Empirical Study of Non-Reversing Magic Mirrors for Augmented Reality Anatomy Learning

Felix Bork*
Technische Universität München
Munich, Germany

Roghayeh Barmaki†
Johns Hopkins University,
Baltimore MD, United States

Ulrich Eck‡
Technische Universität München
Munich, Germany

Kevin Yu§
Technische Universität München
Munich, Germany

Christian Sandor¶
Nara Institute of Technology
Nara, Japan

Nassir Navab‖
Technische Universität München, Munich Germany
Johns Hopkins University, Baltimore MD, United States

(a) **(b)**

Figure 1: *"Spot the difference!"* While traditional Augmented Reality Magic Mirrors (a) naturally reverse left and right, a non-reversing Magic Mirror (b) reveals the *true* digital mirror image to its observer. Aspiring medical professionals are not used to seeing mirrored anatomy. Left and right are always defined with respect to the patients point of view. Therefore, we study the perceptual benefits of a non-reversing Magic Mirror for the purpose of anatomy learning. In the above Figure, the participant decides whether the superimposed virtual stomach is located anatomically correct or not by means of a presenter device held in his *left* hand. In the reversing Magic Mirror design (a), left and right are reversed and the digital mirror image shown on the TV screen is holding the presenter in the *right* hand. For the non-reversing Magic Mirror case (b), the presenter is shown in the *left* hand of the digital mirror image.

Abstract

Left-right confusion occurs across the entire population and refers to an impeded ability to distinguish between left and right. In medicine this phenomenon is particularly relevant as left and right are always defined with respect to the patient's point of view, i.e. the doctor's right is the patient's left. Traditional anatomy learning resources such as illustrations in textbooks naturally consider this by consistently depicting the anatomy of a patient as seen by an observer standing in front. Augmented Reality Magic Mirrors (MM) are one example of novel anatomy teaching resources and show a user's digital mirror image augmented with virtual anatomy on a large display. As left and right appear to be reversed in such MM setups, similar to real-world physical mirrors, intriguing perceptual questions arise: is a non-reversing MM (NRMM) the more natural choice for the task of anatomy learning and do users even learn anatomy the wrong way with a traditional, reversing MM (RMM)? In this paper, we explore the perceptual differences between an NRMM and RMM design and present the first empirical study comparing these two concepts for the purpose of anatomy learning. Experimental results demonstrate that medical students perform significantly better at identifying anatomically correct placement of virtual organs in an NRMM. However, interaction was significantly more difficult compared to an RMM. We explore the underlying psychological effects and discuss the implications of using an NRMM on user perception, knowledge transfer, and interaction. This study is relevant for the design of future MM systems in the medical domain and lessons-learned can be transferred to other application domains.

Index Terms: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, augmented, and virtual realities; H.5.2 [Information Interfaces and Presentation]: User Interfaces—Ergonomics;

1 Introduction

"No, I mean the other left...". For the majority of people, it is second nature to correctly discriminate between left and right. However, a substantial amount of the population is not blessed with this

*e-mail:felix.bork@tum.de
†e-mail:rl@jhu.edu
‡e-mail:ulrich.eck@tum.de
§e-mail:kevin.yu@tum.de
¶e-mail:chris.sandor@gmail.com
‖e-mail:nassir.navab@tum.de

presumably innate skill [41, 48, 55]. Multiple neuro-psychological factors including visuospatial processing, memory, and sensory information [41] as well as hemispheric asymmetry of the brain [24] appear to be involved in the discrimination process. While left-right confusion predominantly has negligible impact on our daily lives such as a delayed arrival after taking wrong turns with a car, laterality errors in the medical domain have the potential to cause patient's harm and lead to disastrous outcomes, e.g. in case of wrong-sited interventions or wrong-sided diagnosis and therapy [38, 40, 46]. One particular factor contributing to left-right confusion in medicine is that left and right are always defined with respect to the patient's point of view, such that the doctor's right is the patient's left. Whether studying from anatomy textbooks, conducting a patient examination, or executing a surgery, medical professionals always perform a mental rotation by exchanging their own left and right with the patient's left and right. Gormley et al. studied the left-right discrimination ability of medical students and measured the greatest difficulty among female students and those students aiming to become general practitioners or psychiatrists [19]. Other studies by Thomas et al. and McKinley et al. found that left-right confusion was higher in cases of additional mental rotation [51] and various forms of distraction [37]. All of these studies argue that medical professionals should be trained to be aware of left-right confusion from undergraduate education level.

With the recent advances in Mixed Reality (MR), novel systems have been developed for the purpose of complementing existing anatomy learning resources. One category of such systems are Augmented Reality (AR) anatomy Magic Mirrors (MM), which mimic real-world physical mirrors by superimposing virtual information about the anatomy on top of the user's digital mirror image on a large display. However, due to the employed mirror paradigm, such anatomy MM systems present an unfamiliar view to their users: a virtually augmented gallbladder for example, located on the right side of the human abdomen, is displayed by a MM on the left side of a user's digital mirror image. In images produced by a *non-reversing mirror*, this apparent reversal does not occur: a non-reversing MM (NRMM) would thus display the gallbladder from our previous example more intuitively on the right side of a user's digital mirror image, cf. Figure 1.

