



**Expectations and Mental Rotation:**

**Different effects on gender explored through the (unlikely) case of  
female superiority in mental rotation rate of large objects**

Kari-Ann Kristiansen

Master thesis for the Cand. Psychol. Degree

Spring 2010

Psy - 2901

Main supervisor:

Professor Bruno Laeng

Department of Psychology, Faculty of Social Sciences

University of Oslo

Co-supervisor:

Associate Professor Torstein Låg

Department of Psychology, Faculty of Health Sciences

Department of Psychology

University of Tromsø

N-9037 Tromsø

## PREFACE

This project is a case of serendipity; it started with an idea emerged under a lecture given by Professor Bruno Laeng in Tromsø spring semester 2007. The topic was vision and cognition. After the lecture I approached him and told him about a project concerning a very large, tiled display at the Department of Computer Science. As it happened, I had been approached some time before this by research scholar Bernt I. Olsen, whose project concerned potential usages of this novel technology. In this context, I became familiar with the recent literature where technologists had used psychometric methods in controlled experiments to test for effects of display size for different usages of novel, large and/or exceptionally-wide displays. Professor Laeng and Olsen subsequently met and decided to cooperate in order to try and extend the research concerning display size. Professor Laeng suggested that mental rotation (MR) could be investigated in this context, since (1) previous research concerning MR had not taken into account increased field of view in this context and that (2) research design could well be improved in a novel experiment. Laeng also suggested that we should test for expectations among the participants, especially considering the “novelty factor” for this technology.

Except for the initial “facilitator” role, my contribution to this project has been to collect all data from both experiments by personally testing all participants. I have also contributed to develop the questionnaires and been mainly responsible for recording the qualitative data (for both experiments). As the experiment was performed in a computer lab, normally inhabited by technicians and computer science researchers, my work has also been to attempt to impose psychological research-standards for this environment when it was “transformed” into a venue for behavioral experiments. This work has

consisted of both trying to control the physical appearance of the room (keeping it as consistent as possible throughout the experimental period (which endured for several months in both experiments), e.g. lighting conditions, distance to the display, etc. – as well as ensuring that participants were exactly and consistently informed in all experimental conditions. I have also been closely involved in the analysis of the data, punching in the data, preparing the data files for the statistical analyses as well as performing some of the analysis and drawing the figures/plots, in particular those regarding Experiment 2.

I would like to thank Professor Bruno Laeng, first of all for his receptivity to the project and for his excellent choice of methodology and design for both experiments, but most of all for his kind and experienced advise and guidance along the way. I also want to acknowledge the immense support from the Department of Computer Science and their staff, especially Ken-Arne Jensen, Kai-Even Nilsen, John-Markus Bjørndalen, Otto Anshus, Daniel Stødle, Tor-Magne Stien Hagen, Espen Johnsen, Jon Ivar Kristiansen, Maria Hauglann, and Gunnar Hartvigsen. I thank my co-advisor Torstein Låg, whom gave me the idea to dwell on the delay in the large display condition. Finally I thank my husband for his love and support, and for all his input and our discussions.

## ABSTRACT

The present study aims to examine the relationship between expectation about performance in a mental rotation task of 3-D geometrical shapes on a small (standard) and large (novel) display and for each sex. In Experiment 1 we explicitly induced expectations about whether participants would perform better on either the large or small displays. Results revealed that women mentally rotate the 3-D shapes faster than men on the large display, but both genders were affected by the expectation variable. Apparently, only females expecting to do better in the large display condition experienced superior performance relative to men. Males appeared affected by expectations toward task difficulty (superior conditions), where they scored lower when expecting a superior condition. Experiment 2 confirmed the effect of display size on gender and female superiority was again observed in the large display condition – however without any explicit manipulation in the experimental design of expectations (for superiority of a display condition over the other).

*Expectations and Mental Rotation:*

*Different effects on gender explored through the (unlikely) case of female superiority in mental rotation rate of large objects*

Introduction

Spatial ability would seem essential in order to interact with objects in our daily environment, not only to locate them, manipulate them, and navigate around them and within the space containing them, but also for several technical and specialized activities like designing and building a house (e.g., architecture design and engineering), medical procedures and analysis of radiology images (e.g., brain scans used by neurologists). Moreover, spatial ability may be important for high-level cognition, since there has been observed a high correlation between spatial ability and the ability to solve problems in geometry, mathematics, chemistry, and physics (Delgado & Prieto, 2004).

Spatial ability could be measured in an infinity of ways, but within psychology, the Mental Rotation Tests (MRT) have been established as a relatively robust and easy-to-collect measure of at least one important component of spatial ability: The ability to mentally hold the image of a three-dimensional figure and rotate it in space so that one can decide whether it is identical to another object (or previous event). This test has shown a consistent behavioral pattern consisting in a strongly linear increase in response time and error rates as the angular disparity between objects in the stimulus pair increases. In other words, the efficacy of the mental rotation process can be operationalized by the slope of the linear function described above (Shepard & Metzler, 1971). Shepard and Metzler measured mental rotation in two planes; “picture plane” and

in “depth” (corresponding to “X-Y” plane and “Z-Y” plane) and found no difference in regards to response times and linear slopes. The original motivation for their study was to reveal a human “trait” that is able to identify two identical three-dimensional objects from two-dimensional drawings (representations), despite the fact that they may be seen in very different orientations. Hence, in order to be able to “mentally rotate” the objects, an internal representation would have to be created from a two-dimensional image before the process of mental rotation could be executed. The finding that objects that were rotated in “depth” took no longer to rotate than those that were rotated in the picture plane was interpreted as support to the claim that a three-dimensional representation of the two-dimensional portrayed image was created by the participants because “they could imagine the rotation around whichever axis was required with equal ease” (ibid.).

Remarkably, since its discovery as a test and mental ability, the mental rotation is the spatial trait that has shown the largest and most consistent sex difference in human subjects, favoring male performance in speed and accuracy compared to what is observed in other spatial tasks (Hirnstein, Bayer, & Hausmann, 2009; Linn & Petersen, 1985; Daniel Voyer, Voyer, & Bryden, 1995).

In the attempt to explain such a gender difference in mental rotation tasks, several accounts have been put forward. One is the “biological” account, which actually comprises a family of accounts that can more or less stress the existence of evolutionarily-selected traits. For example, it has been suggested that the difference originates already in the mother’s womb, where the fetus is exposed to specific hormones, and afterwards post-natally to the exposure of the fetus’ sex hormones, all of which we know affect brain development (Kerns & Berenbaum, 1991). Also, the time-

lines of sexual maturation (which differ in the two sexes) would seem to be crucial for performance in MRT; specifically, those who mature later than the average will score higher than those who mature earlier (Sanders & Soares, 1986).

Differences in brain volumes, within target brain regions of each sex, have also been observed and there is evidence favoring female in grey matter volume in the parietal lobe, which was shown to be disadvantageous for women regarding performance on mental rotation task. The parietal lobes are considered to play an important role in spatial processing, especially for mental rotation. Men show predominantly parietal activation whereas females show additional inferior frontal activation, and their activation in the parietal lobe seems to be more symmetrically organized than in males in MRT e.g. females use more brain regions which require more activation, and hence more time (Koscik, O'Leary, J.Moser, Andreasen, & Nopoulos, 2009). Further examples are lateralization, where subjects with low spatial ability showed a left field advantage, medium ability showed no field advantage and high ability showed right field advantage (Daniel Voyer & Bryden, 1990), which is in line with the expectation that mental rotation should be one of the expressions of a right hemisphere's general superiority in spatial ability (Linn & Petersen, 1985). A sex difference in mental rotation in human infants has also recently been reported by two studies (Moore & Johnson, 2008; Quinn & Liben, 2008), interpreted as further support for a biological explanation.

An alternative explanation for gender differences in the MRT is based on a hypothetical cultural influence. One frequent proposal is that, from early on boys are either driven towards or given toys and activities that foster development of spatial abilities, like Lego toys, sports activities, and video and computer games (I. D. Cherney

& Voyer, 2010). This difference in training and experience of spatial abilities accelerates as the time goes by, since boys also choose more mathematics classes and therefore experience other spatially-relevant mental activities and training. These real life experiences, enhanced by the school curriculum, may foster the development of spatial abilities in boys while girls may perceive most of the spatial tasks as “masculine.” Possibly, girls are even discouraged and/or feel intimidated by the thought of engaging in tasks that demand excellence in basic spatial skills (Parsons et al., 2004). Although a girl who chooses math classes in school has likely received a level of training that demonstrates her excellence in math, studies have shown that outside the school environment, the same girl may engage in fewer spatial experiences than the boys (I.D Cherney, 2001). Boys on the other hand may perceive themselves as generally competent and would tend to do well in tasks which include spatial abilities. In a study by Moè (2009), men were affected positively when they believed that they were better than women and also expected the task to be easy, while women’s performance were more linked to the expectations based on the self image than on the difficulty of the task.

The outperformance of males in mental rotation can also be accounted for by the choice of strategies that each gender takes when they perform the actual mental rotation task. It seems likely that women would prefer using an analytical strategy, which involves breaking apart the stimulus object and then comparing, in a sequential process, certain features from the sample figure with certain features on the target figure. In contrast, men may prefer a holistic approach such as they would image the whole object and rotate the whole representation in their mind (M. M. Peters et al., 1995). The genders can also vary in their response strategies, in the situation where they compare the sample



figure to the target figure. Men seem to have a leaping strategy, they continue immediately with the next item as soon as they have discovered whether it is a match or a non-match, without verifying it. Women tend to have a conservative response strategy, so that they might do a “double check” of whether the item is a match or non-match, which is more time consuming (Hirnstein, Bayer, & Hausmann, 2009). Evidence from event-related brain potentials research have shown that response preparation begins before mental rotation is finished, and thereby will a conservative strategy cost enormously (Heil, Rauch, & Hennighausen, 1998). Certain features of the mental rotation test stimuli has also been proven more difficult for women, e.g. long trajectories multiline or multispotted (Birenbaum, Kelly, & Levi-Keren, 1994). When alternated the standard block figures in MRT with three –dimensional human figures performance improved more for women than men (Alexander & Evardone, 2008).

Interestingly, mental rotation tasks have also yielded robust and reliable sex differences for decades already, with men typically outperforming women in the task (Hedges & Nowell, 1995; Linn & Petersen, 1985; M. M. Peters et al., 1995; D Voyer & Hou, 2006; Daniel Voyer, Voyer, & Bryden, 1995); although such a male superiority has not been observed consistently with all types of stimuli (Jansen-Osmann & Heil, 2007)) or any mental rotation task (M. Peters, 2005; Daniel Voyer, Voyer, & Bryden, 1995), and as shown above from different perspectives, there is evidence enough to question if men always outperform women.

For example, Moè (Moè, 2009) used a version developed by Vandenberg and Kuse, which is an adaptation of the original mental rotation test developed by Shepard & Metzler (Vandenberg & Kuse, 1978; Daniel Voyer & Saunders, 2004). Moè’s research

show that women's performance was affected by positive instructions about gender (e.g. that they were "better than men"), so much that they reached men's scores in MRT, regardless if they expected a easy or difficult task. Men, on the other hand, were affected by instructions about gender and the task difficulty combined. When the instructions were that they were better than women and the task was easy, men improved their performance but in the condition with the opposite instructions ("men are better than women" and that the the task is "difficult") the female superiority effect disappeared. This suggests that an increase in women's performance in MRT is possible when given positive beliefs about self, to the point of reaching men's performance, while men's performance is more effected by the complexity of the task than the belief about self. Thus, it seems that women's spatial abilities are vulnerable and easy to modify trough attitudinal and experimental factors (Quaiser-Pohl & Lehmann, 2002).

In fact, the magnitude of sex difference has decreased in recent years (Daniel Voyer, Voyer, & Bryden, 1995). Scali et al (2000) found that men outperformed women in MRT only when scored in a strict manner combined with the instruction to focus on accuracy. Men did not outperform women when the focus was on speed (Scali, Brownlow, & Hicks, 2000). Quaiser-Pohl & Lehmann found that even though there was an effect size of gender differences in MRT, it varied the most with students in arts, humanities and social sciences and smaller in computational visualistics. They also showed a correlation with spatial abilities and computer experiences and between test performance and achievement related self-concept which depended on gender e.g. evident mostly in the female sample.

Even though research has shown that men outperform women, because of biological, cultural, strategically or situational factors, there is the suggestion that it is easier to undermine cognitive performance than to improve it, even for men (Wraga, Duncan, Jacobs, Helt, & Church, 2006). The power of effort attribution was proven by Moè and Pazzaglia (2010) in an experiment where they told their participants that performance could either be “effort-dependent” or “genetic-dependent”; namely, e.. “anyone can succeed on this task by putting in effort” versus “performance on this test depends on genetic determinants”. As their findings showed, the ‘effort’ instruction had a substantial positive effect on performance, which was significantly higher than in the other two conditions. It also occurred independently of gender, although males scored higher than women. This finding indicates that it is possible to improve on MRT performance as long as there is a belief that what matters above all is effort and practice and, hence, there is a chance to improve performance (Moè & Pazzaglia, 2010).

