This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2021.3096494, IEEE Transactions on Visualization and Computer Graphics

Virtual replicas of real places: Experimental investigations

Richard Skarbez, *Member, IEEE*, Joseph L. Gabbard, *Member, IEEE*, Doug A. Bowman, *Member, IEEE*, J. Todd Ogle, Thomas Tucker

Fig. 1: Side-by-side images of the same room. Left image is a photo of the real room, right image is a computer-generated replica.

Abstract—As virtual reality (VR) technology becomes cheaper, higher-quality, and more widely available, it is seeing increasing use in a variety of applications including cultural heritage, real estate, and architecture. A common goal for all these applications is a compelling virtual recreation of a real place. Despite this, there has been very little research into how users perceive and experience such replicated spaces. This paper reports the results from a series of three user studies investigating this topic. Results include that the scale of the room and large objects in it are most important for users to perceive the room as real and that non-physical behaviors such as objects floating in air are readily noticeable and have a negative effect even when the errors are small in scale.

Index Terms—Virtual reality, Virtual environments, Presence, Psychophysics, User studies.

1 Introduction

In recent years, virtual reality (VR) technology has become increasingly accessible to a wide array of users in various application areas. These applications include cultural heritage [3], real estate, and architecture [18], which share the goal of recreating real places—past, present, or future—in VR.

However, to date, there has been very little research into how users perceive and interact with such replicated spaces. Some important questions in this area include: How accurate do replicated spaces need to be? Are there some elements of replicated spaces for which accuracy is more important than others? Are there elements of a VR space that generally go unnoticed, and thus do not need significant technical investment in terms of modeling, scanning, and/or reconstruction? We report on the design and results of a series of three user studies which had the goal of determining which characteristics of virtual rooms were most important for users to have the same "feeling of reality" as in an identical real physical room.

Note that we have chosen to focus on the subjective experience that a user may have in a replicated space. A pixel-perfect recreation may not be possible or practical; consider a cultural heritage application where creators have to make various assumptions about what a place may have looked like. Even something as "mundane" as constructing a detailed computer model of a room that actually exists is non-trivial; manual modeling requires substantial human skill and effort, and even state-of-the-art scanning and reconstruction techniques are imperfect. In any case, the process of recreating the space *will* introduce errors. That said, it is certainly the case that users will consider some errors more noticeable, distracting, or implausible than others. In this work, we have tried to identify some errors (or classes of errors) that are particularly important to users, so as to enable creators to spend their efforts more productively. In order to do this, we have "parameterized" the virtual replica room, such that changes to various room parameters—such as length, width, amount of clutter, lighting conditions, etc.—can be programmatically varied.

Results from these studies include that the scale of the room and large objects in it are most important for users to perceive the room as real, that non-physical behaviors such as objects floating in air are readily noticeable and have a negative effect even when the errors are small in scale, and that differences in lighting quality seem to have a minimal effect on users' perceptions of a replicated space.

2 Previous Work

"Feeling of reality" can be mapped onto existing constructs, such as Baños et al.'s Reality Judgment [2] or Slater's Plausibility Illusion (Psi) [24]. We adopt Slater's term, which he defines as, "the illusion that what is apparently happening is really happening (even though you know for sure that it is not)." Skarbez, Brooks, and Whitton argue that Plausibility Illusion is a user's subjective experience of an objective construct they call coherence, which is essentially the extent to which a scenario complies with a user's expectations [22]. (Elsewhere, Gilbert uses the term authenticity for essentially the same construct [7].)

If—as in the studies presented in this paper—the virtual environment purports to represent the real world, any behavior not consistent with a user's experiences in the real world would decrease the environment's

- *Richard Skarbez is with La Trobe University. E-mail: r.skarbez@latrobe.edu.au.*
- *Joseph L. Gabbard is with the Grado Department of Industrial and Systems Engineering and Center for Human-Computer Interaction at Virginia Tech. E-mail: jgabbard@vt.edu.*
- *Doug A. Bowman is with the Department of Computer Science and Center for Human-Computer Interaction at Virginia Tech. E-mail: bowman@vt.edu.*
- *J. Todd Ogle is with Applied Research in Immersive Environments and Simulations and Center for Human-Computer Interaction at Virginia Tech. E-mail: jogle@vt.edu.*
- *Thomas Tucker is with the School of Visual Arts and Center for Human-Computer Interaction at Virginia Tech. E-mail: jogle@vt.edu.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2021.3096494, IEEE Transactions on Visualization and Computer Graphics

coherence, which in turn should cause their feeling of Plausibility Illusion to decrease. Some examples include objects floating above the ground (inconsistent with prior experience with gravity), objects interpenetrating one another (inconsistent with prior experience with physical objects), room dimensions being too big or too small (inconsistent with prior experience of the built world), and objects being significantly moved from their positions in the real place (inconsistent with prior experience with the real original room).

There has been little research into the effect of "realism" on user experience in virtual environments, and what has been done to date has primarily focused on its impact on presence. Bouchard et al. demonstrated that participants' belief that the scenario represented the real world, as opposed to being a virtual recreation, resulted in higher presence scores [5]. Hvass et al. describe a significant (but small) effect of geometric realism (polygon count + texture resolution) on physiological and questionnaire-based measures of presence [9]. Slater et al. argue that increased realism (real-time ray tracing as opposed to ray casting) increased stress in a stressful virtual environment, and therefore increased presence [25]. However, a follow-up study indicates that the increase in presence was due to the addition of dynamic behavior to the environment (shadows and reflections), rather than the illumination quality itself [30]. Welch et al. found that changing pictorial realism had a significant effect on presence, but anecdotally, the effect was less important than that of interactivity or latency [28].

That said, there have been some investigations into the effects of visual realism on other aspects of user experience and behavior. Thompson et al. investigated the effects of visual realism on distance judgments in virtual environments, and found that increasing the realism of the virtual scene had no significant effect on distance judgments [27]. Lee et al. explored the effects of visual realism on performance of search tasks in mixed reality environments, and found no significant effect of visual realism on task performance or on presence [14]. Ragan et al. explored the effects of visual complexity on performance and training transfer in a visual search task, and found that high visual complexity led to worse performance during the training trials, but better adherence to the search strategy during the evaluation trials [19].

Another relevant area of research is the study of how human perception, and particularly visual perception, changes when the perceived stimuli are virtual rather than real. The most widely studied such phenomenon is distance perception. It is a known phenomenon that distances are underestimated in immersive virtual environments. (For a survey of this topic, see [20].) In a series of studies that share much in common with our research, Interrante, Ries, Anderson, Lindquist, and Kaeding investigated distance perception in photorealistically replicated virtual rooms. The first paper reported that distances were *not* underestimated in such a room [10]. The second paper investigated this phenomenon further, exploring two hypotheses that might explain this effect: (1) exposure to the real environment enabled participants to better *calibrate* their judgments of size and distance in the virtual environment, and (2) exposure to the real environment generated a greater *sense of presence* in the virtual environment, enabling participants to treat it "as if" it were a real environment [11]. Their results do not support the calibration hypothesis, but leave open the possibility of the presence hypothesis. (This is in line with the thinking of Slater, who argues that presence—defined as a combination of Place Illusion and Plausibility Illusion—can lead one to "respond-as-if-real" in a virtual environment [24].)