In this paper, we explore the use of a NRMM setup in the context of anatomy learning and study whether such a design yields perceptual benefits over a traditional reversing Magic Mirror (RMM). To the best of our knowledge, no NRMM anatomy learning system has been published before. Based on a previously proposed preliminary pilot study [7], we conducted a user study during which we asked medical students to identify correct and incorrect placement of virtually augmented organs both for an RMM and NRMM condition. Participants achieved significantly higher percentages of correct answers for the NRMM conditions. Interestingly, we also found significant performance differences depending on the seniority level of students. By implicitly introducing participants to both MM concepts by means of an interaction game instead of a visual and verbal explanation as done during our pilot study [7], we avoided misunderstanding amongst participants which resulted in higher correct organ identification rates. Interaction in NRMM environments remains a challenge as participants were not able to achieve comparable results for the NRMM design during a simple interaction game. Through our experiment, we gained novel insights into how perception in NRMM systems differs compared to RMM designs and which underlying psychological effects could have impacted those.

In section 2, we provide an overview about related work in the fields of general and MR anatomy learning, interactive AR mirrors, and general mirror perception. Section 3 contains a description of our user study, followed by the results (section 4) and a discussion (section 5). Section 6 summarizes the findings of our study and discussed future work.

2 Related Work

2.1 Anatomy Learning

For aspiring medical professionals, gross anatomy courses are the foundation of a solid undergraduate education. Anatomy is still considered to be medicine's most relevant basic discipline during daily clinical activity by both medical students and specialists [1]. Though traditional teaching resources such as anatomy lectures, physical organ models, and text-books are still an integral part of today's curricula, major technological advances in recent years have opened up the door for novel teaching paradigms, even starting to partially replace the established ones. Virtual dissections instead of cadaver ones are just one example of this. Singh and Kharb have described this paradigm shift by moving from passive and teacher-centered to active and student-centered education [47]. Despite ongoing debates about best practices for anatomical knowledge delivery, a recent review by Estai and Bunt calls for the development, integration, and evaluation of multimodal teaching resources, which complement each other to yield the best possible learning experience for students [16]. Animated anatomy videos [2] as well as web-based learning platforms such as Zygote[1] present valuable online resources for this purpose. Such pre-generated content systems require their users to mentally map the anatomy onto their own body, which can be challenging for inherently dynamic processes such as muscular activity. Many systems using Virtual Reality (VR) for medical education have been proposed and proven helpful [12, 14, 17, 26, 33]. Recently, VR avatars as virtual patients and virtual therapists have been a topic of intense research for physical rehabilitation and anatomy education [8, 29, 52]. However, these systems suffer from the same mental mapping issue and reduced user-centrality. Therefore, more successful approaches use AR instead of VR specifically in motor rehabilitation and patient satisfaction [25]. Especially anatomy MM systems, as proposed previously by Ma et al. [34, 35] or Bauer et al. [3] overcome the VR-related limitations by directly mirroring the movements of users in real-time.

2.2 Interactive AR Mirrors

MM systems have been widely used in fashion apparel simulations and virtual clothing. Kim and Cheeyoung [30] presented a fashion coordination prototype that combined user recognition and the augmentation of face styles, make-up, glasses and dress fitting simulations in a mirror-like image representations. For cosmetics and grooming, the Smart Makeup Mirror system [28] was introduced by Iwabuchi and Siio to facilitate and support wearing make up in form of a virtual dressing table. Two other examples were published by Rahman et al. in form of a prototype assisting in the selection of cosmetic products [43], and by Chu et al. who presented an advanced MM jewelry shopping tool [10]. Another application area for MM's are intelligent fitting rooms [57], superimposing virtual garments onto the user, e.g. shirts [23] or virtual shoes [15]. Among all reported cosmetic and clothing Magic Mirrors, only the Smart Makeup Mirror system [28] implemented both a traditional Magic Mirror and a non-reversing one as two distinct views of the user. They stated that professional makeup artists always recommend to validate a person's appearance from the viewpoint of another person standing in front using a non-reversing mirror. Another group of AR interactive mirrors are built using semi-transparent displays. Saakes et al. introduced *Mirror Mirror*, a systems for collaboratively designing fashion items using spatial augmented reality [45]. Similar systems have been proposed for midair gestures [42] and combined interaction spaces [36].

[1] www.zygotebody.com

2.3 Mirrors: Perceptual Issues

According to Bertamini, mirrors are a window into a completely virtual world [4]. Everything we see inside a mirror is completely virtual, which in a sense makes a mirror the perfect VR system. Our brain is tricked into thinking that people or objects we see inside a mirror physically exist. But even though we gaze into mirrors multiple times a day, most of us do not have an in-depth understanding of what exactly happens on the surface of a mirror, a phenomenon known as *illusion of explanatory depth* [31, 44]. Research confirmed that this phenomenon holds true for mirrors as well. People were unable to judge the size of their mirror image's face [5] and the moment their mirror image appears when approaching a covered mirror from the side [32]. Especially the question about *why mirrors reverse left and right, but not up and down?* has been a controversial topic for decades [13, 20, 49, 50]. In mathematical terms, mirrors reverse across the axis perpendicular to their surface, such that front and back are reversed, similar to a glove being turned inside out. However, Ittelson et al. showed that the reversal is perceived across the axis of greatest perceived symmetry [27]. Due to the bilateral symmetry of the human body, this axis coincided with the left-right axis. People tend to believe that their mirror image is formed by a rotation around the vertical (up-down) axis, i.e. by walking around the mirror to become the virtual self [4, 6]. This, however, resembles exactly the image produced by a non-reversing mirror. In non-reversing mirrors, lifting your right hand corresponds to your mirror image also lifting the right hand. Such non-reversing mirrors can be built physically by placing two mirrors perpendicular to each other to form two sides of an equilateral triangle [53], or digitally in an AR application by rearranging the columns of a digital camera image from left to right.