#### *Human-computer interaction and spatial abilities*

From an applied point of view, scientific interest for the topic of spatial ability and visual perception has received “renewed” interest, resulting in an increasing attention from fields of study related to technological advances for computer displays. This has been partly motivated by an increase of several orders of magnitude in computational power over the last decades, but specifically in the relatively recent advent of much higher quality of displays, particularly with the introduction of the Liquid Crystal Display (LCD) in the late 1990s. Qualitative improvement in display technology has basically happened along two dimensions: an increase in the pixel count (and smaller size of pixels); and a considerable increase in the size of the displays. A novel display technique

that facilitates “tiling” of displays, making many displays “work together” as one display surface basically eliminates restrictions regarding size of the display – making us technologically capable of producing “unlimited sized displays”. In light of this development, from a technological – and economical perspective rather obvious questions have been raised; how big is “enough”; how big is useful – and the consequential: what role does size play (in visual perception)?

A larger area seems quite useful for many work applications, and a large display has e.g. been shown to yield some productivity-gains over smaller (Czerwinski, 2003). Interestingly, Czerwinski, Tan, and Robertson (2002) found that the “gender gap” in 3D-world navigation (spatial tasks) diminishes and seem to almost disappear when users are given a wider field of view – meaning a broader display than what we are typically used to. Specifically, they set up two projection displays side by side and minimized the “seam” between them in order to have a continuous and coherent display that was twice as wide as an ordinary one. They made a virtual 3D-environment in which the participants engaged in way-finding tasks in a virtual world, a task in which males typically outperform females. They found significant effects of field of view and display size resulting on trial times. There was also a significant effect of gender in the small screen setup where males outperformed females, while in the large screen setup there was no observed difference and men and women performed equally.

In a follow-up study Tan, Czerwinski, & Robertson (Tan, Czerwinski, & Robertson, 2003) found that with a large display setup of 100 degrees field of view (normally around 30 degrees) and with the presence of optical flow cues ( i.e., the effects of moving within the environment has on the change in image on the retina of immobile

objects residing within environment), female performance improved to the point of evening out gender differences. They suggested that a wider field of view and optical flow cues could separately contribute to female performance enhancements.

In a recent study, Tan et al. hypothesized that large displays result in immersing the users in the 3D-world, making them feel more “in” the world, and may bias them to adopt more efficient cognitive strategies when performing spatial tasks (Tan, Gergle, Scupelli, & Pausch, 2006). Specifically, they hypothesized that a large screen biases users in adopting egocentric strategies for those tasks (i.e., “moving” their own person/head within the virtual world or “being there”, as opposed to exocentric strategies – rotating the environment around them). In testing this hypothesis they also used the classical mental rotation (MR) test. However, they found no sex differences in task accuracy or response time. Interestingly, in this study the relative size of the display was constant (i.e., only manipulating physical size as variable), so that the retinal size of the object was also a constant. In a quite small study Suzuki and Nakata (Suzuki & Nakata, 1988) investigated whether the size of objects (retinal size) can affect RTs. Participants (N= 6; 2 females) viewed either small, medium, or large object stimuli, where the retinal objective sizes, measured between the centers of the objects, ranged from 2.9 degrees to 5.7 degrees to 11.5 degrees. They found that angular differences in figures had a clear effect on RTs, as did the differences in retinal size. In accordance with Tan and colleagues (Tan et al., 2006), they found no effect of constant retinal size of objects (i.e. manipulating both the size of the display and distance from it in order to keep retinal size constant). Furthermore, Suzuki and Nakata found that RTs increased with the smaller sized objects, compared to the medium and large sized objects, arguing that this might be explained that

the objects might have to be mentally “scaled up” (going through a “size-normalization process”) before rotation on the smaller object.

In the first study to our knowledge to investigate the issue of object size and mental rotation, Shwartz (1979) examined whether the size of an object and its complexity affect rotation rates. Shwartz had 20 subjects (no mention of their sex) making comparisons between two-dimensional objects, which were shown briefly and in succession on the screen, the first stimulus being also followed by an orientation cue in the direction of which the subject were to visualize the object being rotated. When the second stimulus was given, the subject was to make a same/different judgment about the two objects. Shwartz found that increasing size of the objects resulted in increased rotation time. He also found that larger objects required increasingly more time to rotate farther. One should note that the amount of pixels on a screen generally affects the amount of “inherent information” within the displayed image. That is: the more pixels, the more information can be displayed simultaneously. The physical size of the display however affects the visual angle with which we perceive the image. A small display gives a narrow angle, meaning that the displayed image occupies a smaller fraction of our field of view – or the image “captured” on our retina; a larger display provides a larger visual angle. In our work, we have concentrated on the physical size of the display and how this affects the cognitive aspects of our perception

### *Hypothesis*

Based on the relative scant previous research on mental rotation tasks of large objects and/or large displays, we wanted to investigate both the effect of display size and how this factor may affect the gender bias in the task of mental rotation. Since the

previous findings suggested that the size of the objects should influence rotation times (in some direction), and that display size seems to have a positive effect on female performance, we expected to find effects of screen size on gender. Specifically, we predicted an increase in female performance relative to male performance, given the above mentioned effects of field of view on other spatial tasks. We assume that females prefer an analytic approach to the MR task; for instance, by comparing the figures piecewise, this strategy requires comparing at least two parts of the objects in order to make a call whether they are the same. The preferred male approach of holistic processing requires making an internal representation of the complete objects, rotating one of them in order to compare with the other. Females, on the other hand, would encode only parts of the stimulus (at a time), and then mentally rotate another part. If encoding an object requires visual focus (thought to occupy a minimum 1° of our visual field is in focus at any time; (Eriksen & Hoffman, 1972)), this implies longer MRT response for large objects, given the use of a holistic strategy. The larger the object, the more shifting of focus is required in order to encode the whole object – in order to subsequently perform the mental rotation. A piecewise strategy would require focus-shift behavior by default, and, hence, would presumably not be as affected by object-size, as a holistic approach. In addition, visual working memory could impose restrictions on object size, requiring a down-scaling of larger-than-allowed stimulus objects. The effect should be similar to that predicted by the focus-shift hypothesis.

If we can confirm a positive effect on gender performance regarding display size, then this will be further investigated to establish whether this effect can be a result of

expectancies, e.g. do men or women perform better when they are told they are expected to do so, either in the small display condition or the large display condition?



## Method

### *Experiment 1*

#### *Participants*

Forty-one participants, 23 men and 18 women participated in Experiment 1, with an age range from 18 to 45 years and a mean age of = 29.9, SD= 6.2. Four participants were excluded (one woman and three men) because they didn't respond above the criterion level (i.e., 70% correct). Thus, the descriptive statistics and analyses shown below will consist of responses of the remaining 38 participants. All participants came in voluntarily and were compensated with two lottery tickets, worth 50 NOK. Written inform consent was also obtained from all participants, and at the end of the experiment all were presented with a short questionnaire. Three persons were not native Norwegian, and therefore two were given instruction in English and care was taken to make sure that they understood the task. The third participant comprehended Norwegian well. All participants had normal, or corrected-to-normal, eyesight.

#### *Stimuli and apparatus*

The large display wall consisted of 28 projectors, which projected an image onto a screen surface. There were 7 x 1024 horizontal pixels and 4 x 768 vertical pixels appearing onto the screen, all together forming a display of 22 millions pixels of red, green and blue. The screen size was 230 inches (diagonally).

In contrast, the small screen's size was 14.1 inches on a Dell D600 laptop computer, with the rather standard native resolution of 1400x 1050 pixels and a 24 bit color spectrum. The SuperLab software was running natively (i.e., the local processor and graphics hardware was running and displaying the SuperLab software and program-

interface) 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 was comprised of 28+1 Dell 370 PCs with P4 Prescott, 3.2GHz, 2GB RAM, 1Gbit Ethernet and a 48 port HP switch. See (Jensen, 2006) for a current description of the equipment. The SuperLab interface (with the 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.

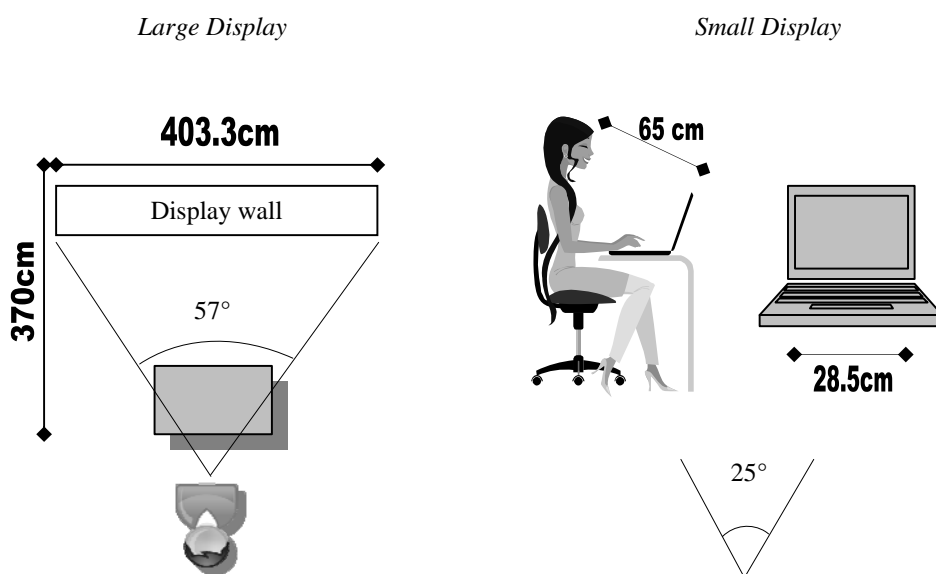
Note that projectors working together to produce a single coherent and continuous image (one “desktop”, if you will) have the unique feature that, if aligned correctly, they can produce one image without the bezels that ordinary displays (LCDs) do when aligned in a matrix. There will, however, be small color-variations between the different projectors, but the resulting “image” can be remarkably coherent. Stacking either projectors or LCD-displays together like this, one can produce a display of almost unlimited sized with a number of pixels proportional to the number of display devices in the configuration.

Measuring an exact viewing distance was not possible, since the participants were instructed to maintain “comfort viewing distance” from their chair and table.

Nevertheless, the table remained at the same point at all times, and was placed 370 cm from the large screen. As a result, viewing distance from large and small screen respectively, hence, was about 370 cm and ca 65 cm. The viewing angle for large and small screen in our experiment is shown in Table 1. ‘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. Figure 1 shows the display setup with corresponding visual angles.

*Table 1 Visual angles in the two display setups*

Angle explanation	Large display	Small Display
Total visual angle display	57.0°	24.7°
Angle Between objects	27.3°	11.8°



*Figure 1 Display setups and corresponding visual angles*

### *Procedure*

The experiment took place in a room at the department of computer science, University of Tromsø. Temperature was set at 20° C and light setting to dark. All participants were tested in the same room with the same equipment within a time period of two month, from 20<sup>th</sup> of June to 1<sup>st</sup> of august 2007. They were randomly assigned to either start with small screen and with positive or negative hypothesis, or with large screen with positive or negative hypothesis. Participants received information about the hypothesis prior to the experiment. There were two contrasting versions of the so-called ‘working hypothesis’: 1) The “positive hypothesis” stated that a large screen can make all people perform better at mental rotation tasks; 2) the “negative hypothesis” stated the opposite; i.e., worse performance was most likely observed with the large screen than with the regular screen. Experimental sets were counter-balanced for gender and practice effect in within-subjects design (M.

M. Peters et al., 1995). The task itself was self-paced and each object remained on the screen until the participant made a choice and pressed a key, indicating a ‘same’ or ‘different’ response. The task was to judge whether two objects which appeared at the screen were



*Figure 2 Example - MR stimuli on the Display Wall (Large display condition)*

identical or not. The computer recorded the result for key pressing using the SuperLab software with 266 trials in a randomly order. Participants were also given a questionnaire

containing questions about their age, gender, level of education and current occupation. Among the queries, participants were probed about whether they had played computer games, for how long and what kind, two or three dimensional. In an attempt of trying to find a relation between performance and a preliminary mind-set or expectation, we also asked what they thought was best to use the small or the large screen. All participants were alone in the room, with the door closed.

#### *Design and statistics*

The design for both phases, large and small screen, was identical, and consisted of 266 trials in each phase. The design was a mixed design where the factors, Screen Size (Small and Large) and Angle ( $0^\circ$ ,  $30^\circ$ ,  $60^\circ$ ,  $90^\circ$ ,  $120^\circ$ ,  $150^\circ$ ,  $180^\circ$ ) were treated as within-subject factors. Gender and Order (Large first/Small first) were between-subjects factors. Order as factor was implemented to control for a learning effect (M. M. Peters et al., 1995) since all participants did the same experiment twice (once for each display size condition).

#### Results from Experiment 1

First, descriptive statistics for each participant were calculated. We computed mean response times (RTs) for correct responses and mean percentage of accuracy for each combination of the experiment variables (Screen Size, Match, Angle). Preliminary analyses showed no main effects or interactions in either accuracy or RTs for Order as a factor. Hence, we ignore Order as a factor in all of the analyses presented below.