There is a smaller body of literature that has specifically explored participants' perception of not only distance, but *space*. In [21], Saleeb reported on a user study wherein participants were asked to evaluate the sizes of physical rooms and virtual replicas of them. Their results suggest that users underestimated all dimensions of the virtual replica compared to the real room. This is in line with what might be expected from the distance perception literature. In [29], Yoon, Byun, and Chung report on a user study investigating whether a real room and a corresponding virtual replica were perceived in the same way. In this study, the perceived length and width of the real room and its virtual replica were not significantly different, while the height of the room was (the direction of this difference was not reported). That said, this study had only 12 participants, the only stimulus was a small roughly cubic room, and participants did not have direct control over their viewpoints. As a result, we are inclined to take the fact that these results do not seem to generalize as evidence of an anomalous result.

A final related phenomenon is the perception of *size* in virtual environments. Kenyon and colleagues investigated size-constancy—the effect where an object is perceived to be the same size regardless of its distance from an observer—in virtual environments, identifying a *loss* of size constancy in VEs [13] [12]. Nguyen and colleagues investigated the combination of scale and distance perception in virtual environments, finding that distance estimates were strongly influenced by familiar size cues [16].

2.1 Definitions

In this work, we accept Slater's conception of presence arising from a combination of Place Illusion and Plausibility Illusion [24]. *Plausibility Illusion*, as discussed earlier, is defined by Slater as, "the illusion that what is apparently happening is really happening (even though you know for sure that it is not)." *Place Illusion*, on the other hand, is defined as, "the...illusion of being in a place in spite of the sure knowledge that you are not there" [24]. Place Illusion, then, is akin to the traditional conception of presence as "being there," while Plausibility Illusion is a new construct that results from the "believability" of the experience.

Both Place Illusion and Plausibility Illusion are subjective experiences. We believe that these subjective experiences are, at least in part, driven by objective characteristics of the virtual reality system and scenario. Following Slater, we refer to the system characteristics that drive Place Illusion as immersion [24]; and following Skarbez, we refer to the system characteristics that drive Plausibility Illusion as coherence [22]. *Immersion*, in turn, is defined as the set of valid actions supported by a virtual reality system, while *coherence* is defined as the set of objectively reasonable circumstances that can be demonstrated in a virtual scenario. (For a more detailed discussion of these terms, their definitions, and the rationales for those definitions, see Section 2 of the Skarbez, Brooks, and Whitton survey [22].)

3 Overview of User Studies

The first user study, described in Section 4, was a pilot study of which the goals were to: demonstrate the feasibility of our virtual experimental testbed, check whether any parameters important to users were not considered in the creation of the testbed, generate a "first pass" ranking of the importance of various parameters, and most critically, identify which parameters merited further investigation in User Studies 2 and 3.

The second study, described in Section 5, was a classical psychophysical study designed to evaluate the subset of parameters identified in Study 1 with respect to one another. Specifically, this user study sought to identify perceptual equivalences between different parameters so that the parameters could be correctly valued in the budget-based Study 3.

The third user study, described in Section 6, used a budget-based method derived from the method introduced by Slater in [26] and employed by Azevedo, Jorge, and Campos [1], Bergström et al. [4], and Skarbez et al. [23], among others. In this study, participants were presented with the replica environment in some initial (degraded) state, and were given the opportunity to upgrade the environment through manipulation of parameters (the same from Study 2) in whichever order they saw fit and to whichever extent they saw fit, given an overall budget constraint. This user study sought to establish which parameters participants considered most important, and "how correct" those parameters needed to be to satisfy participants.

4 Study 1 (Pilot)

As described in Section 3, this was a pilot study of which the primary goal was to identify room parameters that required further investigation. Thus, we sought to investigate a large set of parameters that we could confidently winnow down for use in subsequent user studies.

For this research, we staged a full-size actual room with furnishings typical of a living room including a sofa, two chairs, end tables, coffee table, lamps, a television, bookshelf, art, and several other smaller

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2021.3096494, IEEE Transactions on Visualization and Computer Graphics

items. We added clutter to many of the horizontal surfaces (e.g., coffee and end tables, bookshelves, the floor) that would representative in a real living room setting. These clutter included a set of bongos, a metal pig sculpture, free standing picture frames, books, pamphlets, remote controls, drinking cups, small toys, and many others. In all there were just over 100 items in the room. A photograph of the real room can be seen in Figure 1 at left.

The experimental testbed was implemented using version 5.6 of the Unity game engine. The base room model was generated using a combination of laser scanning to obtain global information, photogrammetry scanning of individual small objects, and additional processing using Pixologic ZBrush, Autodesk Maya, and Autodesk 3ds Max.

In order to investigate which parameter changes were perceptually important to participants, modified replica rooms were created by changing several parameters of the "most accurate replica" (MAR) room—the most accurate version of the virtual replica room, as seen in Figure 1 (right). (Note that the "stripes" in the MAR room result from light coming through Venetian blinds on the wall behind the virtual camera.)

It is important to note that the types of errors considered in this analysis are not simply the errors associated with construction of a triangulated mesh. Construction of such meshes has its own difficulties, such as how to avoid holes in the output mesh, or how to deal with "difficult" objects in the scene, such as those with reflective, transparent, or anisotropic materials. Here, we assume that "perfect" triangulation is possible; however, this is still not necessarily sufficient for VR. In the general case of a real place containing multiple objects, a triangulated mesh generated from a finite number of camera angles is insufficient if a user is enabled to freely move about the scene: The user may view the scene from an angle where the mesh has holes or is unnaturally stretched over occluded concavities. Furthermore, if a user is to interact with objects in the scene, each object must be represented as a separate model. A "VR-suitable" model of a real scene, then, requires the scene to be composed of individual object models. Construction of such VR-suitable models cannot at present be done automatically. Efforts to do so are in the realm of computer vision, and specifically 3D semantic segmentation and reconstruction. Naseer, Khan, and Porikli have published a recent survey on indoor scene understanding with a focus on autonomous agents, but the methods and problems discussed therein also apply to virtual reality [15]. In this work, we have manually created such a segmented scene representation, and introduced errors that we believe could possibly occur in this process.