3 USER STUDY

To investigate the potential of an NRMM system and the perceptual benefits such a design could provide over a traditional RMM, we implemented both of these visualizations in an AR anatomy learning demo application, enabling the augmentation of 3-dimensional organ models on top of the user standing in front of the system. We designed a user study to compare the performance of medical students in identifying correct placement of virtual anatomical structures in these two setups. Following a number of pre-tests, participants were introduced to both the NRMM and RMM design by means of a simple interaction game. During the main part of our user study, five different virtual organs were augmented on top of the participants' bodies for both the NRMM and RMM visualizations on either the anatomically correct or opposite side of their mirror images, see Figure 2.

3.1 Experimental Platform

In accordance with all of the previously mentioned AR MM systems, the hardware components of our anatomy teaching application are a video camera and a large, 60 inch display device. For the former, we chose the Microsoft Kinect v2 sensor which combines both an RGB and a depth camera in a single housing. The Kinect was mounted on top of the display device at a height of two meters facing downwards. We positioned participants 150 cm away from the display device during the entire time of the user study. For the purpose of augmenting virtual organs on top of the user's digital mirror image, we employed the Kinect skeleton tracking API. During the experiments, participants were asked to decide whether these organs are displayed on the anatomically correct side of the body or not. This decision process was controlled by means of two buttons on a Logitech R400 presenter. A third button on the bottom was programmed to switch to the next condition during the experiments, such that the entire experimental procedure was controlled by the participant.

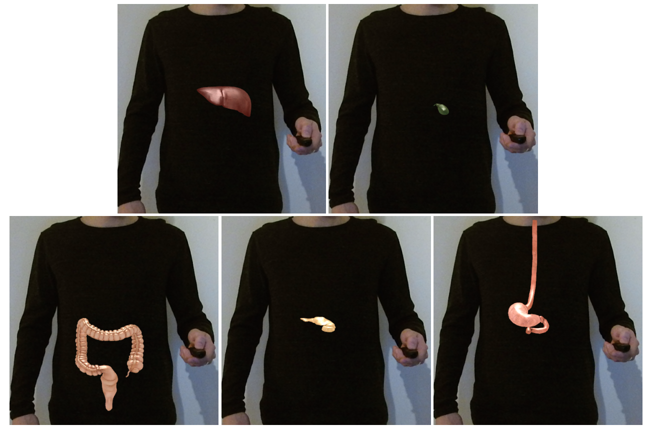

Figure 2: Image sections of screenshots depicting the five different organs virtually augmented during the NRMM vs. RMM organ identification study. From top to bottom, left to right: liver, gallbladder, colon, pancreas, and stomach. The participant held the presenter in his right hand standing in front of the system. As the presenter is in the left hand of the mirror image, all views illustrate anatomically correct RMM overlays.

3.2 Participants

We recruited twenty-five medical students to participate in the user study. A pre-test was used to assess whether they had sufficient anatomy knowledge, see section 3.3.1. Five participants did not manage to pass the test, which resulted in their exclusion from the user study, leaving a total of twenty participants (10M, 10F). We considered half of the participants as juniors and the other half as seniors, with the criterion whether or not they had passed their written preliminary medical exam. The mean age of participants was 26.4 ± 3.1 years and all participants were right-handed.

3.3 Task & Procedure

Our user study consisted of three distinct parts: i) three paper-based pre-tests, ii) interaction game, and iii) the main part of the experiment, the RMM vs. NRMM organ identification study. An overview about all of them is depicted in Figure 3.

3.3.1 Pre-test I: Anatomy Knowledge

During the first pre-test, we asked participants to outline the location of five different organs in an illustration of a frontal view of a patient. The organs of interest corresponded to the ones virtually augmented during the main part of our experiment, namely the *liver*, *gallbladder*, *colon*, *pancreas*, and *stomach*. All of these organs have a distinct laterality inside the human body. Only those participants who were able to correctly outline all five organs were considered for the rest of our experiment.

3.3.2 Pre-test II: Mental Rotation

A second pre-test was used to assess the mental rotation ability of participants. For this task, we presented participants with a total of 10 pairs of 3-dimensional Shepard and Metzler-like block stimuli images proposed by Ganis and Kievit [18]. Each stimuli consisted of 7 to 11 cubes composed of 4 different arms including computer-generated shading and foreshortening depth cues. An example of block-stimuli is shown in Figure 3 a). Participants had to decide in one minute of time, whether the 10 pairs of block-stimuli were the same or mirror images of each other, with the second shape rotated by either 0°, 50°, 100°, or 150° with respect to the first shape.

a) b) c)

Figure 3: Overview of the three steps of the user study: **a)** Depiction of all three paper-based pre-tests. Upper left: anatomy knowledge pre-test. All five different organs with distinct laterality had to be outlined correctly to ensure sufficient anatomy knowledge (liver *(red)*, stomach *(yellow)*, pancreas *(white)*, gallbladder *(green)*, and appendix *(orange)*); Bottom: exemplary block stimuli used during the mental rotation pre-test; Upper right: hand stimuli with distinct pointing for left-right discrimination pre-test. **b)** Participant during interaction game. The dominant hand is tracked and marked with a non-solid green circle. Virtual green circles *falling* from the top have to be caught, while red ones should be avoided. In this case, the participant's dominant hand is the right hand. As the mirror image is also lifting the right hand, a NRMM condition is depicted in this Figure. **c)** Comparison of the four different conditions during the RMM vs. NRMM organ identification study. Upper left) NRMM-NF; Upper right) NRMM-F; Bottom left) RMM-NF; Bottom right) RMM-F. The participant is lifting the right hand and chooses by means of a pointer device whether the presented augmentation is anatomically correct, which is the case for the two *non-flipped (NF)* conditions in the left column. To depict the difference between NRMM and RMM conditions, the former are surrounded by a window frame, as an NRMM can be thought of as a see-through window. RMM conditions are surrounded using a mirror frame. Note: these surrounding frames were not shown to participants during the user study.