#### *Response Times (RTs)*

Mean RTs were collected for all correct trials and for the Match conditions (same shape) collated for each of the angular conditions. Differently-shaped objects have no

point of alignment and, hence, cannot be meaningfully included in an analysis of RTs.

We then performed a repeated-measures ANOVA analysis with Screen size

(Small/Large) and angle ( $0^\circ$ ,  $30^\circ$ ,  $60^\circ$ ,  $90^\circ$ ,  $120^\circ$ ,  $150^\circ$ ,  $180^\circ$ ) as within subjects factors and

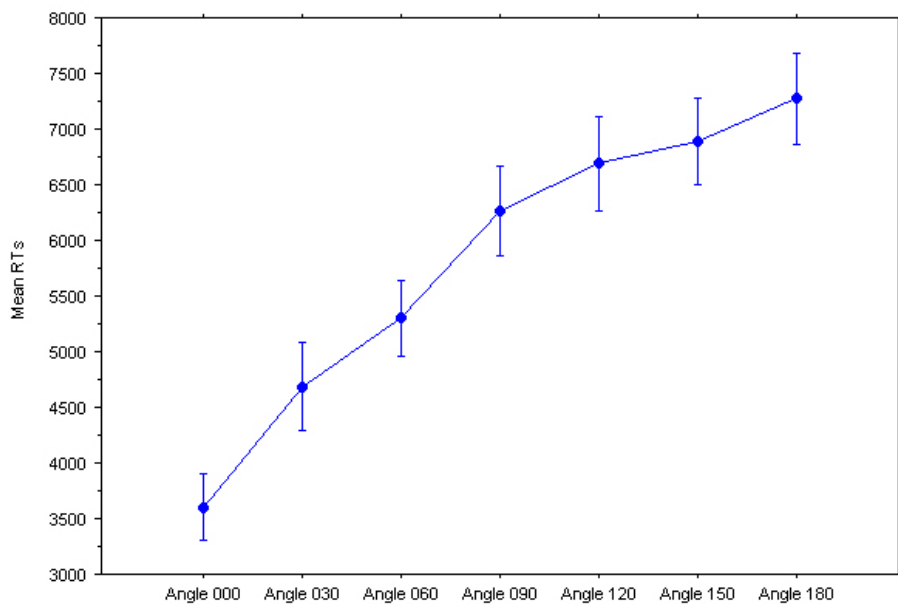
Sex (female/male) and Expectation (For/Against the hypothesis) as a between-subjects

factor. The “For” hypothesis was the statement that the Large display should yield

superior performance, while “Against” stated the Small screen as the superior condition.

The ANOVA analysis revealed a main effect of Screen Size  $F(1,36) = 10.41$ ,  $p =$

.003, and an interactive effect of Sex with Screen and Expectation,  $F(1,36) = 4.15$ ,  $p = .04$ .



*Figure 3 Positive Linear Effect of Angle on RTs (in milliseconds; bars represent the Standard Error).*

A main effect of Angle was also present  $F(1,6) = 78.75$ ,  $p < 0.001$ , generally

validating the experiment with respect to the original Shepard and Metzler results with

the MRT. Figure 3 illustrates a positive linear effect on RT s with increasing angles of rotation. The complete ANOVA table can be found in Appendix B3.

In Figure 4 we plotted mean RTs for the four groups of participants in each of the Screen Size conditions, so as to illustrate the interactive effect of Sex with Screen Size and Expectation. From the Figure 4 it is clear that there was a general slowing of performance for all groups in the Large Screen condition (LS), relative to the Small Screen (SS), with exception of the group of females expecting to do better in the LS. Figure 5 shows this plot in a different fashion, with Screen Size and Sex are along the X-axis with separate plots for hypothesis only.

Given that only the group of those women who expected to do better in the LS than in SS performed faster in the former condition, we suspected that these subjects were simply trading Accuracy for Speed. This would be evident from a decrease in Accuracy in the large Screen condition for this group. In order to assess whether or not this was the case we first computed the overall RTs for each participant.

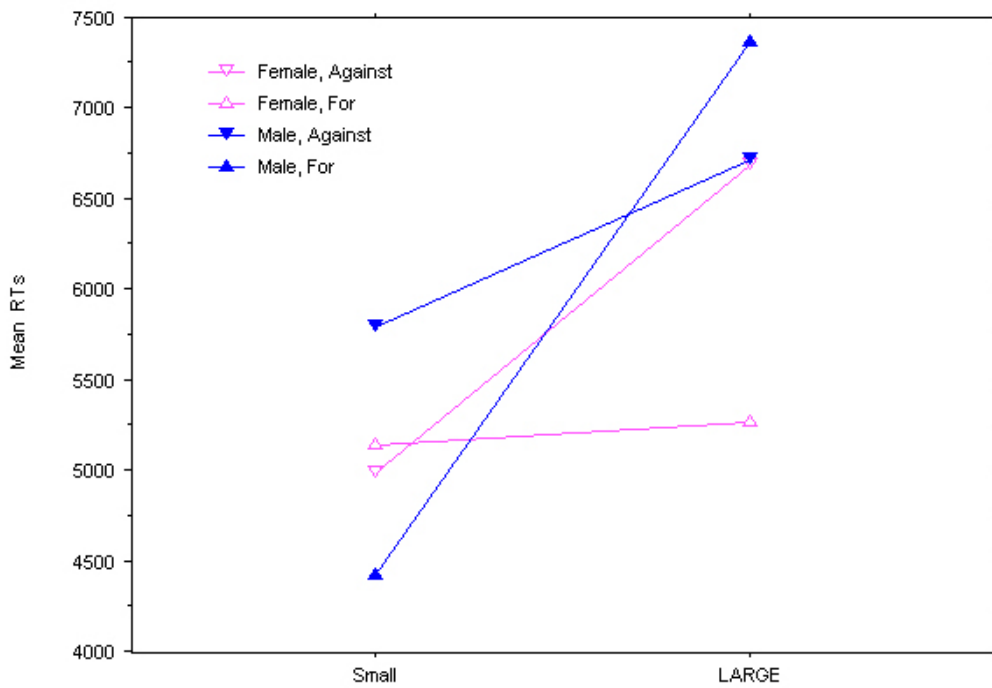


Figure 4 Plot: Interactive Effect of Expectation with Sex and Display Size

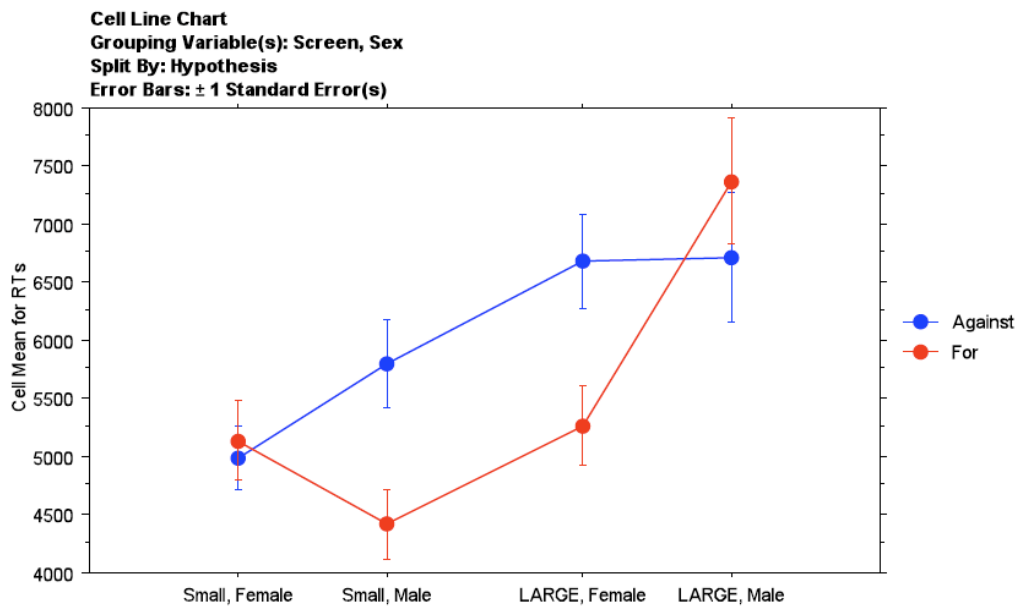
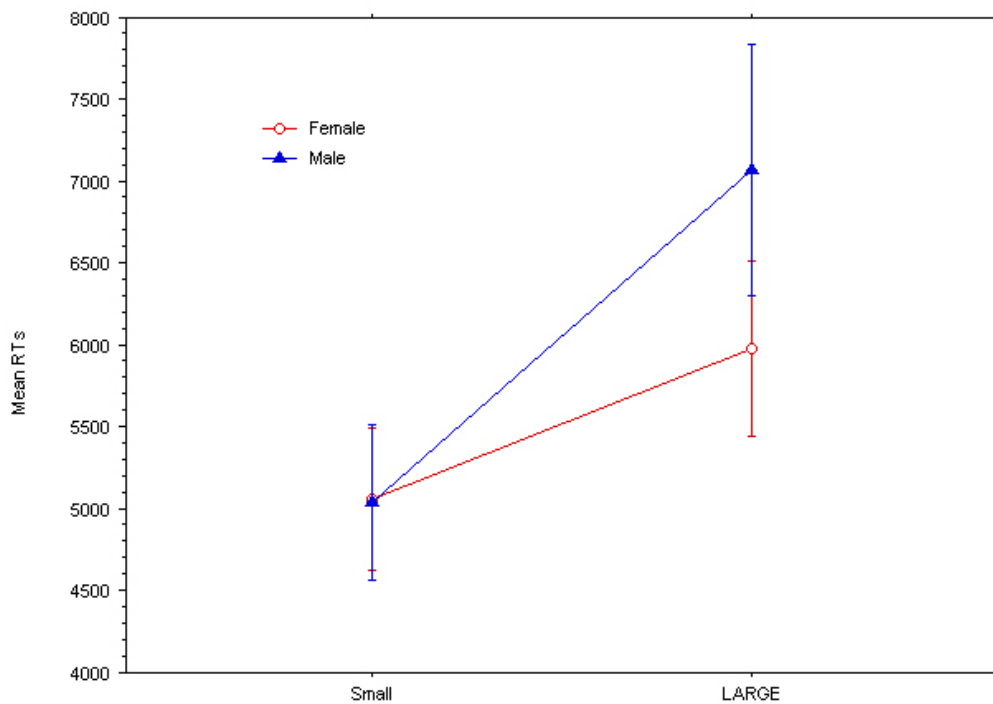


Figure 5 Mean RTs for Screen Size and Sex, split by Hypothesis



Next, we performed a simple regression analysis with accuracy as the regressor and RTs as the dependent variable. We found no evidence of any speed-accuracy tradeoff for the whole group (N=40), since  $r^2 = .04$ ;  $F(1,38) = 1.4$ ,  $p = .24$ . The slope's coefficient was also negative (-45.7), indicating that RTs tended to be shorter with increasing accuracy. The same regression analysis was repeated for the group of women expecting to do better in the LS than SS. Again, no speed-accuracy tradeoff was indicated in this group, as the slope's coefficient was still negative – and larger (-189.7), thus showing a significant correlation:  $r^2 = .42$ ;  $F(1,7) = 5.1$ ,  $p = .05$ .

The working hypothesis in Experiment 1 was that females should benefit from the Large Display. In general, this was discovered to be the case, as Figure 6 illustrates (Women: mean RT= 5975; SD= 2736; Men: mean RT= 6578; SD= 3828). We also computed a 95% confidence interval, using the formula of Loftus and Masson (1995) for within-subjects design (C.I. = 523 ms). A female advantage in the large screen condition of 603 ms exceeded this C.I.



*Figure 6 Female advantage in the Large Screen condition*

In the Small Screen condition there was virtually no difference in RTs between the sexes.

### Discussion

On the basis of previous research, suggesting that large displays can improve the performance of female participants (Czerwinski, Tan, & Robertson, 2002), we predicted that a large display would result in better performance for women compared to men in the mental rotation task.

Accuracy scores showed no significant effects for sex, or any interactional effects for this factor with the others. However, we found that women performed the MRT faster than men in the large display condition than with the small screen. Most remarkably,

women outperformed men in speed in the display wall condition, despite the mental rotation task has typically revealed a robust sex difference favoring men (e.g., Hedges & Nowell, 1995; Voyer et al., 2006). In contrast, the two sexes did not differ in either accuracy or speed of correct responses with the small screen. The latter null finding may not be unusual in studies where only two shapes are compared within a single trial, since several studies with the classic Shepard-Metzler-like mental rotation tasks have not found sex differences (e.g., Butler et al., 2006; Voyer et al., 2006). As Peters and Batista (2008) point out, robust sex differences are typically observed with another variant of the mental rotation task, originally designed by Vandenberg and Kuse (1978), where one target figure is compared to other four comparison figures (of which only two are rotated versions of the target and the others are differently-oriented mirror images).