After significant iterative internal testing, we decided to alter most parameters by $\pm 10\%$, $\pm 25\%$, $\pm 50\%$, or $\pm 75\%$ depending upon the parameter. The goal was to create obvious, noticeable differences between parameter levels for this study, so as to evoke strong responses from Study 1 participants, knowing that in future studies participants would be able to manipulate the parameters with fine granularity. Modifications were drawn from the following list:

- **Ceiling height** changed by +25%, +10%, -5%, and -10%. (The asymmetry is due to the fact that internal testing revealed that for some users, -25% put their head at ceiling level, which was immediately noticeable and disturbing.)
- **Room length** changed by +25%, +10%, -10%, and -25%.
- **Room width** changed by +25%, +10%, -10%, and -25%.
- **Door removed**. The door was replaced by a blank wall.
- **Window removed**. The window was replaced by a blank wall.
- **Wall material** changed in color or texture.
- **Furniture removed**. One of the following was removed at a time: couch, coffee table, side tables, TV stand, and a stuffed chair. When a piece of furniture was removed, all clutter that was "on" that furniture object was also removed, to eliminate the incoherent stimulus of floating objects.
- **Furniture quality reduced**. For both the couch and the stuffed chair, models were generated that contained 10%, 25%, 50%, and 75% of the original number of vertices. If "decimated furniture" was one of the errors in a given room, the sofa and the stuffed chair always varied together.
- **Furniture mismatched**. One of the following was replaced at a time: coffee table, side tables, stuffed chair, couch, lamps, and a wooden chair. When a piece of furniture was mismatched, it was replaced with another similar object of the same type taken from the ShapeNet database [6]. (Note that ShapeNet objects were not necessarily obtained by photogrammetry, so it may have been easy for participants to notice that they came from a different source.)
- **Furniture repositioned**. Furniture objects were either globally raised by 10cm, globally lowered by 10cm, or globally moved outward (away from room center) by 10%. The corresponding "moved inward by 10% condition" was not tested due to experimenter error.
- **Furniture rescaled**. One of the following scaling errors occurred: All furniture was 25% larger, all furniture was 10% larger, all furniture was 10% smaller, the sofa was 25% larger, the sofa was 10% smaller, the coffee table was 25% larger. When a furniture object was scaled up, all clutter that was "on" that furniture object was also scaled up, to avoid giving participants conflicting context cues.
- **Clutter removed**. Either 10% (9) of clutter objects were hidden, 25% (22) of clutter objects were hidden, or 50% (44) clutter objects were hidden. Clutter objects were hidden at random, and in different random orders for each of the three conditions. Physical constraints were not enforced; for example, it was possible for the middle book in a stack of three to be hidden.
- **Clutter quality reduced**. For the mask, the white pitcher, the bongos, and the pig statue, models were generated that contained 10%, 25%, 50%, and 75% of the original number of vertices. If "decimated clutter" was one of the errors in a given room, all four objects always varied together.
- **Clutter repositioned**. One of the following positioning errors occurred: Four objects—always the dumbbells, the coffee mug, the brochure, and the white pitcher—were elevated by 10cm, the four objects were elevated by 5cm, the four objects were lowered by 5cm, the four objects were lowered by 10cm, all clutter was moved inward by 25%, or all clutter was moved outward by 10%.
- **Clutter rescaled**. One of the following scaling errors occurred: the four objects were 25% larger, the four objects were 10% larger, the four objects were 10% smaller, the four objects were 25% smaller, all clutter was 25% larger, all clutter was 10% larger, all clutter was 10% smaller, all clutter was 25% smaller.
- **Lights missing**. One of the following lights was "turned off": "sunlight", ceiling lights, or both lamps. Note that "sunlight" was implemented as a directional light outside the room, and the ceiling lights did not actually light the room, since area lights could not be enabled—turning the ceiling lights on or off only changed whether the ceiling light panels appeared to "glow" or not.
- **Light brightness changed**. All lights were either 4 times as bright, 2 times as bright, ½ as bright, or ¼ as bright. All lights always varied together.
- **Light color changed**. All lights were interpolated 25% toward blue, 10% toward blue, 10% toward red, or 25% toward red. All lights, including the ambient lighting, always varied together.

In total, ten different modified replica rooms were created, each of which had between five and nine parameters modified. Screenshots for each of these rooms appear in Figure 2.

4.1 Participants

Six participants were recruited from the students, faculty, and staff of Virginia Tech's Grado Department of Industrial and Systems Engineering, in which the study was being conducted. For this study, participant

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2021.3096494, IEEE Transactions on Visualization and Computer Graphics

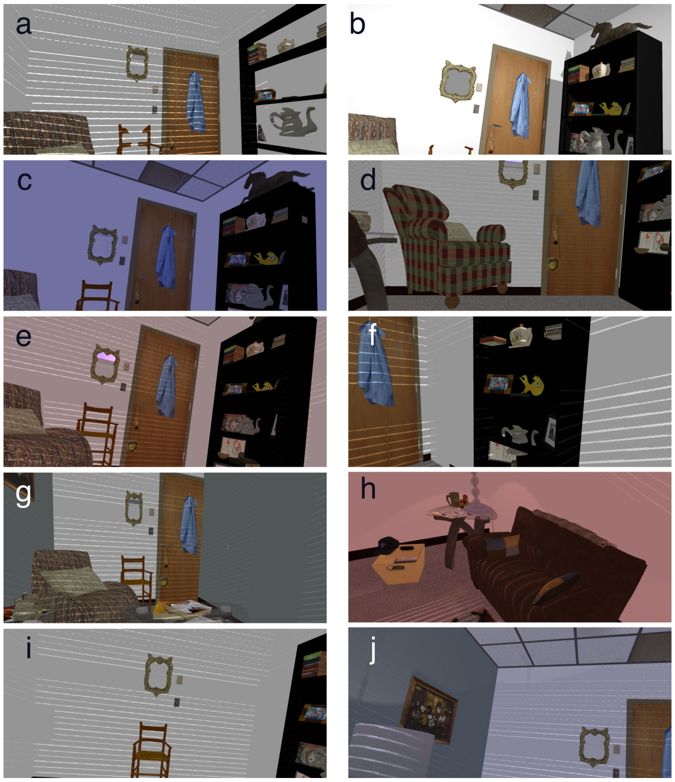

Fig. 2: Screenshots from each of the ten modified replica rooms participants experienced in Study 1.

descriptors such as age, gender, etc. were not recorded, and participants were not compensated.

4.2 Materials

The virtual environment was displayed using an Oculus Rift CV1 head-worn display (HWD). The Rift has a nominal 110° field of view, and a resolution of 1080x1200 pixels per eye. It weighs 470 grams.

For tracking, the Rift tracking system was used in a roomscale (3 camera) configuration. The size of the tracked space was approximately 2m x 3m, and was comparable to the navigable space available in the real room.

4.3 Measures

For this study, participants were asked to "think aloud" as they explored the environment, and specifically to comment on things that seemed unusual about the virtual environment. These comments were transcribed by an experimenter, and these comments were analyzed in order to generate descriptive statistics regarding (1) what the "unusual" things that were noticed by participants were, and (2) in what order they were commented upon.