3.3.3 Pre-test III: Left-Right Discrimination

With our third and last pretest, we compared the left-right discrimination ability of participants. Similar to a study presented by Brandt and Mackavay [9], stimuli of left and right hands were shown during the procedure [11]. In the first part of this pretest, we used a 5×5 grid of hands with the index finger always pointing either up or down. Participants were asked to read out loud in which direction the corresponding hands point as accurately and fast as possible. For the second part, an identical task had to be performed, this time with all hands pointing either to the left or right side. The difference between these two measured task-completion-times was used as a metric for participants left-right discrimination ability.

3.3.4 Interaction game

In order to make participants familiar with both the RMM and NRMM design used during the main part of the user study and to compare interaction performance for both conditions, we developed a simple AR interaction game. Green and red colored circles were augmented on top of the RGB video stream of the Kinect V2, *falling* from top to bottom. Another green circle was displayed at the screen coordinates of the user's dominant hand. The task was to *catch* all green circles and to avoid the red ones. The maximum number of green circles was 20. Two parameters were varied during the experiment: Time-to-next-circle decreased after every 5 green circles by $0.5\,s$, starting from $3.0\,s$. Similarly, the time it took for circles to fall down was decreased by the same amount after every 5 green circles, starting from $3.5\,s$. We recorded both the number of green and red circles caught. The experiment was run in two passes: in the first pass, participants executed the experiment for a regular RMM visualization. Subsequently, we switched to an NRMM design and the task was repeated. Figure 3 b shows an image section of a screenshot during the interaction game. Following this experiment, participants knew the difference between an RMM and NRMM design, which were subject to investigation during the following organ identification study.

3.3.5 RMM vs. NRMM organ identification study

During the last and most important part of our user study, we investigated the ability of participants to identify correct placement of virtually augmented organs for both an RMM and NRMM setup. During the course of the experiment, five virtual organs *(liver, gallbladder, colon, pancreas, and stomach)* were augmented individually onto the participant's digital mirror image, either on the anatomically correct side of the body or on the opposite side. Thus, four different conditions were traversed:

NRMM-NF: Non-Reversing MM, Organ Non-Flipped

NRMM-F: Non-Reversing MM, Organ Flipped

RMM-NF: Reversing MM, Organ Non-Flipped

RMM-F: Reversing MM, Organ Flipped

The two conditions for which the virtual organ was not flipped corresponded to the anatomically correct placements. An AR view of all four different conditions is shown in Figure 3 c. After each time participants provided an answer for a certain condition by means of the hand held presenter, we displayed a black screen and the participant was asked to continue with the next condition by pressing another button on the presenter. Only after this, the camera image became visible again. This design choice was made in order to avoid too obvious switches between the NRMM and RMM conditions. We asked participants to provide an answer as quickly as possible, while prioritizing correct answers at the same time. By not providing the participants with prior training sessions and any performance feedback, we eliminate training effects from our study. The main goal of this part of the user study was to investigate whether an NRMM provides perceptual benefits over the traditional RMM design and whether these in return yield an increased overall rate of correct answers.

3.4 Design

There were three independent variables for the organ identification part of our experiment. The type of virtual organ augmented onto the participant's digital mirror image had five levels, corresponding to the five aforementioned organs of interest shown in Figure 2. Organs could be either displayed on the anatomically correct side of the body (not flipped) or one the opposite one (flipped). The third independent variable signaled which MM design was used: either a traditional RMM or an NRMM. Consequently, our experiment had a $5 \times 2 \times 2$ within-subjects design. We employed a balanced Latin square matrix for randomizing these conditions across the study participants [54].

3.5 Hypotheses

We formulated five hypotheses prior to designing the user study. Those were subject to an extensive statistical evaluation:

- **H 1.** The overall percentage of correctly identified virtual organs is higher for the NRMM conditions in comparison to the regular RMM conditions.
- **H 2.** The average decision time (in seconds) is smaller for the NRMM conditions compared to the RMM conditions.
- **H 3.** The percentage of correctly identified virtual organs for the RMM conditions is higher for less experienced participants.
- **H 4.** Participants with higher mental rotation or left-right discrimination ability perform better for the RMM conditions.
- **H 5.** During the interaction game, the total amount of errors (missed green circles, hit red circles) is significantly higher for the NRMM compared to the RMM.

Furthermore, we expect the vast majority of participants to qualitatively prefer the NRMM conditions and to quickly establish the link between these views and the familiar patient examination and textbook view.

4 Results

In this section, we present a detailed analysis of the results obtained from our user study. Table 1 summarizes the findings for the organ identification study. Overall, participants achieved higher percentages of correct answers for the NRMM conditions compared to both RMM conditions , see Figure 4. Combining both NRMM and RMM conditions yielded a significant difference in achieved percentage of correct answers ($F_{1,78} = 10.8, p < 0.01, \eta^2 = 0.12$).