However, even though females appear to outperform males in this study, when we split the mean RTs by hypothesis we found that only the group of females expecting to do better on the large display performed significantly better than the other groups, as illustrated by Figure 4 (bottom pink line) – as indicated by the interactive effect of Sex with Screen Size and Hypothesis. The finding that women with prior expectations in favor of Large displays led us to suspect that they were trading speed for accuracy, which was ruled out by simple regression analyses. We also need to note that there was a general slowing of performance on the large display, as illustrated in the Figures. In this regard we also need to note that there was a small delay induced by the Large Screen condition, since the stimuli needed to be transferred via a network connection to the Display Wall cluster, where the stimulus would be enlarged within a VNC server to fit the larger display area. Unfortunately, it was not possible to measure exactly this delay in

this experiment but we estimated it to vary from 500 ms to above 1000 ms. In the next Experiment we were able to measure this confound more precisely. At any rate, a delay would be equal for all participants and therefore would not affect differences in the between-subjects results. However, the within-subjects significant main effect of Screen was not reliable for this experiment. Basically, these results made us conclude that yes – females do perform better than males on a large display that provides a wider field of view – as expected. However, the two variables Display Size and Expectation showed a significant interactive effect in this first experiment. In order to tease apart the effect of Expectation it may be clarifying to look at Figures 4 and 5 focusing on each of the genders individually. Concentrating on men, we see that in the Small Screen condition, those men who (should) expect better performance with a large screen outperform the group with small screen superiority-expectation with well over 1 second (mean RT). Turning to the large display condition, it would seem that this “phenomenon” is present in the large display condition as well: men with small screen superiority-expectation outperform those with large screen superiority-expectations, however with somewhat less margin (ca 500ms). At first glance it looks almost as men “resisted” the presented hypothesis. Moé (2009) found that males’ performance was affected by task difficulty instructions, arguing that men were susceptible to induced “pressure” in expectations of task complexity. Men seem to have “improved performance in the non-threatening situations, that is when nothing was said about the gender and failure could be explained by the difficulty of the task” (ibid. p. 25). An important difference between effects of “expectation” in our experiment compared to Moé’s is that our instructions were completely gender-neutral. Hence, we do not have potential “stereotype threats” in our

experimental design per se. However, instructions were that either Small or Large screen condition should provide superior performance, which may well constitute a “pressure” situation, providing a fear of not confirming such expectations. Furthermore, Moé (ibid.) found that females were unaffected by expected difficulty of task, but were affected by expectations based on the self, e.g. that they are better than men. Glancing to our Figures 4 and 5 again, we notice that females’ expectation did not affect performance in the small screen condition; both groups performing equally. However, in the large screen condition the groups diverged notably: females expecting better performance in the large display condition did considerably better than those expecting better performance in the small screen condition (differing by more than 1 second in mean RTs). It is already clear that effects of expectation are different between genders in our Experiment 1. It may well be that men are affected by fear of not confirming expectation of superior display condition, while for females it is rather unclear why the novel display condition is affected by expectation while a “standard” display condition is not. We may hypothesize that novelty – alone – (in the large display condition) is the factor that elicits an effect of expectation among females. If we assume a traditional gender-effect for large displays (equal to that of small-), we would expect either male superiority or equal performance between genders with regards to RTs. This would imply that a “novelty” effect, combined with expectations provide females with a positive, performance-enhancing effect. Conversely, if we assume female superiority in the large display condition, the reverse effect would be observed: “novelty” effect combined with expectation only provides a negative effect (i.e. worse performance) in the large display condition.

Hence, the interaction of Sex with Display Size and Expectation is ambiguous, since there was little way for us to actually know if it was only in the presence of an explicitly-stated positive or negative expectation that this effect would reveal itself. In order to pursue our research question and disentangle the effects of expectation from the effects of display size, we needed to replicate the first experiment's results, while omitting any explicit suggestion of a working hypothesis prior to the experiment.

*Experiment 2 Introduction*

In Experiment 1 we hypothesized that display size would affect performance on gender, and if so if it was caused by instructions about what screen they were supposed to perform their best on. In Experiment 2 we left out the instructions about expected performance and instead asked afterwards what they believed they performed their best on e.g. the small display condition or the large display condition.

*Participants*

Sixty-eight participants, 36 men and 32 women participated in a second experiment, with an age range from 18 to 51 years and a mean age of = 24.0, SD 5.44. Nine participants were excluded; four woman 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 didn't 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.

A crucial difference from the first experiment was that the 'working hypothesis' was not made explicit to the participants (in either the positive or negative form). In other words, we did not inform the participants (as it is the standard procedure in psychological research) about what, according to the experimenters, could be expected as the effects of display size or of gender. We also modified some of the questions in the questionnaire presented at the end of the mental rotation task. From the results of Experiment 1 we did

not know if participants would perform better or worse without such an explicit positive expectation given by the experimenter to the participant. However, in order to evaluate for subjective or implicit expectations among the participants, we included questions regarding this in the post-test questionnaire that participants answered after finishing testing in both display conditions.

Another change in the method was to reduce the number of angles of rotation that were tested within the mental rotation test in order to reduce the total number of trials. Comments from the participants in the 1<sup>st</sup> Experiment made us realize that the within-subjects design for display size made the total number of trial strenuous for the participants. Hence, we reduced the number of angular conditions from 7 to 5, including only angular differences of 30°, 60°, 90°, 120° and 150°.

Finally, with the help of a high-speed camera we were able to exactly measure the mentioned delay in the large-screen condition in the second experiment. We used a Casio Exilim EX-F1 for this purpose, basically videotaping one complete experiment with both displays at 300 fps, counting frames between updates of the small screen and the large – hence, providing us with a set of exact delay timings. The rest, design and statistics were the same as for the first experiment.



## Results, Experiment 2

As done previously, we calculated descriptive statistics for each participant so as to obtain mean RTs for correct responses and mean % accuracy for each combination of variables (Screen Size, Match, and Angle). Nine subjects (4 females, 5 males) were removed from the analysis since they scored below the 70% cut-off.

The Large Screen delay was measured as occurring on average after 848 ms (using the mentioned high-speed camera), which – in the analysis below has been subtracted from the mean Large Display RTs for all participants.

In order to assess the effects of Sex and Screen size, we performed an ANOVA with the angular condition (30°, 60°, 90°, 120°, 150°) and Display Size as within-subjects factors and Sex and Expectation as between-subjects factors, using SPSS16 for the purpose. Expectation was the variable recorded in the debriefing questionnaire where we asked which screen size condition the participant thought that they performed best on, with respect to time (RTs).

The ANOVA did not reveal any main effects or interactive effects for Accuracy in this second experiment either.

The previously observed linear relation between angular disparity and RTs was replicated as a main effect of Angle was observed  $F(1,4)=36.28$ ,  $p < .001$ . In the Between-subjects effects we also observed a significant effect of Expectation  $F(1,1)=12.64$ ,  $p=0.001$ . An interactive effect of Sex with Display Size was also registered,  $F(1,55) = 5.88$ ,  $p=.019$ , along with an interactive effect of Display Size \* Expectation,  $F(1,1) = 9.89$ ,  $p=0.003$  and an interactive effect of Angle \* Expectation  $F(1,4) = 2.89$ ,  $p=0.03$ . These tables are given in Appendix B2.

To illustrate the main effect of Angle we have plotted mean RTs for angular disparities in Figure 8, split by sex – one graph per display condition. On the left hand side we observe that in the Large Screen condition, female performance followed a slope well below the 95% confidence interval of male’ performance in all angular conditions. The male slope appears also somewhat steeper towards increasing angular disparities. In the Small Screen condition (right hand side) there is virtually no difference in RTs for the two genders illustrating equal performance on the small display.

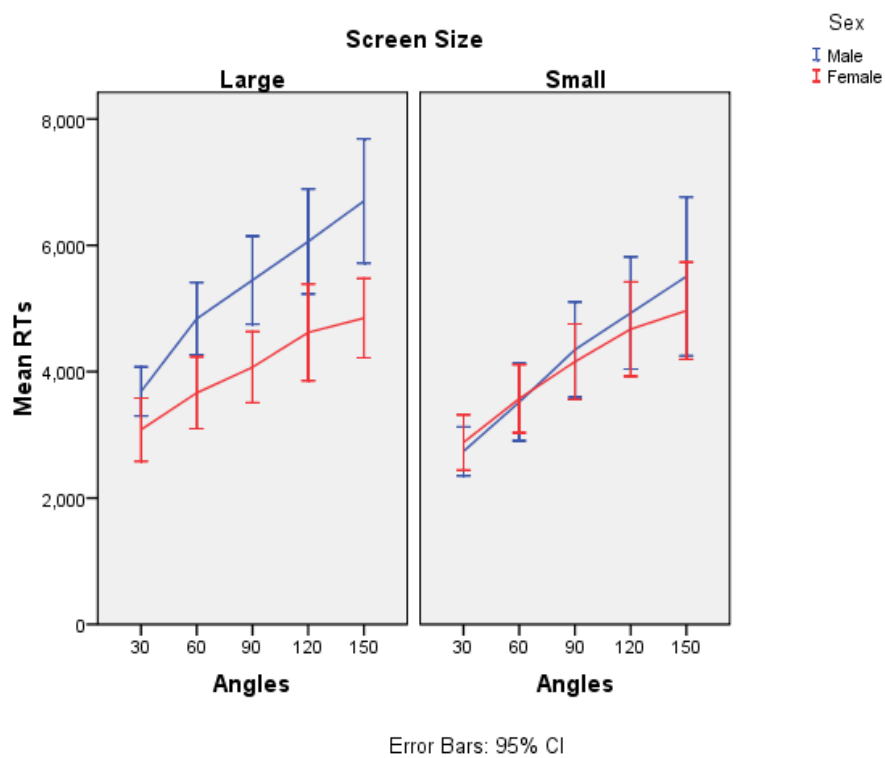
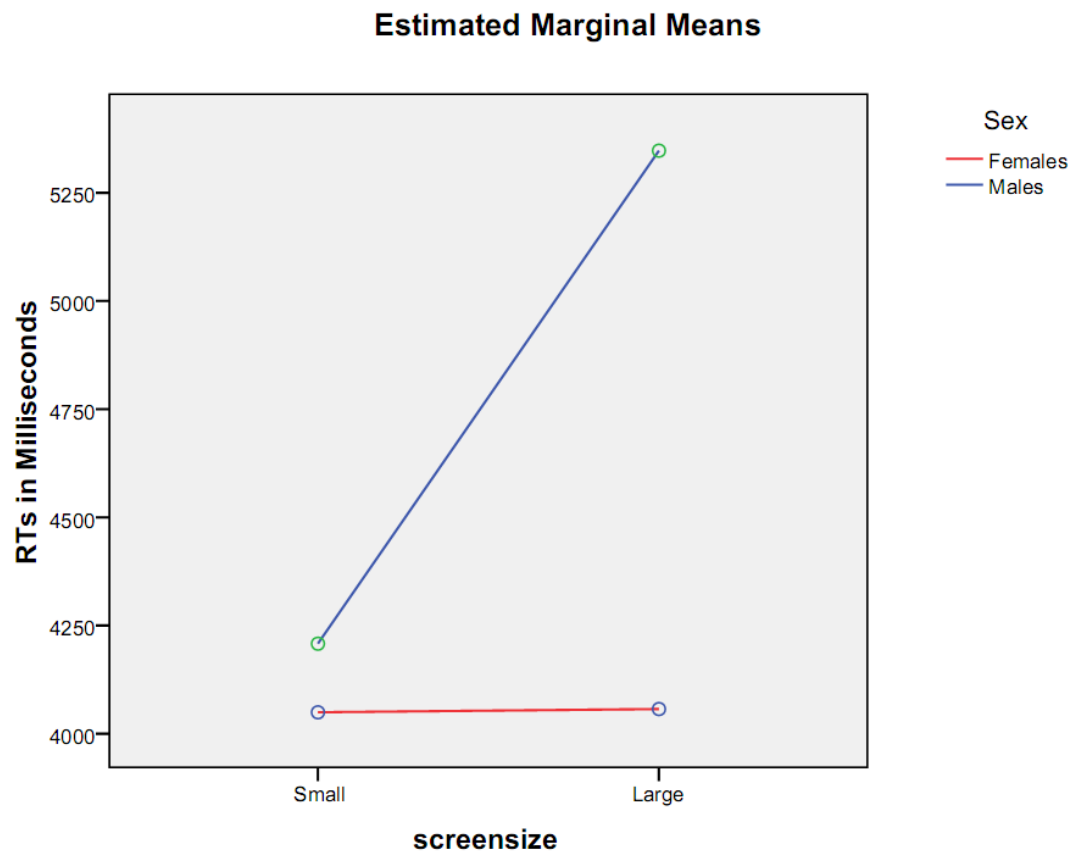
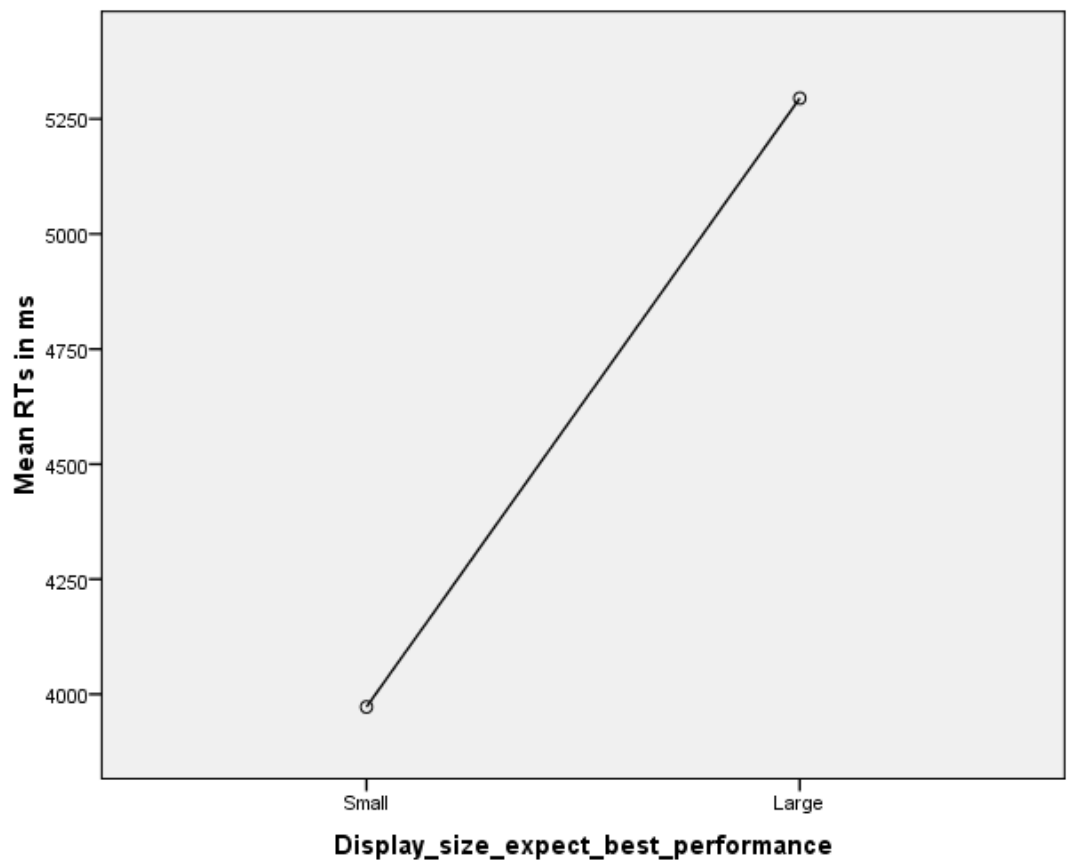


