education. We also included a question to try and record subjective expectations towards which display-condition the participants felt they felt produced better results with regards to speed *and* accuracy. This question was added to the questionnaire in this experiment in order to have something to compare the Expectation variable from the previous experiment with. This time, however, the “Expectation” would be a subjective feeling, rather than something externally produced (by us; the experimenters) *and* recorded *after* the experiments (both large and small screen conditions finished).

Design and statistics. We used a mixed design where Sex (female/male) was the between-subjects factors and Screen (large/small) and Angle (30°, 60°, 90°, 120°, 150°) were within-subject factors. An additional between-subjects factor was Order (large first/small first). Data were analyzed using Statview® (5.0) and SPSS® (v.16).

3 Results

We calculated descriptive statistics for each participant to obtain mean RTs for correct responses and mean % accuracy for each for each combination of variables (Screen Size, Match, Angle). The analyses did not reveal any main effects or interactive effects for Accuracy as factor in the second experiment either. We did have to remove 9 subjects (4 females, 5 males) from the analysis since they scored below a 70% cut-off. Excluded subjects typically had chance performance in one or both screen size conditions. The Large Screen delay was measured on average to be 848ms, which – in the analysis below has been subtracted from the mean Large

Display RTs for all participants. We found a main effect of Screen Size, $F(1,55)=6.69$, $p=.012$, and an interactive effect of Sex with Screen Size, $F(1,55) = 6.51$, $p=.014$. The linear relation between angular disparity and RTs were again reproduced ($P<0.00$, $F(1, 4)=77.14$). Descriptive statistics confirmed that females were on average 1290ms faster (males: 5347ms; females: 4057ms) than males in the large display condition, while there was no significant difference between the sexes in the small display condition (m: 4208ms; f: 4050ms).

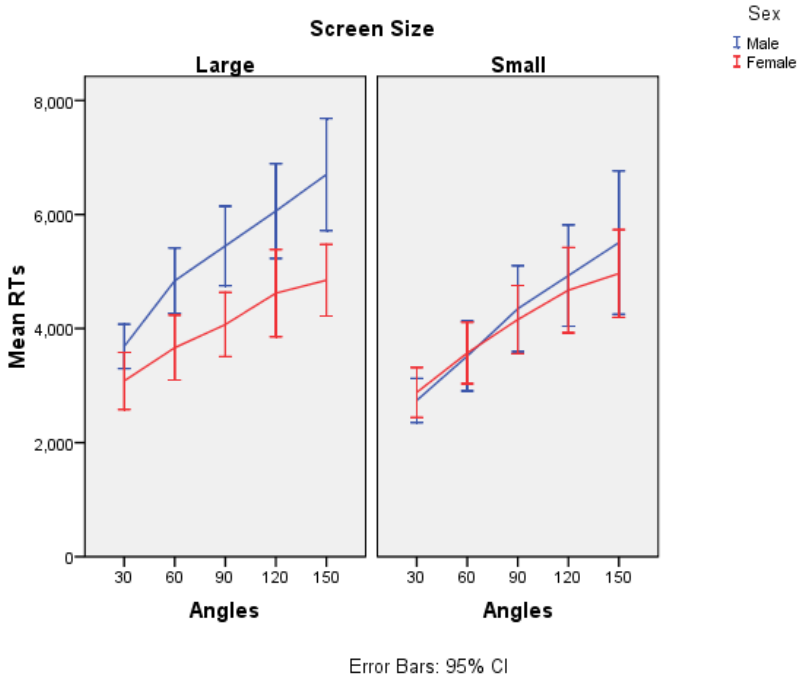


Fig. 3. Mean RTs for the two display conditions for all angular conditions, split by sex. Vertical lines represent 95% confidence intervals for the respective means.

In Figure 3, we observe how the plots of mean RTs for the males and females differ in the large display condition, but do not significantly differ for the normal sized stimulus. The red line in the left half of Figure three (females) follow a curve well below the 95% confidence interval-lines of the blue line (male observed mean RTs with corresponding CIs), while in the regular display (Small) there is no such effect.

As for the Expectation factor, this time around there was no observed interactional-effect of Expectation with Sex in this experiment, as neither Sex ($P>0.24$) nor Sex * Display Size ($P>0.19$) showed any significant interaction with Expectation as factor. Even if there was no apparent difference between the genders in their “expectation” towards which screen size they performed better with – in order to investigate

participants' expectations towards the displays further, we plotted the mean RTs for each group for all angular conditions of degree of rotation (again: participants responded to whether they thought their performance was better on the large or the small display), presented in Figure 4. What we can see from Figure 4 is that the participants that expected their results to be better on the large display (red line) performed slower than the group that expected their results to be better on the small display.