Similar to the approach pursued in our preliminary pilot study [7], we examined whether the seniority level of participants affects the percentage of correctly identified organs during the experiment. We split all participants into two groups: To the first group, hereafter referred to as the *junior group*, we assigned those medical students who did not yet take the medical preliminary examination. Consequently, all other medical students formed the *senior group*. The two groups were balanced and both contained 10 participants. For correct answer percentages, we observed an interesting difference among juniors and seniors. Juniors performed significantly better for the RMM conditions than seniors ($F_{1,38} = 8.67$, $p < 0.01, \eta^2 = 0.19$), cf. Figure 4.

Participants were slightly faster in the two NRMM conditions compared to the RMM conditions. Table 1 shows the mean decision times for all four conditions. As there was substantial variation in the observations across participants, the difference was not statistically significant as revealed in an analysis of variances ($F_{1,78} = 1.37, ns$).

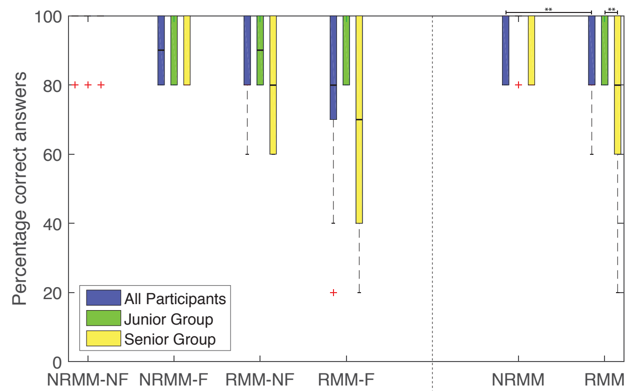

Figure 4: Combined results for organ identification experiment. Participants were asked to identify correct placement of virtual organs in four different MM views displayed on the x-axis. Junior medical students performed significantly better for the RMM conditions compared to the group of experienced medical students. Combined results for both NRMM and RMM conditions are shown on the right.

Comparing the percentage of correct answers for each of the five different organs among all participants revealed the following results: in the NRMM conditions, all five organs had comparably high identification results (liver 100.0%, colon 95%, gallbladder 92.5%, pancreas 92.5%, and stomach 90.0%). Overall, results for the RMM conditions were worse: the stomach was detected correctly in only 60.0% of the cases, followed by the pancreas (82.5%), liver (87.5%), colon (90.0%), and gallbladder (90.0%).

For evaluating the results of our interaction game, we defined the total error count in both conditions as the sum of missed green circles and hit red circles. For the RMM case, we measured a mean error of 1.7 ± 1.26 compared to 5.0 ± 1.59 for the NRMM case. These differences were statistically significant at the $p < 0.01$ level, ($F_{1,38} = 52.92, p < 0.001, \eta^2 = 0.58$).

For the two paper-based pre-tests, we measured a mean percentage of correct answers for the mental rotation test of $85.0 \pm 11.47\,\%$, and an average time difference between up-down and left-right discrimination of $4.36 \pm 2.05\,s$. Correlations between these pretest data and the participants' performance during the interaction game and organ identification study were calculated using Spearmans rank correlation coefficient. Pearsons correlation is sensitive to outliers as well as skewness, hence we used nonparametric metrics to report correlations. A moderate correlation was observed for mental rotation (MR) score and left-right discrimination (LRD) score, with the latter defined as the time difference between up-down and left-right discrimination tasks, see section 3.3.3, ($r_s = -0.46, n = 20, p < .05$). Secondly, we found a moderate correlation between LRD score and RMM error count during the interaction game ($r_s = 0.56, n = 20, p < .05$). However, this was not the case for LRD score and the NRMM error count ($r_s = 0.39, n = 20, ns$). MR score was not strongly correlated to error counts during the interaction game. However, it was slightly related to the total number of mistakes during the organ identification experiment ($r_s = -0.42, n = 20, p = .062, ns$).

Post-experiment interviews were conducted with all participants after the user study. Only two male, junior participants did not express a preference for either the NRMM or RMM view. They stated that both are equally well suited for anatomy learning. The other participants expressed strong preference for the NRMM view, stating that it is the more natural choice for an anatomy learning AR system and resembles a more familiar view of the anatomy.

Table 1: Comparison of the average correct answers and decision times among all 20 participants during the organ identification study for the four individual conditions as well as combined NRMM and RMM conditions. The NRMM conditions provide a higher overall average of correct answers with slightly lower decision times compared to the RMM conditions.

Conditions		**Correct Answers**		**Decision Times**	
		Mean μ	SD σ	Mean μ	SD σ
(NRMM-NF)	Non-Reversing Magic Mirror, Organs Not Flipped	98.00%	6.16%	5.06 s	2.78 s
(NRMM-F)	Non-Reversing Magic Mirror, Organs Flipped	90.00%	10.26%	4.59 s	1.82 s
(RMM-NF)	Reversing Magic Mirror, Organs Not Flipped	85.00%	15.73%	5.32 s	3.53 s
(RMM-F)	Reversing Magic Mirror, Organs Flipped	79.00%	25.53%	5.97 s	3.97 s
(NRMM)	Non-Reversing Magic Mirror Combined	94.00%	5.99%	4.83 s	2.17 s
(RMM)	Reversing Magic Mirror Combined	82.00%	18.24%	5.64 s	3.62 s

5 Discussion

This is the first study which compares a traditional reversing mirror design as generally employed in MM applications to a non-reversing mirror in the context of AR anatomy learning. We investigated the potential such a design has on the learning ability of medical students, underlying psychological effects, and possible implications on user perception and interaction.