4.4 Procedures

Upon arriving at the lab, participants received a brief description of the task, and then donned the Rift HWD. Participants first experienced the MAR room; this was done so that participants would have a baseline experience against which to compare the subsequent degraded virtual rooms. Participants were then told that they would experience a series of modified versions of these rooms, and that they should "think aloud" and comment on the differences they noticed, if any, between the modified room and the MAR room. Participants were re-exposed to the MAR room after each exposure to a modified room so as to refresh their memory. (We acknowledge that ideally participants would have been exposed to the actual room rather than the MAR room, but this was not possible due to time and access constraints.) The entire session lasted approximately 30 minutes.

4.5 Results

In this section, we summarize the results of this user study by listing the parameters that were not selected for further investigation in Studies 2 and 3, followed by those that were, with associated justifications.

4.5.1 Parameters not selected for further investigation

These parameters were generally always noticed by all participants, or rarely noticed by any. In either case, it was decided that we should not spend additional effort in trying to determine their relative importance.

- Ceiling height (Rarely noticed)
- Door present or absent (Always noticed)
- Window present or absent (Always)
- Wall material (Always)
- Furniture removed (Always)
- Furniture quality reduced (Rarely)
- Furniture mismatched (Always)
- Furniture position - translation (Rarely)
- Clutter quality reduced (Never)
- Clutter position - translation (Rarely)
- Clutter scale (Rarely)
- Light brightness (Never)
- Light color (Always)

4.5.2 Parameters selected for further investigation

These parameters differed in how early or how often they were noticed depending on the level of the parameter. As a result, we decided that we would further investigate these parameters to determine at which levels they are important and how important they are relative to one another.

- Room length and width (combined in Study 3 as *room scale*)
- Furniture position - elevation
- Furniture scale
- Clutter (removed several pieces at a time)
- Clutter position - elevation
- Lights ("turned off" one at a time)

5 Study 2

As described in Section 3, this user study was a psychophysical study of which the primary goal was to identify subjective equivalences between different parameters of the VE, specifically the seven parameters identified in Section 4.5.2. This was needed in order to appropriately value these parameters in the planned Study 3 (Section 6).

In order to accomplish this, we designed the study as follows. In each trial, a participant would experience three versions of the virtual replica room. First, the *MAR room* as in Study 1, to give the participant a point of comparison. Second, an *Exposure room*, in which one of the seven parameters was set to one of five levels (including unchanged), and the participant was asked to verbally rate how different the Exposure room felt from the MAR room, on a scale from 1 to 7. Finally, the participant experienced a *Test room*, in which they were able to control one of six parameters (six because the parameter from the Exposure room could not be reused), and were asked to adjust that parameter using the Oculus Touch joystick until the Test room felt "as different" from the MAR room as the Exposure room had. Note that while many of the Exposure rooms had parameter values that could vary both above and below the correct value, in the Test rooms, parameter values were constrained to always vary in one direction—above or below, but not both. For example, in the Exposure rooms, room width took on values between 0.5x and 1.5x the veridical value, but in the Test rooms, room

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2021.3096494, IEEE Transactions on Visualization and Computer Graphics

width could only vary between 1.0x and 1.5x. This was done to simplify the participants' decision space and the subsequent analysis.

As an example, consider that the parameter that is varied in the Exposure room is room width, and that it was set to 50% of the true room width. The participant might consider this room very different from the MAR room, and assign a difference rating of 7. Then, in the Test room, the participant-controlled parameter is the number of clutter objects. They are asked to adjust this parameter until the Test room feels as different from the MAR room as the Exposure room did; that is, until it feels like a 7. They then remove all the clutter objects from the room (setting number of clutter objects to 0), and declare a match. This ends the trial.

Each participant underwent 210 trials (7 exposure room parameters x 5 stimulus levels of exposure room parameters x 6 test room parameters). The total duration was approximately 3-4 hours. The system automatically suggested breaks after every 20 trials (approximately every 20 minutes). Screenshots from the Exposure and Test rooms appear in Figure 3.

Fig. 3: Screenshots from the Exposure Room (a, b) and Test Room (c, d) for one trial in Study 2.

5.1 Participants

Eight participants (six female) were recruited from the student population of Virginia Tech's Grado Department of Industrial and Systems Engineering, in which the study was being conducted. Participants were compensated at a rate of $5/half hour. This user study was approved by the Virginia Tech Institutional Review Board, #17-491.

5.2 Materials

The materials used in this study were the same as in Study 1, with the addition of an Oculus Touch controller. The controller was used so that the user could manipulate environment parameters by moving the joystick left and right.

5.3 Measures

For each trial, two data points were collected. The first was a "difference rating" (on a scale from 1-7, as shown in Figure 3b) of the Exposure room compared to the MAR room, and the second was the "point of equivalence" in the Test room—that is, the level at which the participant deemed the parameter in the Test room to feel "as different" from the MAR room as the Exposure room did.

5.4 Procedures

Upon arriving at the lab, participants read and signed an informed-consent form and completed a short demographic questionnaire. Participants were informed both verbally and in writing that they were free to withdraw from the study at any time and for any reason. After completing this process, participants donned the Oculus Rift HWD and were exposed to a familiarization environment, in which they calibrated the Rift for their IPD and familiarized themselves with moving in the Rift and the Guardian system. After completing the familiarization process, participants began the trials, which proceeded as described in Section 5.

5.5 Results

The results from Study 2 are summarized in the graphs in Figures 4 and 5.

Figure 4 shows a grid of subgraphs, each plotting the values of a Test room parameter corresponding to an Exposure room parameter. This figure is quite complicated, and included primarily for completeness. That said, there are some insights that can be gleaned from it. One is that, if a parameter behaves "normally", the graphs for that parameter's column in the figure should be roughly V-shaped. This is due to the fact that if an Exposure room parameter value is much smaller than normal (the left-most point in each graph), or much larger than normal (the right-most point in each graph), the corresponding Test room parameter value should be large; meanwhile, if an Exposure room parameter value is true to the MAR room (the center value in each graph), the Test room parameter should also be true to the MAR room (minimum). From these graphs, it can be observed that room length and room width behave this way, as does clutter pieces included to a degree. The other parameters are not so well-behaved. Specifically notice the light level column, in which light level is rarely changed from its base level.