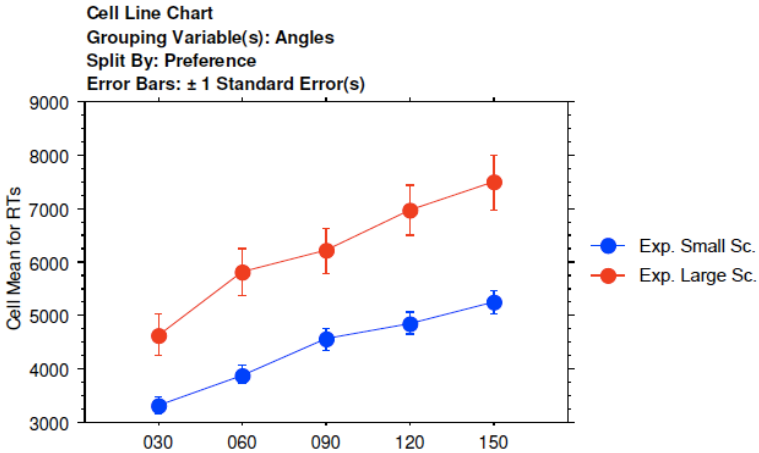


Fig. 4. Experiment 2: mean RTs for different angular conditions, split by which display condition the participants believed their performance was best on

The points on X-axis represent screen size; while points represent observed mean RTs for the different groups with bars representing ± 1 SE. We also observe that the positive linear trend for increasing angular rotations is apparent from Figure 4.

4 Discussion

We have confirmed in this replication of our previous study [3] that females outperform males with respect to task efficacy (speed) in the task of mental rotation on large objects (i.e., large displays that cover about 50 degrees of our visual field). Moreover, this effect does not seem to imply females performing faster on larger displays, but rather that the larger (than normal) objects have a detrimental effect on male performance, so as to produce this unusual effect of female superiority on the task of mental rotation.

In our previous work [3] we did observe a female advantage with identical large stimuli as in the current experiment. However, we could not reach a conclusion, based on those data, about whether this female advantage was due to the objects's size or if

it was due to expectations alone – or a combination of these factors. From Figure 1 of the present study, it seems that those females who spontaneously expecting to do better with a large display did disproportionately better than those expecting to do worse with a large display than with a normal sized laptop computer display. The current results lead us to believe that Expectation in the previous work had in fact *little or nothing* to do with the “improved” performance in that experiment for females relative to males in the large display condition. In fact, there is no actual “improvement” to be gained from perceiving large objects in the mental rotation task, for either females or males – with respect to either speed or accuracy. What we have observed in this experiment is a significant detrimental effect of large objects on male performance on the mental rotation task, while female performance on average remains virtually unaffected by the larger object size.

Acknowledgements. We would like to thank the Display Wall project and professor Otto Anshus in particular for the use of the equipment and their lab. Our gratitude is also extended to the technical personnel at the Department of Computer Science, University of Tromsø.

References

1. Wallace, G., Anshus, O.J., Bi, P., Chen, H., Chen, Y., Clark, D., Cook, P., Finkelstein, A., Funkhouser, T., Gupta, A., Hibbs, M., Li, K., Liu, Z., Samanta, R., Sukthakar, R., Troyanskaya, O.: Tools and Applications for Large-Scale Display Walls. *IEEE Computer Graphics and Applications* 25, 24–33 (2005)
2. Olsen, B.I., Dhakal, S.B., Eldevik, O.P., Hasvold, P., Hartvigsen, G.: A large, high resolution tiled display for medical use: experiences from prototyping of a radiology scenario. *Studies in health technology and informatics* 136, 535–540 (2008)
3. Olsen, B.I., Laeng, B., Kristiansen, K.-A., Hartvigsen, G.: Spatial Tasks on a Large, High-Resolution Tiled Display: Females Mentally Rotate Large Objects Faster Than Men. In: Harris, D. (ed.) *EPCE 2009. LNCS*, vol. 5639, pp. 233–242. Springer, Heidelberg (2009)
4. Peters, M., Battista, C.: Applications of mental rotation figures of the Shepard and Metzler type and description of a mental rotation stimulus library. *Brain and Cognition* 66, 260–264 (2008)
5. Shepard, R.N., Metzler, J.: Mental rotation of three-dimensional objects. *Science* 171, 701–703 (1971)
6. Peters, M.M., Laeng, B.B., Latham, K.K., Jackson, M.M., Zaiyouna, R.R., Richardson, C.C.: A redrawn Vandenberg and Kuse mental rotations test: different versions and factors that affect performance. *Brain and Cognition* 28, 39–58 (1995)