During our user study we found that participants in general identified correct organ locations significantly better in NRMM conditions compared to RMM conditions, which confirms our hypothesis **H1**. As expected, participants scored significantly better during the interaction game in the RMM condition, which confirms hypothesis **H5**. These results demonstrate the potential application of the NRMM paradigm for an AR anatomy teaching system, however user interaction with NRMM systems still remains a challenge.

We also defined hypothesis **H3** based on the results of our preliminary pilot study [7], where we found significant differences in RMM performance between junior and senior medical students. The results of this user study verify our previous observations. In fact, senior students more often provide incorrect answers in RMM conditions than junior students, which confirms hypothesis **H3**. A possible explanation to this is the *mere-exposure effect* [56, 39], a psychological phenomenon by which a person develops a strong preference for a certain stimulus through continuous exposure as seen in anatomy learning from textbooks. More experienced medical students have been exposed continuously to the concept of exchanging their own left and right with the patient's left and right, either from textbook illustrations, patient examinations, or even surgical interventions. However, as junior students achieved comparable percentages of correct answers for both NRMM and RMM conditions, is it intriguing to discuss which long-term effects RMM anatomy learning has. One possible outcome is improved left-right discrimination ability due to the continuous exposure to both NRMM and RMM views. However, the exact opposite could also hold true, as medical students could be confused by seeing non-corresponding views in different learning resources (i.e. textbook vs. RMM). An RMM could even introduce the risk of learning incorrect anatomy. Before such systems should be integrated into the medical curriculum, the risk for such negative outcomes should be eliminated.

Decision times between the conditions were not significantly different, which rejects our hypothesis **H2**. We observed that most participants first tried to understand if an RMM or NRMM condition is currently displayed, for example by raising their hands or touching their bodies, and only then continued to reason about the correctness of the augmented organs, which could explain the high variance in the measured decision times.

Although we found moderate correlations between the participants' mental rotation and left-right discrimination scores, we did not find significant correlations between these pre-tests and the results of our main study. Therefore, we did not have enough evidence to support our hypothesis **H4**. However, the weak, non-significant correlation between mental rotation score and overall error count during the organ identification study should be subject to further investigation.

Another interesting perceptual difference between RMM and NRMM designs is user perception. One senior, female participant mentioned that she imagined her digital mirror image for the NRMM conditions to be decoupled from herself. Instead, she imagined it to be a patient standing in front of her. A user-centric, personalized AR overlay is specific to RMM systems. Therefore, it would be interesting to study in future work whether or not users of an NRMM system perceive their digital mirror image as themselves, or as decoupled. Furthermore, the same aspects would be interesting to study in case of a third person observer, e.g. in collaborative anatomy learning sessions.

In future work, we would like to further investigate the potential and risks of NRMM systems for anatomy learning. While NRMM systems seem to be better suited for such tasks it remains unclear how users should interact with them. Users naturally have great difficulty to interact in reversed coordinate frames [22, 21]. One possible approach is to decouple the interaction from the displayed image, for example by displaying a cursor that is controlled as if the system would operate as in RMM mode. This will create a discrepancy between the displayed image of users on the screen and the cursor, however it could allow for efficient user input. Another option would be to switch between NRMM and RMM modes depending on the current task. During interaction, for example the selection of the currently displayed body system, the system would operate in RMM mode and once the system is displayed, it would then switch into NRMM mode. Lastly, we also want to evaluate picture in picture visualization which shows both modes at the same time.

6 Conclusions

In this paper, we have explored the potential of non-reversing Magic Mirror (NRMM) systems in the context of anatomy learning and discussed perceptual benefits such a design could provide over traditional reversing Magic Mirror (RMM) systems. While the latter present a digital mirror image comparable to the one produced by a real-world physical mirror, an NRMM shows the *true* digital mirror image of a person standing in front of the system, such that left and right are not reversed. We conducted a user study comparing both visualizations to each other. Medical students achieved significantly better results for the task of identifying anatomically correct placement of virtual organs in an NRMM. This coincided with participants' qualitative opinions, who found that the NRMM visualization is a more natural fit for the context of AR anatomy learning. However, several perceptual challenges remain to be solved in NRMM environments including interaction and user perception. Developers of AR MM systems should consider the intent of their application and the type of expected users before deciding on the type of mirror visualization for their system. Previously acquired domain knowledge and lateral importance, as in the case of anatomy learning, as well as the mere-exposure effect, can make an NRMM design the better choice for screen-based AR applications.

Acknowledgements

We thank Anna-Maria von der Heide, Severine Habert, and Alexander Keppler for their help during the user study design and execution. Anatomy models were obtained from http://www.plasticboy.co.uk.

References

[1] L.-A. Arráez-Aybar, I. Sánchez-Montesinos, R.-M. Mirapeix, B. Mompeo-Corredera, and J.-R. Sañudo-Tejero. Relevance of human anatomy in daily clinical practice. *Annals of Anatomy-Anatomischer Anzeiger*, 192(6):341–348, 2010.

[2] S. A. Azer. Can youtube help students in learning surface anatomy? *Surgical and radiologic anatomy*, 34(5):465–468, 2012.

[3] A. Bauer, A.-H. Dicko, F. Faure, O. Palombi, and J. Troccaz. Anatomical mirroring. In *the 9th International Conference*, pages 113–122, New York, New York, USA, 2016. ACM Press.