Figure 7 Mean RTs (in milliseconds) for angular disparities in the two Display conditions (Large on the left, Small Screen on the right), split by Sex.



*Figure 8 Mean RTs for the two Screen Size conditions, split by Sex*

In Figure 8 we have plotted the interactive effect of Sex with Screen Size. We can see in this figure how males' performance increased significantly in the large screen condition (blue line) of about 1 second, while female performance seemed unaffected by screen size (red line).



*Figure 9 Mean RTs for Expectation of which better screen size condition*

In Figure 10 we have plotted mean RTs for the participants' Expectation (for which display the participants expected shorter RTs) for better Display Size condition, illustrating the main effect of Expectation mentioned above. We observe that those participants expecting the Small Display to have yield superior performance (with regards to speed) are, on average, over 1 second faster than the group expecting better performance in the Large Display condition.

Figure 10 plots expectation in the two display conditions and illustrates the interactive effect of Expectation with Screen Size. From this plot we generally note, that participants expecting to do better in the Large Display condition are overall slower than

those expecting better performance in the Small Display condition as also seen in Figure 9. However, we also note that expectation of small display superiority have a noticeable increase (positive ‘slope’) in RTs from Small to Large screen of about 1 second. For the group expecting large screen superiority there is no such increase, but rather a small decrease in mean RTs (negative ‘slope’) for the Large Screen condition.

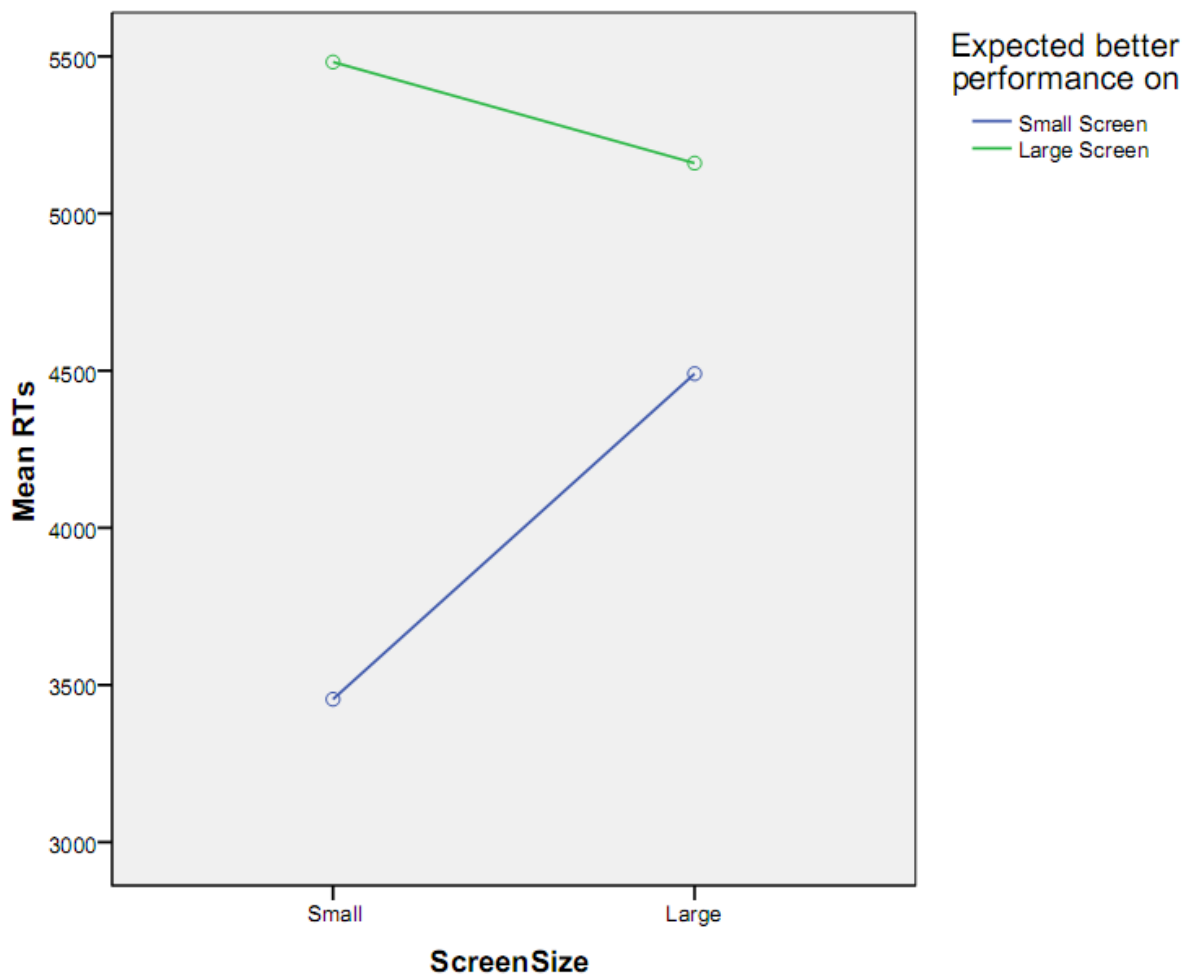


Figure 10 Mean RTs (in milliseconds) for Screen Size conditions, split by expectations for shorter RTs.

In Figure 11 we have plotted the observed interaction of Expectation with Angle. Apparent in this figure are the shorter mean RTs for those expecting small screen superiority, while those with large screen superiority also seemed to have a steeper slope – that is, increasingly higher RTs for increasing angular disparities.

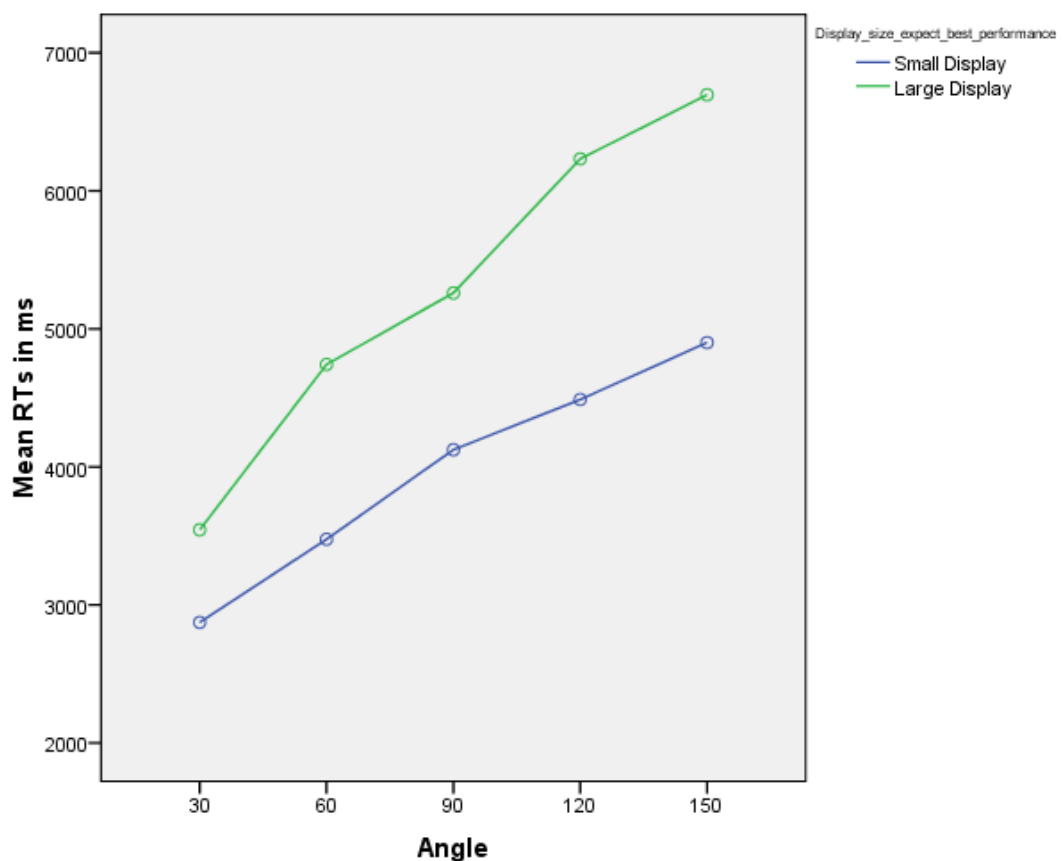
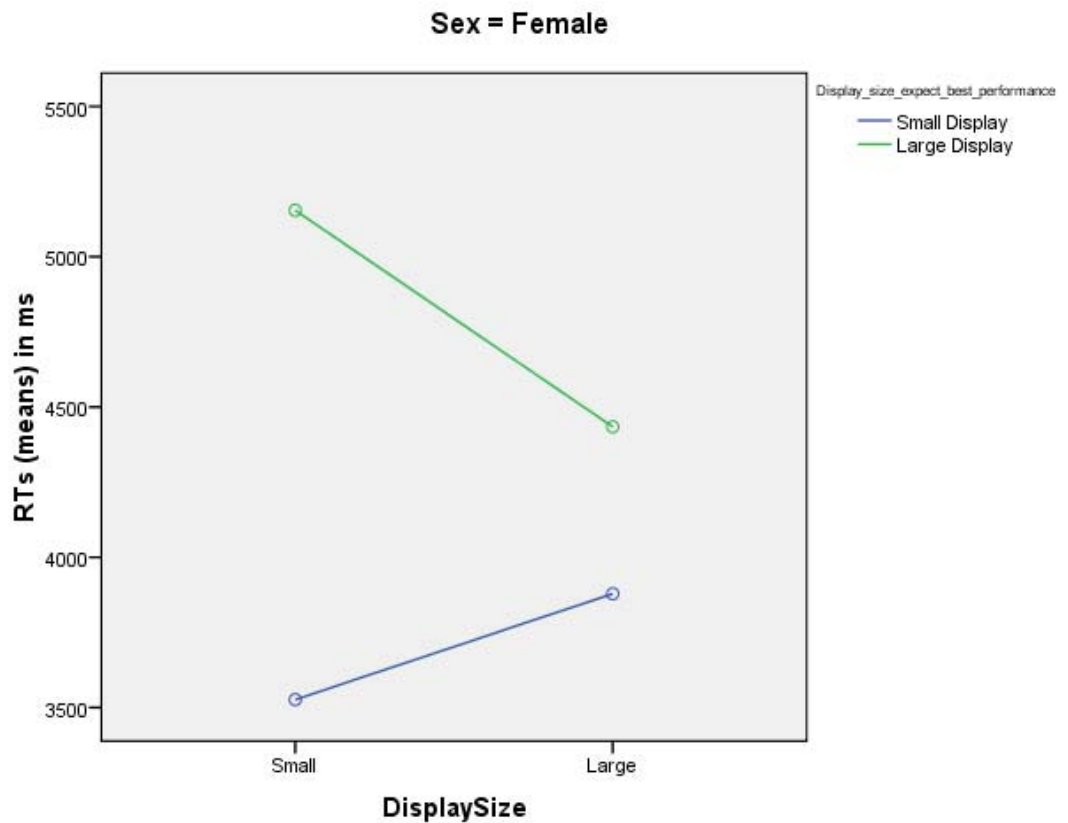