Figure 5 shows, for each Exposure room parameter, the resulting difference ratings at each level of that parameter. Note that these graphs only refer to the "difference rating" measurements discussed in Section 5.3, not to the "point of equivalence" measures from that same section. The distinct V-shape in graphs *a*-*e* (with graph minima very close to 1) shows that participants did in fact recognize when Exposure rooms were unchanged from the MAR room. This enables the results to be interpreted with greater confidence. The asymmetry in graphs *a*-*b* indicates that where room scale is concerned, participants perceive shrunken rooms to be substantially "more different" when compared to enlarged rooms. By comparison, the relatively symmetric graphs in subfigures *c* and *d* seem to indicate that furniture height and furniture scale are comparably noticeable whether the parameters are increased or decreased, and graph *e* indicates that clutter that is floating above a surface is more noticeable than clutter that is sunken into it. Graph *f* indicates that the Exposure room was (correctly) perceived as most similar when all clutter was present and most dissimilar when all clutter was absent, but once half the clutter was missing, removing even more did not make it more noticeably different. Graph *g*, which illustrates the lights being turned on or off, on the other hand, barely rises above the minimum value, regardless of how the parameter was manipulated. Subfigure *h* shows all the other graphs superimposed, so their relative magnitudes can be more readily compared. Here it can be seen that room scale and furniture scale parameters have the greatest impact, and lighting by far the least.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2021.3096494, IEEE Transactions on Visualization and Computer Graphics

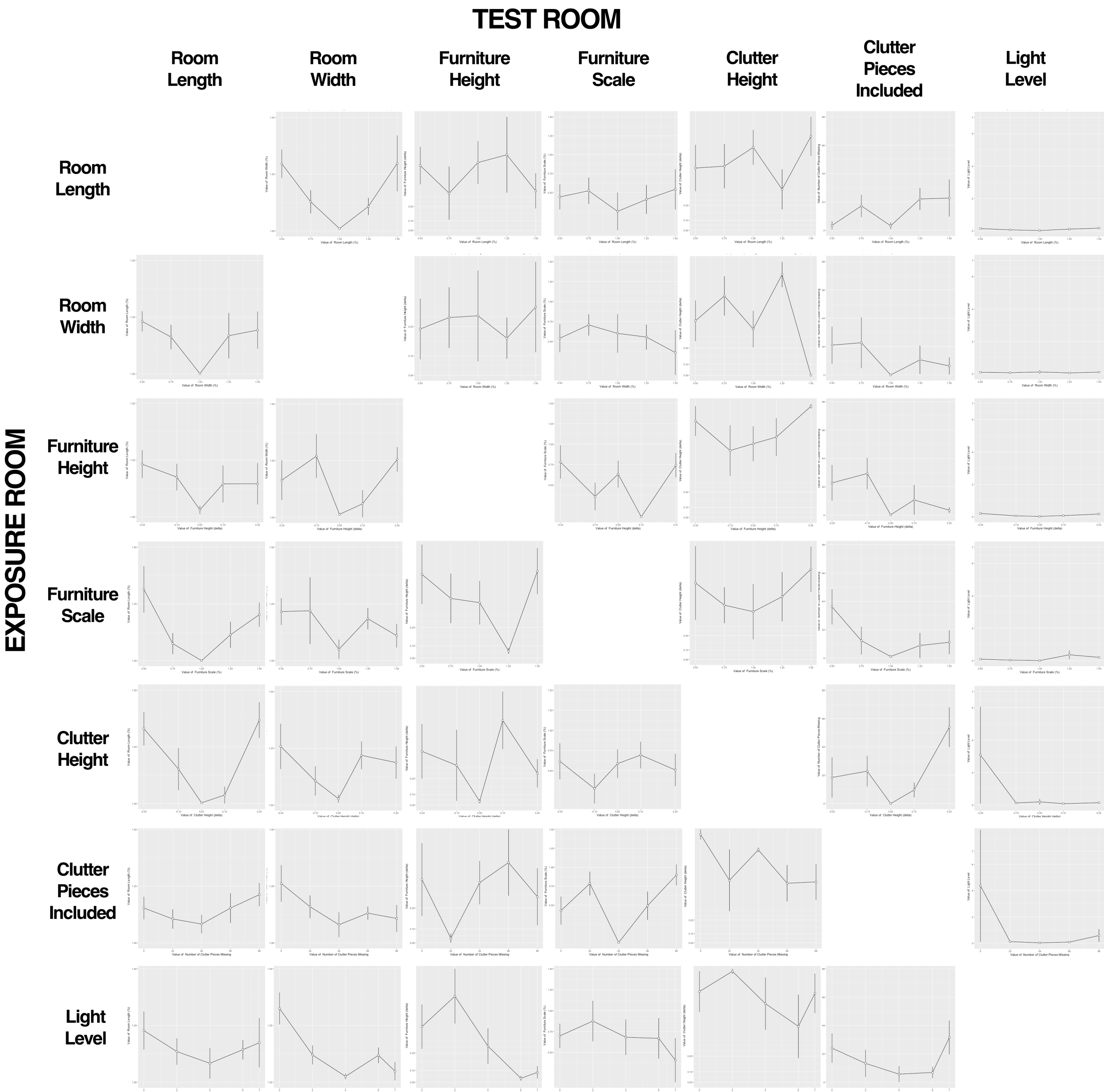

Fig. 4: Graphs of Test room parameter level by Exposure room parameter level.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2021.3096494, IEEE Transactions on Visualization and Computer Graphics