[4] M. Bertamini. Mirrors and the mind. *Psychologist*, 23(2):112, 2010.

[5] M. Bertamini and T. E. Parks. On what people know about images on mirrors. *Cognition*, 98(1):85–104, 2005.

[6] M. Bertamini, A. Spooner, and H. Hecht. Naive optics: Predicting and perceiving reflections in mirrors. *Journal of experimental psychology. Human perception and performance*, 29(5):982–1002, 2003.

[7] F. Bork, R. Barmaki, U. Eck, P. Fallavollita, B. Fuerst, and N. Navab. Exploring non-reversing magic mirrors for screen-based augmented reality systems. In *IEEE Virtual Reality Conference (to appear)*. IEEE, 2017.

[8] M. Botezatu, H. Hult, M. K. Tessma, and U. G. Fors. Virtual patient simulation for learning and assessment: superior results in comparison with regular course exams. *Medical teacher*, 32(10):845–850, 2010.

[9] J. Brandt and W. Mackavey. Left-right confusion and the perception of bilateral symmetry. *International Journal of Neuroscience*, 12(2):87–94, 1981.

[10] M. Chu, B. Dalal, A. Walendowski, and B. Begole. Countertop responsive mirror: supporting physical retail shopping for sellers, buyers and companions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2533–2542. ACM, 2010.

[11] E. Chudler. Right-Left Confusion? Online: https://faculty.washington.edu/chudler/java/hands1.html (accessed: 2017-03-03).

[12] D. Coffey, N. Malbraaten, T. Le, I. Borazjani, F. Sotiropoulos, and D. F. Keefe. Slice WIM. In *Symposium on Interactive 3D Graphics and Games*, pages 191–8, New York, New York, USA, 2011. ACM Press.

[13] C. Corballis. Much ado about mirrors michael. *Psychonomic Bulletin & Review*, 7(1):163–169, 2000.

[14] J. W. V. de Faria, M. J. Teixeira, L. de Moura Sousa Júnior, J. P. Otoch, and E. G. Figueiredo. Virtual and stereoscopic anatomy: when virtual reality meets medical education. *Journal of Neurosurgery*, 125(5):1105–1111, Nov. 2016.

[15] P. Eisert, P. Fechteler, and J. Rurainsky. 3-d tracking of shoes for virtual mirror applications. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–6. IEEE, 2008.

[16] M. Estai and S. Bunt. Best teaching practices in anatomy education: A critical review. *Annals of Anatomy*, 208:151–157, Nov. 2016.

[17] J. Falah, S. Khan, T. Alfalah, S. F. Alfalah, W. Chan, D. K. Harrison, and V. Charissis. Virtual reality medical training system for anatomy education. In *Science and Information Conference (SAI), 2014*, pages 752–758. IEEE, 2014.

[18] G. Ganis and R. Kievit. A New Set of Three-Dimensional Shapes for Investigating Mental Rotation Processes: Validation Data and Stimulus Set. *Journal of Open Psychology Data*, 3(3972):701–7, Mar. 2015.

[19] G. J. Gormley, M. Dempster, and R. Best. Right-left discrimination among medical students: questionnaire and psychometric study. *BMJ*, 337(dec16 1):a2826–a2826, Dec. 2008.

[20] R. L. Gregory. *Mirrors in mind*. Penguin Harmondsworth, UK, 1998.

[21] A. F. Healy, E. L. Wohldmann, and L. E. Bourne Jr. How does practice with a reversed mouse influence subsequent speeded aiming performance? A test of global inhibition. *Journal of Cognitive Psychology*, 23(5):559–573, Aug. 2011.

[22] L. M. Hiatt and R. Simmons. Coordinate frames in robotic teleoperation. In *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, pages 1712–1719. IEEE, 2006.

[23] A. Hilsmann and P. Eisert. Tracking and retexturing cloth for real-time virtual clothing applications. In *International Conference on Computer Vision/Computer Graphics Collaboration Techniques and Applications*, pages 94–105. Springer, 2009.

[24] M. Hirnstein, S. Ocklenburg, D. Schneider, and M. Hausmann. Sex differences in left–right confusion depend on hemispheric asymmetry. *Cortex*, 45(7):891–899, July 2009.

[25] S. Hoermann, L. Hale, S. J. Winser, and H. Regenbrecht. Patient engagement and clinical feasibility of augmented reflection technology for stroke rehabilitation. *International Journal on Disability and Human Development*, 13(3):355–360, 2014.

[26] H.-M. Huang, S.-S. Liaw, and C.-M. Lai. Exploring learner acceptance of the use of virtual reality in medical education: a case study of desktop and projection-based display systems. *Interactive Learning Environments*, 24(1):3–19, 2016.

[27] W. H. Ittelson, L. Mowafy, and D. Magid. The perception of mirror-reflected objects. 20(August 1990):567–584, 1991.

[28] E. Iwabuchi, M. Nakagawa, and I. Siio. Smart makeup mirror: computer-augmented mirror to aid makeup application. In *International Conference on Human-Computer Interaction*, pages 495–503. Springer, 2009.

[29] M. Kallmann, C. Camporesi, and J. Han. Vr-assisted physical rehabilitation: Adapting to the needs of therapists and patients. In *Virtual Realities*, pages 147–168. Springer, 2015.

[30] M. Kim and K. Cheeyong. Augmented reality fashion apparel simulation using a magic mirror. *International Journal of Smart Home*, 9(2):169–178, 2015.