Figure 11 Mean RTs for the angular conditions split by Expectations of superior display size condition

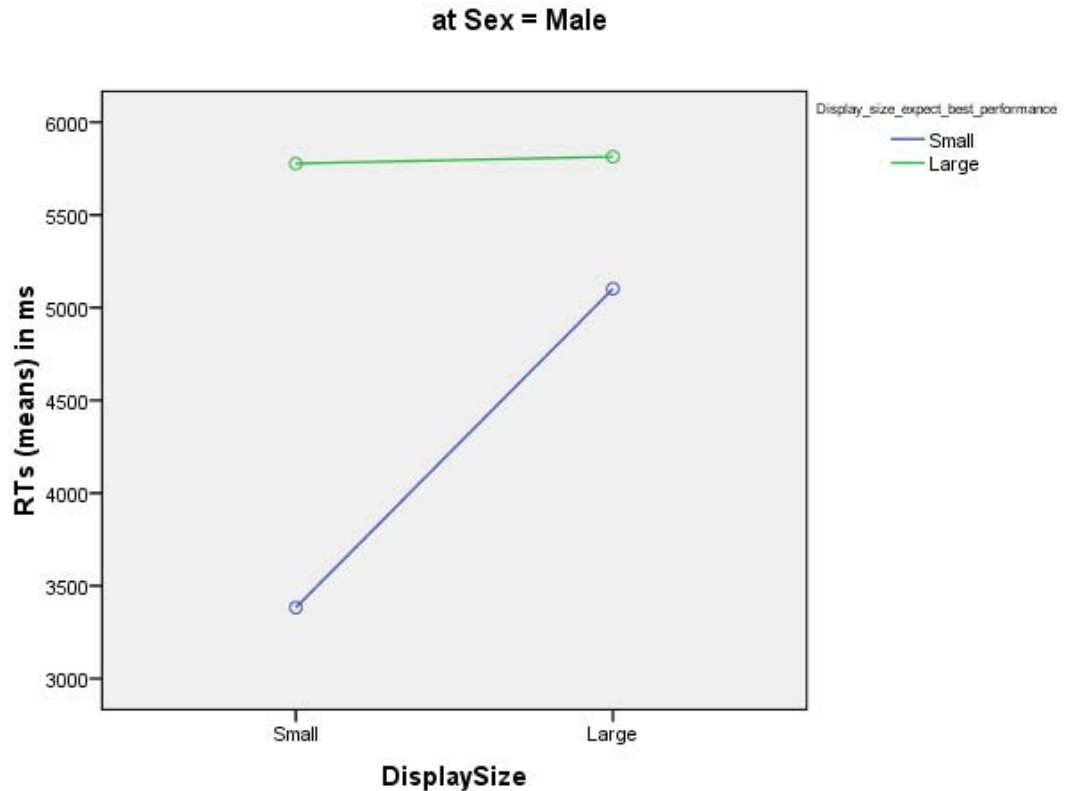
From Experiment 1, we observed different effects of *induced* expectations among the genders in the two display conditions. We note that in this second experiment, the expectation was not induced, but were rather subjective *impressions* from participants of which display condition they thought that their performance would have occurred faster. From the ANOVA in Experiment 2 we did not observe any significant interaction, since

$F(1,1)=0.48$ ,  $p=0.49$ . At any rate, we provide in Figure 12 and 13 the plots for this (null) interaction.



*Figure 12 Mean RTs for Females in the two display condition, split by Expectation (of superior display condition)*

From figure 12 we note that Expectations of small screen superiority (SSS) for females is associated with shorter mean RTs in both display conditions. However, the difference between the mean RTs differs in the two display conditions; above 1.5s in the small screen condition and little over 0.5s in the large screen condition. We also observed an increase in the mean RTs for the SSS group from Small to Large display condition, while for the group expecting Large Screen Superiority (LSS) the tendency was shorter RTs for Large Screen condition compared to their Small Screen condition.



*Figure 13 Mean RTs for Males in the two display condition, split by Expectation (of superior display condition)*

In Figure 13 we have plotted mean RTs for Males in both screen conditions with separate lines for expected Screen size superiority (SSS and LSS). The trend is basically the same as for females: shorter RTs associated with SSS and longer with LSS. Increasing RTs are again observed for the SSS group (as for females) going from Small Screen condition to Large. However, there is virtually no difference between mean RTs in the two display conditions for the LSS group among males. The difference in mean RTs in the Small Display condition was also somewhat larger between the SSS and LSS groups: over 2s difference is observed (left side of plot). Lastly, we note that about 1/3 of the participants expected the large screen condition to yield superior MRT performance (9 out of the 28



females and 10 out of the 29 males). Table 2 below presents the exact mean RTs for the two figures (12 and 13) above.

6. Display\_size\_expect\_best\_performance \* Sex \* DisplaySize

Measure:MEASURE\_1

Display_size_expect_best_performance	Sex	DisplaySize	Mean	Std. Error	95% Confidence Interval	
					Lower Bound	Upper Bound
liten	kvinne	1	3526.337	337.891	2848.614	4204.060
		2	3878.674	365.704	3145.165	4612.183
	mann	1	3382.674	337.891	2704.951	4060.397
		2	5101.958	365.704	4368.449	5835.467
stor	kvinne	1	5154.511	490.944	4169.802	6139.220
		2	4434.133	531.355	3368.370	5499.897
	mann	1	5776.860	465.750	4842.683	6711.037
		2	5813.920	504.088	4802.848	6824.992

Table 2 Mean RTs and 95% CIs for the Expectation \* Sex \* Display Size interaction

## Discussion Experiment 2

The interactive effect of screen size with gender found in Experiment 2 basically informs us that there is a general slowing of response times in the large display condition (as observed in Figure 5) – meaning a significantly slower rate of mental rotation on the larger, wider-field of view display – for men. Remarkably, females appear to have an equal performance in the two display conditions. Thus, we replicated the main result of Experiment 1 that females mentally rotate large object faster than males. From Figure 5 it is also apparent that there is no significant sex difference in the small screen condition, which is – as mentioned in Experiment 1, not an unusual finding per se in this version of the MRT. From this finding of no sex difference in the small screen condition we may “conclude” that the two sex samples in our study seemed to be quite equal in both experiments. A lack of a sex difference in MRT is generally explained by education, practice, training and field of study (interest). We surmise that our samples may have been more “gender-neutral” with respect to spatial ability (the MRT in particular) than those that are traditionally tested in other studies. In fact, for both experiments we have recruited about half of the participants from students in the hard sciences (i.e., Computer Science, Physics, Chemistry and some from Biology). Such a selection merely reflected the opportunity of recruiting students from departments located nearby the laboratory (e.g., students in the social sciences and Psychology in particular, would be geographically located at a greater distance on campus), given that the experiments were conducted in the Faculty of Science and Technology.

In the analysis of “Expectation” in Experiment 2 – which is the subjective feeling recorded after completing the experiment in both display conditions of which display the

participants expected their performance to be better on with regards to speed we found no interactive effect with gender – contrary to the case in Experiment 1. We did, however, find an effect of expectation, apparent from Figure 8, which was not gender-specific. The results plotted in Figure 9 implies that those expecting to do better in the small screen condition did perform better – overall – than those expecting the large display to yield higher MRT rate. The effect of “Expectation” presented itself as a main effect and in two interactions: with Screen Size and with Angle. As a main effect it is clear that the group expecting to do better in the small screen (SSS) had a much higher mental rotation rate than the LSS group with about 1.25 seconds lower mean RTs. As “Expectation” in this second experiment constitutes a feeling of which display condition felt “easier”, i.e. shorter RTs, it is probable that this variable is tied to the strategic approach used for mental rotation. A holistic approach could be experienced more difficult on a large display (as discussed in the Introduction), since it would require more shift-of-focus behaviour. It could well be that the SSS group (preferring the Small display) could be using holistic processing, while the LSS group use analytical approach. However, since there is no apparent difference for the analytic approach for the two screen size conditions (no preferred screen size), it could well be that many of these (about half) would more or less randomly prefer large or small display. This hypothesis – tying the “Expectation” to MRT strategy (Small screen for holistic, Large screen for analytic), is supported by the lower RTs for the SSS group, since holistic processing is more efficient. Moreover, in the Expectation \* Screen size interaction we see from Figure 10 that the only group negatively affected by the large screen condition compared to the small screen condition was the SSS group, while the LSS group experienced an improved performance

in the large display condition, also consistent with a holistic approach to the MRT. In the interaction with Angle, the Expectation variable confirms that the effect endured over all angular conditions (Figure 11), and also seemed to increase somewhat with increasing angular disparity (i.e., resulting in a slightly steeper slope).

Surprisingly, as illustrated in Figure 13, the males who expected better performance with the large display had no increase in mean RTs in the large display condition compared to the small screen condition. Considering that both experiments found that females mentally rotate large objects faster than men, it is interesting that we can also identify a group of males where this effect does not apply. Possibly, these men were also using an analytic processing strategy, and this strategy is generally not affected by object size.

If we also consider in Figure 12 that the increase in mean RTs from small to large display conditions is rather modest with the SSS group of females (sub 400ms), compared to the male SSS group (over 1700ms increase), one could conclude that (1) there is a substantial part of the female SSS group that use analytic approach and (2) there might be another unknown variable that contributes additionally to this female advantage in MRT rates for large objects. A presence of another, unknown factor is supported by the considerable drop in RTs for the female LSS group in the large screen condition, since the analytic strategy should not involve a performance *advantage* in the large screen condition, compared to the small screen condition. This effect is also not observed with the male LSS group, who has equal performance in both screen conditions, indicating its presence only with females.

### General Discussion

In two experiments we found that, contrary to most findings for the spatial task of mental rotation, females can outperform males on a large display with wider than normal field-of-view. A females' superiority on the large display occurred not only in Experiment 1 where the idea that a large display is superior to the standard one was explicitly implanted in the participants' expectations but also in Experiment 2, where there was no external manipulation of expectations. If the females' superior performance on the large display in Experiment 1 was due to a differential effect of expectation in the two genders, then we should have observed a reduction or no sex difference also in the large display condition of Experiment 2. Thus, in light of the results from the 2<sup>nd</sup> experiment, females are superior to males with regards to efficiency of response (speed) on a large screen. Hence, our previous conclusion that it might have been the females' expectation to do better in the large screen condition that affected their performance (see Figure 4) receives a different meaning in the light of the results from our 2<sup>nd</sup> experiment. That is, most likely, an expectation to do worse in the large screen condition yielded a negative effect on females' performance. In other words, without prior expectations, females' performance on the large screen should be equal to their small screen performance and, in general, females should outperform males in the large screen condition. It must also be considered that in Experiment 1, the Large Display delay (while stimulus was transferred to the large, tiled display) was not subtracted from the analysis. We know that this delay had a mean value of minimum 848ms (since the equipment used in the study could only have improved this delay – not worsen it,

technologically). If we glance at Figure 4 again, we see that if we subtract about 800ms from the results presented, the group of women expecting large screen superiority would be faster on large screen than on small screen (by about ca 800ms), while the other female group expecting worse performance in the large display condition would still be slower in the large screen condition than in the small by over 500ms (which was the CI for these mean values). In the small screen condition there was no difference for the female groups. This could indicate that females *were* in fact affected by expectations (or instruction of such), that do not affect the self image but rather a belief about the task's difficulty – contrary to what Moé (2009) hypothesized. This conclusion, however, requires that females are in a “novel” situation, e.g. using a display wall, such as in this experiment, where they had no prior experience with or prior knowledge. In fact, in the small screen condition there was no such effect of expectation for females (see Figures 4 and 5).

Our investigation of the effect of expectation (after completing the experiment) in Experiment 2 only revealed a difference between those who implicitly preferred or expected results to be better in the one screen condition compared to the other, as shown in Figures 9, 10 and 11. This may very well imply different mental strategies for the mental rotation task (holistic vs. piecemeal), as the ‘slopes’ of RTs over angular disparity of the MRT figures showed different effects in the two screen size conditions. This conclusion is supported by the effects of “expectation” for males in the two screen size conditions: male SSS group (using holistic strategy) is negatively affected by large display condition, while male LSS group is not affected. The female LSS group, however, showed *decreased* mean RTs in the large screen condition, indicating that there

is another factor contributing to the female advantage with mental rotation of large objects, since object size should not affect analytic processing strategy – in either direction (with regards to mean RTs). This additional factor may be related to the extra “pause” that was induced by the network and processing overhead in the large display condition (of 848 ms). A recent study of *interstimulus interval* (ISI), investigating the effect of increasing the pause between stimuli, has proved increased accuracy results from larger a ISI for an auditory task (Gomes, Barrett, Duff, Barnhardt, & Ritter, 2008). It is possible that females are more able to exploit the ISI present in this study than men, and without the delay we would see less of an effect of the display size. However, since the MR task was designed to be self-paced (i.e. participants decided when to respond), participants could easily pause at their own will, regardless of this “enforced” pause.