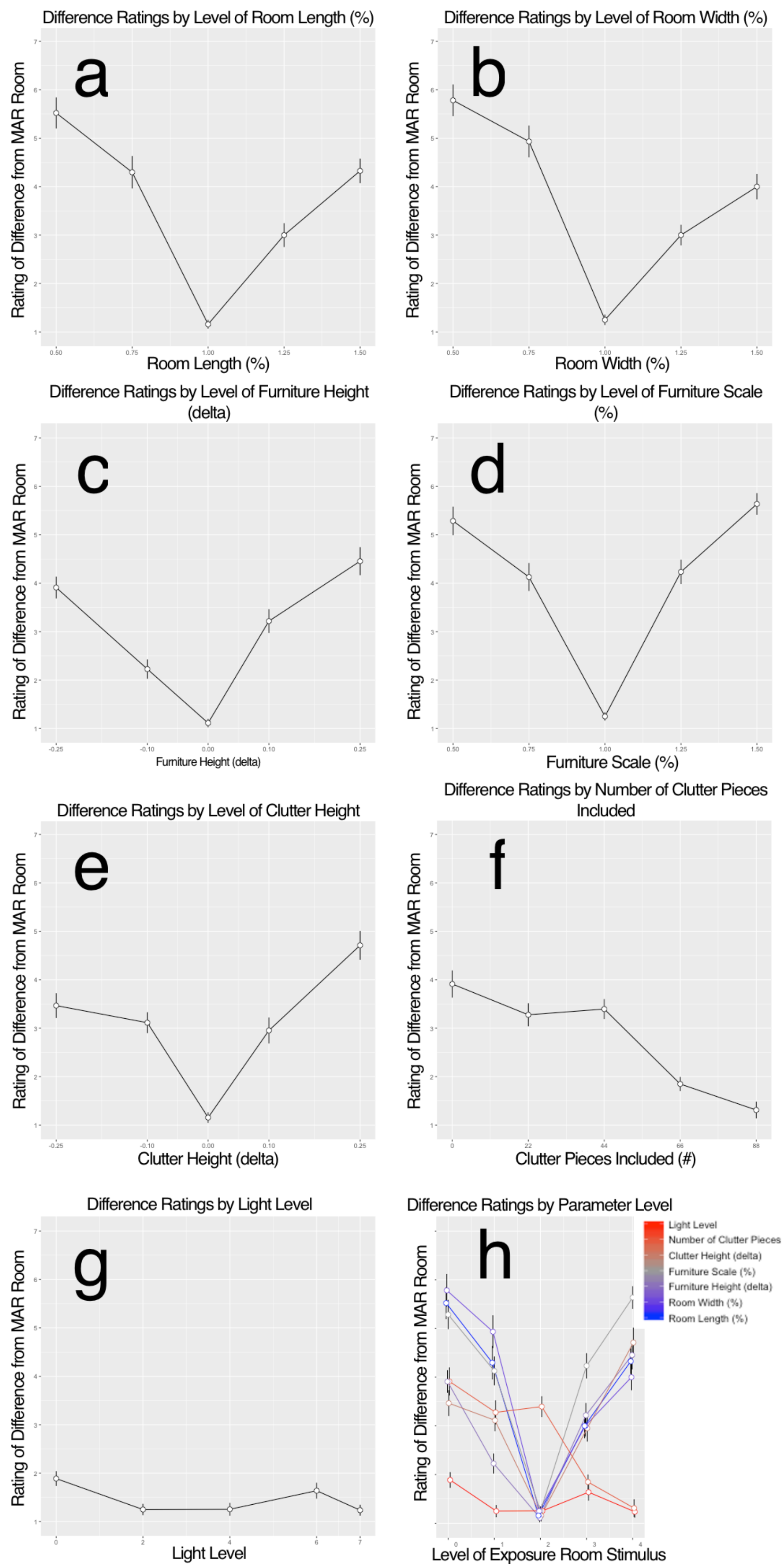

Fig. 5: Graphs of difference rating by each parameter; (a) room length, (b) room width, (c) furniture height, (d) furniture scale, (e) clutter height, (f) clutter pieces, (g) lighting condition, (h) all parameters together.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2021.3096494, IEEE Transactions on Visualization and Computer Graphics

6 STUDY 3

As described in Section 3, this user study was a "budget-based" study of which the primary goals were to generate a rank ordering of the studied room parameters, as well as "how correct" each parameter needed to be, for the virtual replica room to feel as perceptually similar to the real original room as possible. This is a meaningful change from Studies 1 and 2, where participants did not experience the real physical room at all, and comparisons were made against the MAR room.

In this study, we used the same parameters from Study 2, however, the lights were broken up into three separate budget items (sun light, lamp lights, and ceiling lights), and room length and room width were combined into a single parameter, room scale. We refer to each instance of these eight parameters as a *configuration*, and denote a configuration with a property vector of the form C = { RoomScale, FurnitureElev, FurnitureScale, ClutterPieces, ClutterElev, LampLight, SunLight, CeilingLight }. Details regarding the costs associated with each of these parameters can be seen in Table 1; these costs were informed by the results of Study 2—furniture scale is most important—and therefore most costly—followed by room scale, furniture elevation, amount of clutter, elevation of clutter, and lighting. (Note that lighting was included even after being judged as least important in Study 2. This is because the choice of parameters for both Studies 2 and 3 was made before Study 2 was conducted.)

Each participant was first exposed to the real original room, as depicted in Figure 1a. Participants were instructed to "Pay attention to 'how real' this room feels; you are going to experience several copies of this room in virtual reality, and we will ask you to change the virtual room until it feels as real as possible." After this, they were escorted to the virtual reality room, where they donned the Oculus Rift and Touch controllers, and experienced a substantially modified version of the MAR room, and were given a points budget to make improvements to that room. (Upgrading every parameter to the maximum level cost 316 points, as shown in Figure 6 and broken down in Table 1. In the training room, participants were given a budget of 316 points, so as to expose the user to all the possible upgrades and what they felt like. During each recorded trial, participants worked with a restricted improvement budget.) Participants then doffed the equipment, were re-exposed to the real original room, and re-donned the equipment for the next trial in virtual reality. This process was repeated for each recorded trial, of which there were seven for each participant. Each trial started from one of the seven configurations listed in Table 2; these were presented to each participant in randomized order. Note also that each trial began with a different randomly-chosen parameter selected, but the parameters were always listed in the same order. So, for example, a trial could start on any of the eight parameters, but *Furniture Elevation* would always appear between *Room Scale* and *Furniture Scale*.

The virtual environment, along with the budget/upgrade user interface and each of the improvements, is illustrated in Figure 6.

In the virtual reality trials, each participant was instructed to, "spend your improvement budget in order to make the room feel as 'real' as possible as 'quickly' as possible. That is, if one property feels most important to you, you should improve that one first, then the next most important property, and so on." Participants were able to explore the entire parameter space (both which parameter to adjust and how much of the points budget to spend on it), but once they confirmed their expenditure (by orally informing the experimenter), that parameter became "locked", and could not be revisited or further adjusted. This means that, for example, a participant could not upgrade room scale partway, then the number of clutter pieces, then return to room scale. This was done to simplify both the study procedures and the analysis. The trial ended when the participant had spent their entire budget.

6.1 Participants

Forty participants (nineteen female) were recruited from the student population of Virginia Tech, at which the study was being conducted. Participants were compensated at a rate of \$5/half hour. This user study was approved by the Virginia Tech Institutional Review Board, #17-491.

6.2 Materials

The materials for this study were the same as in Study 2.

6.3 Measures

There were two types of dependent variables: (1) the sequence in which the participant chose to improve room parameters, and (2) the amount of their budget they chose to spend on each improvement. By construction, there were always an equal number of measurements of types (1) and (2), but there could be a different number of expenditures for each trial. For example, one participant could choose to improve all eight parameters, while another might spend their entire improvement budget on improving three parameters to the maximum level. Each participant underwent seven trials.

6.4 Procedures

The pre-experiment procedures here were the same as the procedures for the previous study, as described in Section 5.4. After completing the familiarization process, participants began the trials, which proceeded as described in Section 6. At the end of the virtual reality trials, participants completed a short post-experiment questionnaire. The entire study lasted approximately one hour.

6.5 Results

As in [26], we make the simplifying assumption that the results of the seven trials are statistically independent. This is not strictly the case; each participant carried out a series of trials, and learned about the room parameters from trial to trial. However, the study was designed such that each trial began from a different starting configuration, with different pre-selected parameter changes and as a result, different points budgets. Because of these changes, participants would have had to reconsider their parameter choices in every trial.

In the remainder of this section, we report separately on the three types of dependent variables: the parameter sequences chosen in each trial, the amount of budget spent on each parameter, and the post-experiment questionnaire data.