[31] R. Lawson. The science of cycology: Failures to understand how everyday objects work. *Memory & cognition*, 34(8):1667–1675, 2006.

[32] R. Lawson and M. Bertamini. Errors in judging information about reflections in mirrors. *Perception*, 35(9):1265–1288, 2006.

[33] J. Lu, Z. Pan, H. Lin, M. Zhang, and J. Shi. Virtual learning environment for medical education based on vrml and vtk. *Computers & Graphics*, 29(2):283–288, 2005.

[34] M. Ma, P. Fallavollita, I. Seelbach, A. M. Heide, E. Euler, J. Waschke, and N. Navab. Personalized augmented reality for anatomy education. *Clinical Anatomy*, 2015.

[35] M. MA, P. Jutzi, F. Bork, I. Seelbach, A. M. Von der Heide, N. Navab, and P. Fallavollita. Interactive mixed reality for muscle structure and function learning. *Proceedings of the 7th International Conference, MIAR 2016, Bern Switzerland, August 24-26*, pages 117–128, 2016.

[36] D. Martinez Plasencia, F. Berthaut, A. Karnik, and S. Subramanian. Through the combining glass. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*, pages 341–350. ACM, 2014.

[37] J. McKinley, M. Dempster, and G. J. Gormley. 'Sorry, I meant the patient's left side': impact of distraction on left-right discrimination.

Medical Education, 49(4):427–435, Mar. 2015.

[38] K. E. Miller, M. Mims, D. E. Paull, L. Williams, J. Neily, P. D. Mills, C. Z. Lee, and R. R. Hemphill. Wrong-Side Thoracentesis. *JAMA Surgery*, 149(8):774–6, Aug. 2014.

[39] T. H. Mita, M. Dermer, and J. Knight. Reversed facial images and the mere-exposure hypothesis. *Journal of Personality and Social Psychology*, 35(8):597–601, 1977.

[40] NHS. England, London (UK). Never Events reported as occurring between 1 April 2015 and 31 March 2016 - final update. Online: https://improvement.nhs.uk/resources/never-events-data/ (accessed: 2017-03-01).

[41] S. H. Ofte and K. Hugdahl. Rightleft Discrimination in Male and Female, Young and Old Subjects. *Journal of Clinical and Experimental Neuropsychology (Neuropsychology, Development and Cognition: Section A)*, 24(1):82–92, Feb. 2002.

[42] M. Perry, S. Beckett, K. O'Hara, and S. Subramanian. Wavewindow: public, performative gestural interaction. In *Acm international conference on interactive tabletops and surfaces*, pages 109–112. ACM, 2010.

[43] A. M. Rahman, T. T. Tran, S. A. Hossain, and A. El Saddik. Augmented rendering of makeup features in a smart interactive mirror system for decision support in cosmetic products selection. In *Distributed Simulation and Real Time Applications (DS-RT), 2010 IEEE/ACM 14th International Symposium on*, pages 203–206. IEEE, 2010.

[44] L. Rozenblit and F. Keil. The misunderstood limits of folk science: An illusion of explanatory depth. *Cog. science*, 26(5):521–562, 2002.

[45] D. Saakes, H.-S. Yeo, S.-T. Noh, G. Han, and W. Woo. Mirror mirror: An on-body t-shirt design system. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 6058–6063. ACM, 2016.

[46] S. C. Seiden. Wrong-Side/Wrong-Site, Wrong-Procedure, and Wrong-Patient Adverse Events. *Archives of Surgery*, 141(9):931–9, Sept. 2006.

[47] V. Singh and P. Kharb. A paradigm shift from teaching to learning gross anatomy: meta-analysis of implications for instructional methods. *Journal of the Anatomical Society of India*, 62(1):84–89, June 2013.

[48] M. D. STORFER. PROBLEMS IN LEFT-RIGHT DISCRIMINATION IN A HIGH-IQ POPULATION. *Perceptual and Motor Skills*, 81(2):491–497, Oct. 1995.

[49] Y. Takano. Why does a mirror image look left-right reversed? A hypothesis of multiple processes. *Psychonomic Bulletin & Review*, 5(1):37–55, 1998.

[50] Y. Takano. Mirror reversal of slanted objects: A psycho-optic explanation. *Philosophical Psychology*, 28(2):240–259, 2015.

[51] M. S. Thomas, S. Kini, and K. Mala. Factors affecting the right and left discrimination ability among dental students. *Journal of Education and Ethics in Dentistry*, 3(2):66–5, 2013.

[52] J. K. Tworek, H. A. Jamniczky, C. Jacob, B. Hallgrímsson, and B. Wright. The lindsay virtual human project: An immersive approach to anatomy and physiology. *Anatomical sciences education*, 6(1):19–28, 2013.

[53] W. Watson. Non-reversing mirror device with perceived depth perspective, Apr. 8 1986. US Patent 4,580,880.

[54] E. Williams. Experimental designs balanced for the estimation of residual effects of treatments. *Australian Journal of Chemistry*, 2(2):149–168, 1949.

[55] S. M. Wolf. Difficulties in Right-Left Discrimination in a Normal Population. *Archives of Neurology*, 29(2):128–129, Aug. 1973.

[56] R. B. Zajonc. Attitudinal effects of mere exposure. *Journal of Experimental Social Psychology*, 7(4):461–472, 1971.

[57] W. Zhang, T. Matsumoto, J. Liu, M. Chu, and B. Begole. An intelligent fitting room using multi-camera perception. In *Proceedings of the 13th international conference on Intelligent user interfaces*, pages 60–69. ACM, 2008.