Female choice of strategy may also use more brain regions in MRT, hence the more grey matter volume than men, and they are more structurally organized, all factors together may accumulate into an advantage e.g. faster processing, in large display condition. Men on the other hand might have to start making the necessary areas more structure before they can mentally rotate object in the large display condition e.g. slower processing and more time use.

Since female show activation additionally in inferior frontal areas (where executive functions are), which men don't, it might cause them to be more vulnerable than men in situations where variables are manipulated, like in Experiment 1.

## CONCLUSION

A number of factors are associated with performance on MRT and one way of putting it, as Moè and Pazzaglia did (Moè & Pazzaglia, 2010), is to conclude that Biology can explain the baseline in MRT for men and women, but cultural and social factors can explain the enhancement that a belief about self can have in MRT performance. The present study provides additional evidence for an explanation of the sex difference in mental rotation based on the choice of strategy for the task. That is, the favored male strategy, e.g. holistic processing, profits from a smaller or standard (computer) display that is located closer to the body, while the approach favored by females facilitates the MRT in larger displays with a wider field of view, where the observer's body is "immersed" into the visual field (Tan, Gergle, Scupelli, & Pausch, 2006). Large displays favor the female approach to such a degree that they can significantly outperform males, despite the traditionally observed male superiority in MRT.

Women might also have a tendency to feel intimidated in tasks like MRT and might engage in more self evaluative processes than men. Men, however, are more susceptible to task complexity than females. This study adds to the evidence that males *are* affected by expectations of task complexity. In Experiment 1 men performed better in the inferior condition (i.e. better in the small screen condition when told to expect better performance in the large display condition and vice versa), which could be interpreted as reflecting a fear for not confirming the expectations. Since no instructions or expectations regarding the self was "administered" in this experiment it is interesting to note that females were affected by expectations *not* related to the self in a *novel* situation, where



they did not have experience or knowledge of the object of study (the large, tiled display).

On the basis of identifying a female advantage with regards to speed in the MRT task in very large displays with wider field of view, the finding of significant effects of expectations for superior condition in the second experiment indicates that those participants who find the small or standard screen more comfortable may well be using holistic processing, since this condition can yield superior MRT compared to a large display.

#### *Limitations of the study*

The main weakness of the present study is the new technology used, with the delayed embedded in the large display condition. Furthermore, in light of the results of the Expectation variable in Experiment 2 we should have provided participants with the option of having no-preferred (or expected) screen size condition (i.e. no difference in experienced task difficulty). This is important in light of the hypothesis that this “expectation” variable identifies MRT strategy (holistic vs analytic), and that the group using analytic strategy might have randomly chosen screen size with expected smaller RTs.

#### *Further research*

For further research one might consider to study the effects of screen size using the alternative Vandenberg and Kuse stimuli, at the same time ask the participant which strategy they use in both conditions. Another factor worth mentioning is the Interstimulus Interval (ISI), where effects should be investigated for the MRT, looking for a potential sex difference. Even though it seems like an unlikely candidate to explain

the sex differences found in this study (female superiority in the large display condition), it could have value to definitely rule it in – or out.

## REFERENCES

- Alexander, G. M., & Evardone, M. (2008). Blocs and bodies: Sex differences in a novel version of the Mental Rotation Test. *Hormones and Behavior*, *53*, 177-184.
- Birenbaum, M., Kelly, A. E., & Levi-Keren, M. (1994). Stimulus Features and Sex Differences in Mental Rotation Test Performance. *Intelligence*, *19*, 51-64.
- Cherney, I. D. (2001). *Gender-linked differences in the development of incidental and intentional memory for static and dynamic stimuli*.
- Cherney, I. D., & Voyer, D. (2010). Development of a Spatial Activity Questionnaire I: Items Identification. *Sex Roles*, *62*, 89-99.
- Czerwinski, M. (2003). *Toward Characterizing the Productivity Benefits of Very Large Displays*. Paper presented at the Zurich, Switzerland, Zurich, Switzerland
- Czerwinski, M., Tan, D. S., & Robertson, G. G. (2002). *Women take a wider view*. Paper presented at the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves, Minneapolis, Minnesota, USA
- Delgado, A. R., & Prieto, G. (2004). Cognitive mediators and sex-related differences in mathematics. *Intelligence*, *32*, 25-32.
- Eriksen, C., & Hoffman, J. (1972). Temporal and spatial characteristics of selective encoding from visual displays. *Perception & Psychophysics*, *12*, 201.
- Gomes, H., Barrett, S., Duff, M., Barnhardt, J., & Ritter, W. (2008). The effects of interstimulus interval on event-related indices of attention: An auditory selective attention test of perceptual load theory. *Clinical Neurophysiology*, *119*(3), 542-555.
- Hedges, L., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science*, *269*(5220), 41-45.
- Heil, M., Rauch, M., & Hennighausen, E. (1998). Response preparation begins before mental rotation is finished: Evidence from event-related brain potentials. *Acta Psychologica*, *99*, 217-232.
- Hirnstain, M., Bayer, U., & Hausmann, M. (2009). Sex-specific response strategies in mental rotation. *Learning and Individual Differences*, *19*, 225-228.
- Jansen-Osmann, P., & Heil, M. (2007). Suitable stimuli to obtain (no) gender

- differences in the speed of cognitive processes involved in mental rotation. *Brain Cogn.*, 64(3), 217-227.
- Jensen, S. T. (2006). Systems Research on High Performance Distributed Systems. Retrieved March, 2010, from <http://www.cs.uit.no/forskning/DOS/hpdc/>
- Kerns, K. A., & Berenbaum, S. A. (1991). Sex Differences in Spatial Ability in Children. *Behavior Genetics*, 21(4).
- Koscik, T., O'Leary, D., J.Moser, D., Andreasen, N. C., & Nopoulos, P. (2009). Sex differences in parietal lobe morphology: Relationship to mental rotation performance. *Brain and Cognition*, 69, 451-459.
- Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: a meta-analysis. *Child Dev.*
- Moè, A. (2009). Are male allways better then females in mental rotation? Exploring a gender belief explanation. *Learing and Individual Differences*, 19, 21-27.
- Moè, A., & Pazzaglia, F. (2010). Beyond Genetics in mental rotation test performance: The power of effort attribution. *Learing and Individual Differences*.
- Moore, D., & Johnson, S. (2008). Mental Rotation in Human Infants: A Sex Difference. . *Psychological Science*, 19(11), 1063-1066.
- Parsons, T. D., Larson, P., Kratz, K., Thieboux, M., Bluestein, B., Buckwalter, J. G., et al. (2004). Sex differences in mental rotation and spatial rotation in a virtual environment. *Neuropsychologia*, 42, 555-562.
- Peters, M. (2005). Sex differences and the factor of time in solving Vandenberg and Kuse mental rotation problems. *Brain and Cognition*, 57(2), 176-184.
- Peters, M. M., Laeng, B. B., Latham, K. K., Jackson, M. M., Zaiyouna, R. R., & Richardson, C. C. (1995). A redrawn Vandenberg and Kuse mental rotations test: different versions and factors that affect performance. *Brain and Cognition*, 28(1), 39-58.
- Quaiser-Pohl, C., & Lehmann, W. (2002). Girls' spatial abilities: Charting the contributions of experiences and attitudes in different academic groups. *British Journal of Educational Psychology*, 72, 245-260.
- Quinn, P., & Liben, L. (2008). A sex difference in mental rotation in young infants. *Psychological Science*, 19(11), 1067-1070.

- Sanders, B., & Soares, M. P. (1986). Sexual Maturation and Spatial Ability in College Students. *Developmental Psychology*, 22(2), 199-203.
- Scali, R. M., Brownlow, S., & Hicks, J. L. (2000). Gender Differences in Spatial Task Performance as a Function of Speed or Accuracy Orientation. *Sex Roles*, 43(5/6), 359-376.
- Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171(3972), 701.
- Suzuki, K., & Nakata, Y. (1988). Does the size of figures affect the rate of mental rotation? *Perception & Psychophysics*, 44(1), 76-80.
- Tan, D. S., Czerwinski, M., & Robertson, G. G. (2003). *Women go with the (optical) flow*. Paper presented at the CHI '03, Fort Lauderdale, Florida, USA
- Tan, D. S., Gergle, D., Scupelli, P., & Pausch, R. (2006). Physically large displays improve performance on spatial tasks. *ACM Transactions on Computer-Human Interaction*, 13(1), 71-99.
- Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Percept Mot Skills*.
- Voyer, D., & Bryden, M. P. (1990). Gender, Level of Spatial Ability, and lateralization of Mental Rotation. *Brain and Cognition*, 13, 18-29.
- Voyer, D., & Hou, J. (2006). Type of items and the magnitude of gender differences on the mental rotations test. *Canadian Journal of Experimental Psychology*.
- Voyer, D., & Saunders, K. A. (2004). Gender differences on the mental rotations test: a factor analysis. *Acta Psychologica*, 117, 79-94.
- 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(2), 250-270.
- Wraga, M., Duncan, L., Jacobs, E. C., Helt, M., & Church, J. (2006). Stereotype susceptibility narrows the gender gap in imagined self-rotation performance. *Psychonomic Bulletin & Review*, 13(5), 813-819.

## APPENDICES

Appendix A. Documents relating to Experiment 1

Appendix B. Documents relating to Experiment 2

## A.1 Hypothesis stating that small displays are better for MR

Below we present the information that was given to participants explaining that they should expect that the small display should provide superior performance over the large display. The text reads “You are participating in an experiment regarding rotation and comparison of three-dimensional (3D) objects and the relationship between how well people perform this task and the size of the screen. Recent research shows that people perform this task better when using small screens. In this experiment we want to test whether this hypothesis holds truth. We also measure the distance between the pupils of the participants to see whether this makes any difference.”

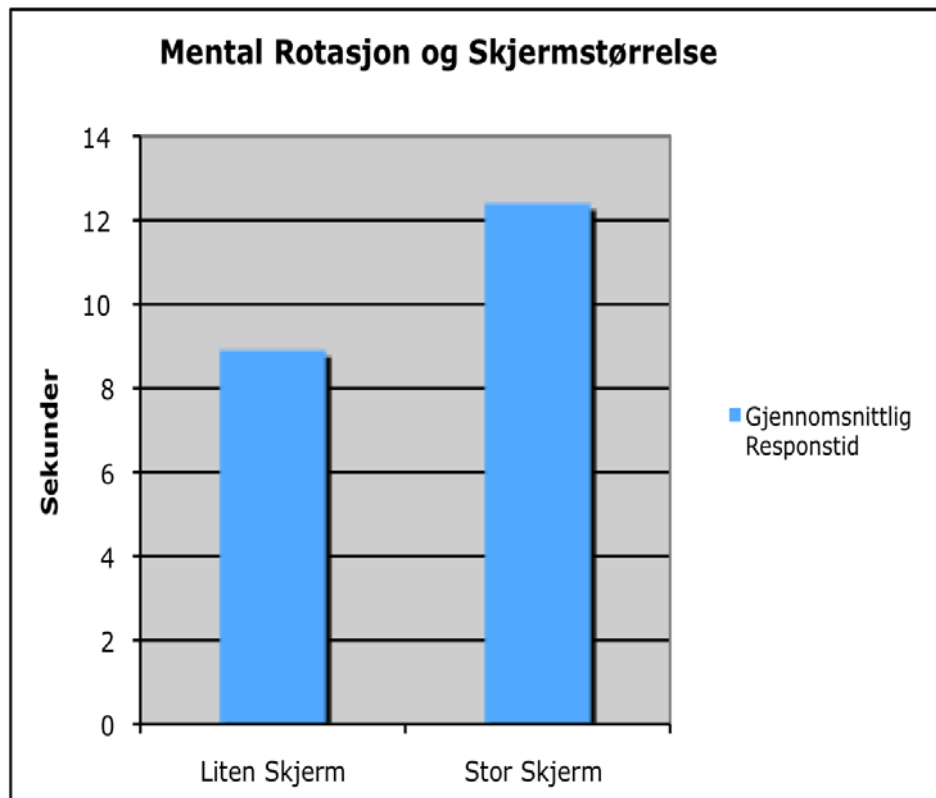
The figure that follows the text illustrates how mean response-times are shorter with a small display compared to a large.

### *Mental Rotasjon Eksperiment*

Du er deltaker i et forsøk som omhandler rotasjon og sammenligning av tredimensjonale (3D) objekter og sammenhengen mellom hvor flinke vi er til å gjøre dette og størrelse på skjermen som viser objektene.

*Ny forskning viser at vi mennesker blir bedre til å utføre denne oppgaven ved å bruke små skjermer. I dette eksperimentet vil vi undersøke om denne hypotesen stemmer eller ikke.*

Vi måler også avstanden mellom pupillene for å videre undersøke om denne har en sammenheng med resultatene.