6.5.1 Transitions

From the parameter sequences chosen in each trial, we constructed a transition probability matrix P. Over the 280 total trials, there were 1727 observed parameter changes, for an average of 6.17 parameter changes per trial. By the design of Study 3, P is a 256x256 matrix. (There are 8 parameters, each of which can be in one of two states: changed (1) or unchanged (0). So there are $2^8 = 256$ possible configurations, and to consider the probability of a transition from any state to any other state, P needs to have 256^2 cells.) Because we placed several restrictions on allowed parameter choices at any given time, P is quite a sparse matrix: there are only 273 distinctly observed state transitions (out of 65536).

Once P is known, it is possible to compute the probability distribution of transitioning to any given configuration from any given configuration. If we choose as a starting configuration $C = 0,0,0,0,0,0,0,0$ (no parameter changed), and define s as a 256-vector of all zeros except the element corresponding to C, then sP gives the probability distribution after one parameter has been changed, sP^2 after 2 changes, and sP^n after n changes. By construction, the configuration $C = 1,1,1,1,1,1,1,1$ is absorbing, so the eighth transition adds no information. Therefore, we report on the first seven transitions in Table 3.

6.5.2 Expenditures

At each step, participants had to decide not only which parameter to change, but how much of their points budget to spend on that parameter. This data was collected in order to estimate "how correct" each parameter had to be. For example, participants regarded RoomScale as the most important parameter, but as indicated in Table 4, they only upgraded it to 0.908 on average, suggesting that this was "close enough" for most participants. Clutter elevation, on the other hand, was less important, but the average participant upgraded it to the maximum

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2021.3096494, IEEE Transactions on Visualization and Computer Graphics

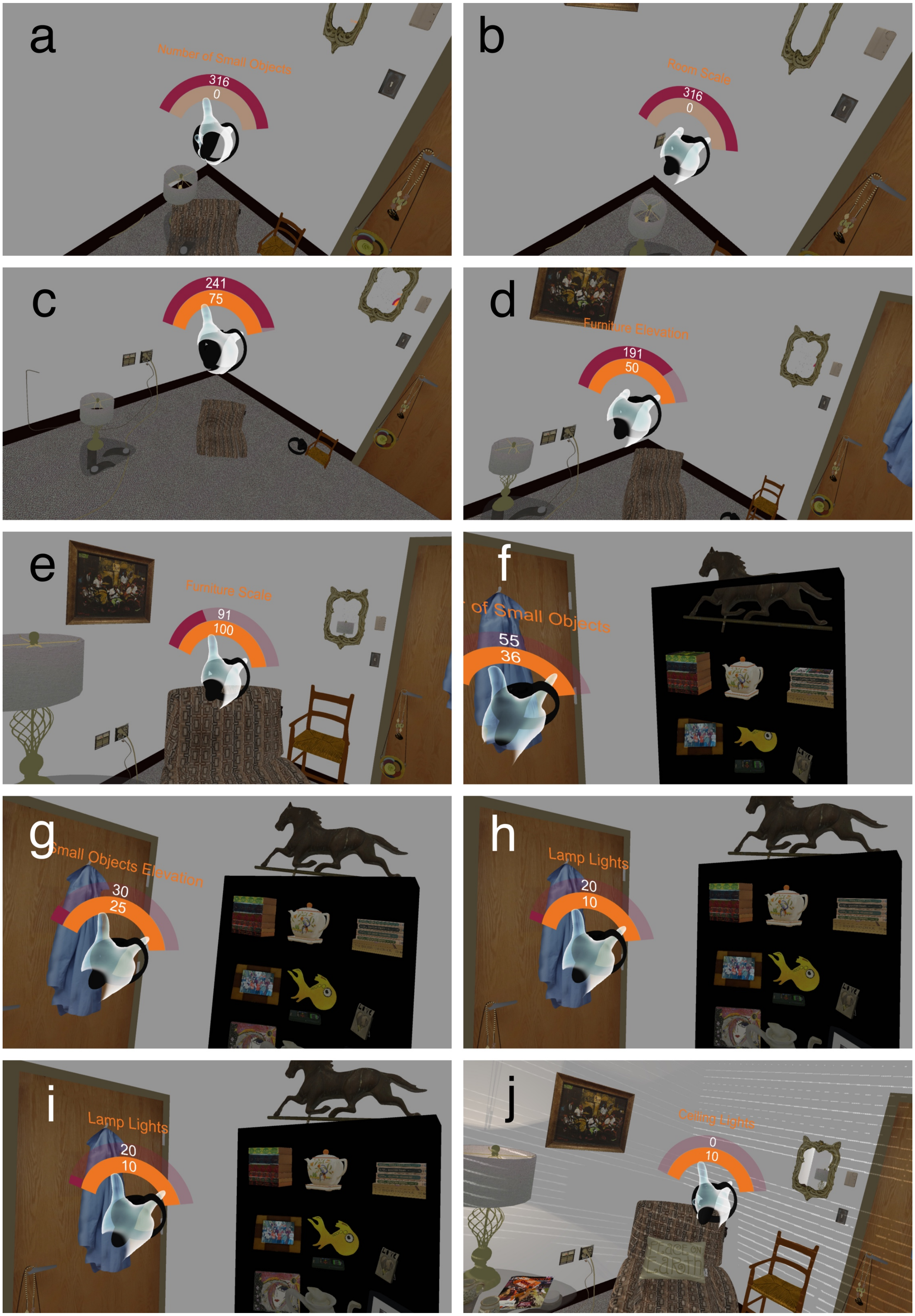

Fig. 6: Screenshots showing a full sequence of all possible improvements in the Training environment of Study 3.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2021.3096494, IEEE Transactions on Visualization and Computer Graphics

Table 1: List of parameters, ranges, increments, and costs for Study 3.

Parameter	Initial value	Max Value	Increment	Cost in points/unit (total cost)
RoomScale	0.5x	1.0x	0.01x	1.5 (75)
FurnitureElev	-25cm	0cm	1cm	2 (50)
FurnitureScale	0.5x	1.0x	0.01	2 (100)
ClutterPieces	0 pieces	72 pieces	1 piece	0.5 (36)
ClutterElev	-25cm	0cm	1cm	1 (25)
LampLight	0 (off)	1 (on)	1	10 (10)
SunLight	0 (off)	1 (on)	1	10 (10)
CeilingLight	0 (off)	1 (on)	1	10 (10)

Table 2: List of starting configurations for trials in Study 3.

Starting configuration	Improvements pre-assigned	Points available
{0,0,0,0,0,0,0,0}	None	250
{1,0,0,0,0,0,0,0}	RoomScale (75)	175
{0,1,0,0,0,0,0,0}	FurnitureElev (50)	200
{0,0,1,0,0,0,0,0}	FurnitureScale (100)	150
{0,0,0,1,0,0,0,0}	ClutterPieces (36)	214
{0,0,0,0,1,0,0,0}	ClutterElev (25)	225
{0,0,0,0,0,1,1,1}	All lights on (30)	220

level, suggesting that there was no elevation difference that was "good enough."