*Figure 14 Mental Rotation hypothesis: positive towards small displays*



## A.2 Hypothesis stating that large displays are better for MR

### *Hypothesis stating that large displays are better for MR*

Below follows the opposite hypothesis presented to the participants – that large displays give better performance than smaller. The text reads “You are participating in an experiment regarding rotation and comparison of three-dimensional (3D) objects and the relationship between how well people perform this task and the size of the screen. Recent research shows that people perform this task better when using large screens. In this experiment we want to test whether this hypothesis holds truth. We also measure the distance between the pupils of the participants to see whether this makes any difference.”

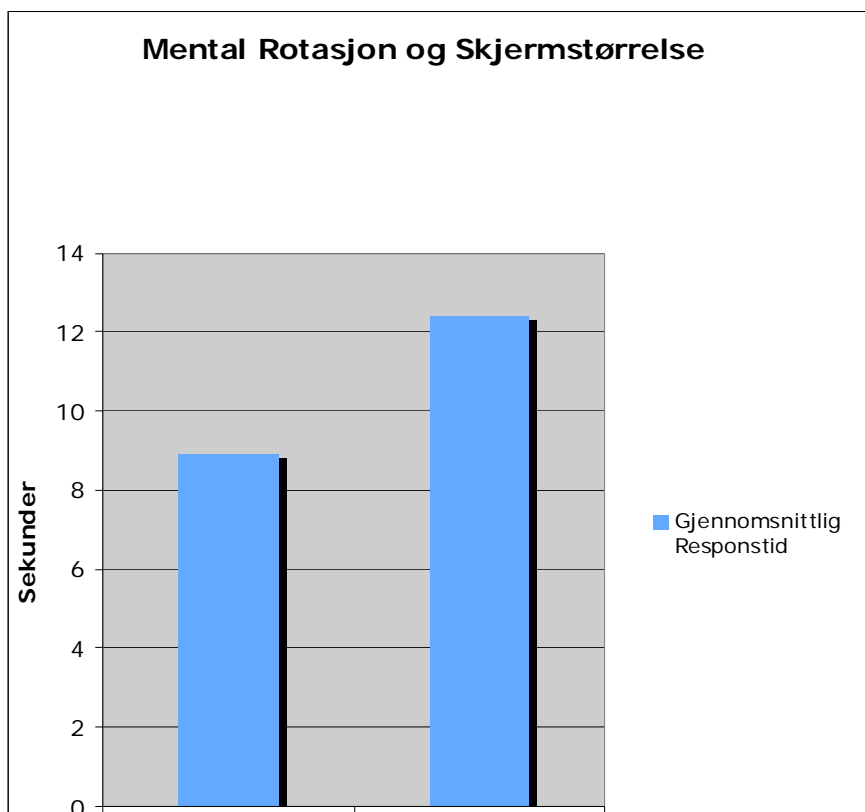
The figure that follows illustrates how response-time shrinks with larger displays.

### *Mental Rotasjon Eksperiment*

Du er deltaker i et forsøk som omhandler rotasjon og sammenligning av tredimensjonale (3D) objekter og sammenhengen mellom hvor flinke vi er til å gjøre dette og størrelse på skjermen som viser objektene.

*Ny forskning viser at vi mennesker blir bedre til å utføre denne oppgaven ved å bruke store skjermer. I dette eksperimentet vil vi undersøke om denne hypotesen stemmer eller ikke.*

Vi måler også avstanden mellom pupillene for å videre undersøke om denne



har en sammenheng med resultatene.

### A.3 Questionnaire for Experiment 1

## Mental rotasjon og skjerm størrelse

deltaker

nr: \_\_\_\_\_

Kjønn: mann    Kvinne

Alder:    år

Pupilleavstand:    mm

Utdanning: 9 årig    Vidreg    Høyskole:    Universitet:

Bransje/yrke:

Spilt dataspill siden:

### Appendix A.4 Tables from ANOVA for Experiment 1

#### ANOVA Table for RTs

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Hypothesis	1	34299642,214	34299642,214	,293	,5917	,293	,081
Sex	1	42535321,682	42535321,682	,363	,5505	,363	,088
Hypothesis * Sex	1	2505287,146	2505287,146	,021	,8845	,021	,052
Subject(Group)	36	4215075859,686	117085440,547				
Screen	1	281153561,425	281153561,425	10,405	,0027	10,405	,898
Screen * Hypothesis	1	1831863,112	1831863,112	,068	,7961	,068	,057
Screen * Sex	1	36113638,889	36113638,889	1,337	,2553	1,337	,192
Screen * Hypothesis * Sex	1	112049253,379	112049253,379	4,147	,0491	4,147	,497
Screen * Subject(Group)	36	972711003,394	27019750,094				
Angle	6	833498522,967	138916420,494	80,013	<,0001	480,077	1,000
Angle * Hypothesis	6	20779190,172	3463198,362	1,995	,0676	11,968	,721
Angle * Sex	6	12528032,552	2088005,425	1,203	,3060	7,216	,462
Angle * Hypothesis * Sex	6	4833909,669	805651,612	,464	,8345	2,784	,185
Angle * Subject(Group)	216	375013953,048	1736175,709				
Screen * Angle	6	2274426,682	379071,114	,415	,8689	2,488	,169
Screen * Angle * Hypothesis	6	7623514,360	1270585,727	1,390	,2199	8,338	,530
Screen * Angle * Sex	6	11210650,050	1868441,675	2,044	,0612	12,261	,734
Screen * Angle * Hypothesis * Sex	6	4357524,010	726254,002	,794	,5753	4,766	,306
Screen * Angle * Subject(Group)	216	197491075,629	914310,535				

*B.1 Questionnaire for Experiment 2*

*Mental rotation and screen size.*

Round 2

Deltaker nr:

Gender :  
Age :  
Pupil distance (eyes) :  
Education : Elementary High-school  
College/University  
Profession/classes :

1. Do you play computer games?
  - a. Yes
  - b. No
  
2. If yes: do you play computer games that involve orienting/moving in 3-dimensional spaces (airplane simulators, World of Warcraft, etc.) – or other 3-D games?
  - a. Games like World of Warcraft (moving in space)
  - b. Games like Tetris (moving/turning/manipulating objects in space)
  - c. Other types of games (examples)
    - i. \_\_\_\_\_
    - ii. \_\_\_\_\_
    - iii. \_\_\_\_\_
  
3. Do you, in everyday life, perform tasks that remind you of the task you did in this experiment?
  - a. Daily
  - b. Weekly
  - c. Rarely
  - d. Never
  
4. Which screen do you think you perform the best on when it comes to number of right rotations and least time use on?
  - a. Small
  - b. Large
  
5. Which screen was more comfortable to do the task on?
  - a. Large
  - b. Small

6. If you would do the same experiment all over again, which screen would you prefer?
  - a. Large
  - b. Small
  
7. If you had to cooperate with someone in other mental rotation tasks, which screen would you prefer then?
  - a. Small screen
  - b. Large screen
  
8. Did you notice that the Large Screen is “Tiled”
  - a. Yes
  - b. No
  
9. If you noticed: did this interfere, for instance when you were solving the tasks in the experiment?
  - a. Yes
  - b. No
  
10. Do you normally use glasses or do you have any vision-impairments, which are not corrected with glasses or contact lenses (answer yes only if you have not used glasses or contact lenses during the experiment)?
  - a. Yes
  - b. No

Appendix B.2 Tables from ANOVA for Experiment 2

Multivariate Test Experiment 2

Effect	Value	F	Hypothesis df	Sig.	Partial Eta Squared	Observed Power <sup>b</sup>	
ScreenSize	Pillai's Trace	.045	2.512 <sup>a</sup>	1.000	.119	.045	.343
	Wilks' Lambda	.955	2.512 <sup>a</sup>	1.000	.119	.045	.343
	Hotelling's Trace	.047	2.512 <sup>a</sup>	1.000	.119	.045	.343
	Roy's Largest Root	.047	2.512 <sup>a</sup>	1.000	.119	.045	.343
ScreenSize * Sex	Pillai's Trace	.100	5.883 <sup>a</sup>	1.000	.019	.100	.663
	Wilks' Lambda	.900	5.883 <sup>a</sup>	1.000	.019	.100	.663
	Hotelling's Trace	.111	5.883 <sup>a</sup>	1.000	.019	.100	.663
	Roy's Largest Root	.111	5.883 <sup>a</sup>	1.000	.019	.100	.663
ScreenSize * Display_size_expect_best_performance	Pillai's Trace	.157	9.893 <sup>a</sup>	1.000	.003	.157	.870
	Wilks' Lambda	.843	9.893 <sup>a</sup>	1.000	.003	.157	.870
	Hotelling's Trace	.187	9.893 <sup>a</sup>	1.000	.003	.157	.870
	Roy's Largest Root	.187	9.893 <sup>a</sup>	1.000	.003	.157	.870
ScreenSize * Sex * Display_size_expect_best_performance	Pillai's Trace	.009	.484 <sup>a</sup>	1.000	.490	.009	.105
	Wilks' Lambda	.991	.484 <sup>a</sup>	1.000	.490	.009	.105
	Hotelling's Trace	.009	.484 <sup>a</sup>	1.000	.490	.009	.105
	Roy's Largest Root	.009	.484 <sup>a</sup>	1.000	.490	.009	.105
Angle	Pillai's Trace	.744	36.281 <sup>a</sup>	4.000	.000	.744	1.000
	Wilks' Lambda	.256	36.281 <sup>a</sup>	4.000	.000	.744	1.000
	Hotelling's Trace	2.903	36.281 <sup>a</sup>	4.000	.000	.744	1.000
	Roy's Largest Root	2.903	36.281 <sup>a</sup>	4.000	.000	.744	1.000
Angle * Sex	Pillai's Trace	.104	1.456 <sup>a</sup>	4.000	.230	.104	.419
	Wilks' Lambda	.896	1.456 <sup>a</sup>	4.000	.230	.104	.419
	Hotelling's Trace	.116	1.456 <sup>a</sup>	4.000	.230	.104	.419
	Roy's Largest Root	.116	1.456 <sup>a</sup>	4.000	.230	.104	.419
Angle * Display_size_expect_best_performance	Pillai's Trace	.188	2.890 <sup>a</sup>	4.000	.031	.188	.741
	Wilks' Lambda	.812	2.890 <sup>a</sup>	4.000	.031	.188	.741
	Hotelling's Trace	.231	2.890 <sup>a</sup>	4.000	.031	.188	.741
	Roy's Largest Root	.231	2.890 <sup>a</sup>	4.000	.031	.188	.741

Angle * Sex *	Pillai's Trace	.068	.913 <sup>a</sup>	4.000	.464	.068	.269
Display_size_expect_best_performance	Wilks' Lambda	.932	.913 <sup>a</sup>	4.000	.464	.068	.269
	Hotelling's Trace	.073	.913 <sup>a</sup>	4.000	.464	.068	.269
	Roy's Largest Root	.073	.913 <sup>a</sup>	4.000	.464	.068	.269
	Pillai's Trace	.051	.671 <sup>a</sup>	4.000	.616	.051	.203
ScreenSize * Angle	Wilks' Lambda	.949	.671 <sup>a</sup>	4.000	.616	.051	.203
	Hotelling's Trace	.054	.671 <sup>a</sup>	4.000	.616	.051	.203
	Roy's Largest Root	.054	.671 <sup>a</sup>	4.000	.616	.051	.203
	Pillai's Trace	.125	1.778 <sup>a</sup>	4.000	.148	.125	.504
ScreenSize * Angle * Sex	Wilks' Lambda	.875	1.778 <sup>a</sup>	4.000	.148	.125	.504
	Hotelling's Trace	.142	1.778 <sup>a</sup>	4.000	.148	.125	.504
	Roy's Largest Root	.142	1.778 <sup>a</sup>	4.000	.148	.125	.504
	Pillai's Trace	.063	.835 <sup>a</sup>	4.000	.510	.063	.247
Display_size_expect_best_performance	Wilks' Lambda	.937	.835 <sup>a</sup>	4.000	.510	.063	.247
	Hotelling's Trace	.067	.835 <sup>a</sup>	4.000	.510	.063	.247
	Roy's Largest Root	.067	.835 <sup>a</sup>	4.000	.510	.063	.247
	Pillai's Trace	.100	1.383 <sup>a</sup>	4.000	.253	.100	.399
ScreenSize * Angle * Sex *	Wilks' Lambda	.900	1.383 <sup>a</sup>	4.000	.253	.100	.399
	Hotelling's Trace	.111	1.383 <sup>a</sup>	4.000	.253	.100	.399
	Roy's Largest Root	.111	1.383 <sup>a</sup>	4.000	.253	.100	.399
	Display_size_expect_best_performance						

a. Exact statistic

b. Computed using alpha = .05

**Number of participants expecting superior display condition for both genders**

		Sex	
		kvinne	mann
Display_size_expect_best_performance	liten	19	19
	stor	9	10

Measure: MEASURE\_1

Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power <sup>a</sup>
--------	-------------------------	----	-------------	---	------	---------------------	-----------------------------

Intercept	1.086E10	1	1.086E10	620.857	.000	.921	1.000
Display_size_expect_best_performance	2.211E8	1	2.211E8	12.643	.001	.193	.937
Display_size_expect_best_performance * Sex	6724891.145	1	6724891.145	.385	.538	.007	.093
Error	9.269E8	53	1.749E7				

a. Computed using alpha = .05