6.5.3 Questionnaires

After completing the study, all participants completed a questionnaire. We do not report data for all responses here, choosing to focus on two aspects of the questionnaire: (1) "feeling of reality" scores, and (2) participants' subjective rankings of the importance of the studied parameters.

Regarding their feelings of reality, participants were given two prompts: "Picture in your mind the WORST version of the virtual room. On a scale of 0 to 100 (100 being equally as real as the real world room), how real did that room feel to you?" and "Picture in your mind the BEST version. . . " The summary statistics regarding these two scores are shown in Table 5; as one would expect, participants rated their memory of the BEST room much higher, with high statistical significance.

Regarding their subjective rankings of parameter importance, these added credence and context for the psychophysical data presented in Sections 6.5.1 and 6.5.2. Participants were given a series of prompts, of the form "When improving the virtual environment, which factor was most important [second-most important, third-most important, etc.] for you?" Participants could choose any of the eight parameters, as well as "No factor was particularly more [less] important than the others." These data are presented in Table 6, similarly to Table 3. Note the similarities between these two tables.

7 Discussion

Throughout this section, we discuss the implications of the results from the three studies organized as claims about the data followed by the supporting evidence for those claims.

7.1 Room Scale is the most impactful of the studied parameters

In both Study 2—in which room length and room width at 0.5x represented 2 of the 3 highest observed difference ratings over the whole parameter set—and Study 3—in which room scale was subjectively the most important parameter to a substantial majority of participants, was the first parameter upgraded in a substantial majority of trials, and had the second highest gross mean expenditure—room scale was observed to be the most impactful parameter.

This is perhaps not surprising, as the scale of the room overall provides the context by which to evaluate the scale of individual objects. (Correctly-sized furniture in an implausibly small room would likely appear even more incoherent than implausibly small furniture in the same room.) From Study 2, we can see that participants rated smaller rooms as more incoherent than larger rooms for the same MAR room (Figures 5a and 5b). In Study 3, the median accepted room scale was 0.92x, indicating that participants were willing to accept a room approximately 8% smaller than the real room as "feeling real."

7.2 Furniture Scale is the second most impactful of the studied parameters

In both Study 2—in which furniture scale of 1.5x represented the second highest overall difference rating—and Study 3—in which furniture scale was subjectively the second most important parameter to a substantial majority of participants, was the second parameter upgraded in a substantial majority of trials, and had the highest gross mean expenditure—furniture scale was observed to be the second most impactful parameter.

Again, this is not particularly surprising. Furniture objects are the biggest objects in the room, save for the room itself, so they might be expected to make the biggest impact on the coherence of the room. In Study 3, the median accepted furniture scale was 0.85x, indicating that participants were willing to accept furniture 15% smaller than MAR as "feeling real."

Note that this is in a room that was accepted as 8% smaller than MAR on average, so 15% overstates the difference. If one considers room scale and furniture scale to be a single percept of "relative scale", the relative scale difference between the two is only 7%. Study 2 provides some evidence that room scale and furniture scale are linked in this way, as while the room scale differences were perceived as worse when the room was smaller than in the MAR room, the furniture scale differences were perceived as slightly worse when the furniture was bigger than in the MAR room (Figure 5d). In both cases, the furniture was "too big" for the room. And in both cases, this was perceived as less realistic than the furniture being "too small" for the room.

7.3 Both relative and absolute scales have a large impact on Plausibility Illusion

Section 7.2 discussed the importance of the *relative* scale of objects in a virtual replica room, as users seem to be keenly aware of the fact that an object is too big or too small compared to its context. However, *absolute* scale is also critically important in virtual reality experiences, in a way that is not the case with other media. In a single image or even a movie—particularly one that is computer generated—it is impossible to judge the absolute scale of the scene, since the camera is just a point, can have any height, and can move at any speed. In VR, however, the user themself provides context for absolute scale: "The characteristics of our body in metric terms, such as size, eye height, walking speed, etc. constitute the frame of reference and standard for the assessment of distances, position of objects, etc., both under a quantitative point of view for obtaining real measures, and a qualitative

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2021.3096494, IEEE Transactions on Visualization and Computer Graphics

Table 3:
Most likely state and next transition at every number of parameters changed (For readability, only transitions with probability ≥ 0.1 are shown).

Starting configuration	Probable transitions			Most likely action
{0,0,0,0,0,0,0,0}	{0,1,0,0,0,0,0,0} (0.1)	**{1,0,0,0,0,0,0,0} (0.75)**		Change RoomScale
{1,0,0,0,0,0,0,0}	**{1,0,1,0,0,0,0,0} (0.56)**	{1,1,0,0,0,0,0,0} (0.36)		Change FurnitureScale
{1,0,1,0,0,0,0,0}	**{1,1,1,0,0,0,0,0} (0.86)**			Change FurnitureElev
{1,1,1,0,0,0,0,0}	{1,1,1,0,0,0,1,0} (0.11)	**{1,1,1,1,0,0,0,0} (0.76)**		Change ClutterPieces
{1,1,1,1,0,0,0,0}	**{1,1,1,1,1,0,0,0} (0.88)**			Change ClutterElev
{1,1,1,1,1,0,0,0}	{1,1,1,1,1,0,0,1} (0.20)	{1,1,1,1,1,0,1,0} (0.37)	**{1,1,1,1,1,1,0,0} (0.43)**	Change LampLight
{1,1,1,1,1,1,0,0}	**{1,1,1,1,1,1,0,1} (0.5)**	**{1,1,1,1,1,1,1,0} (0.5)**		Change remaining lights in either order

Table 4: Summary statistics of expenditures.

	Maximum		Expenditure			Parameter value		
Parameter	Expenditure	Value	Mean	Median	S. D.	Mean	Median	S. D.
RoomScale	75	1.0	61.2	63	11.4	0.908	0.92	0.076
FurnitureElev	50	0	39.7	43	11.2	-0.051	-0.035	0.056
FurnitureScale	100	1.0	69.3	70	18.3	0.847	0.85	0.091
ClutterPieces	36	72	26.5	28.75	9.58	53.0	57.5	19.2
ClutterElev	25	0	20.3	25	7.93	-0.047	0	0.079
LampLight	10	1	4.02	0	4.92	0.402	0	0.492
SunLight	10	1	6.65	10	4.73	0.665	1	0.473
CeilingLight	10	1	4.31	0	4.97	0.431	0	0.497

Table 5: "Feeling of reality" statistics.

Mean (WORST)	Mean (BEST)	Difference	
46.5	81.3	38.2	$p < 0.001$

point of view for subjective considerations: near, far, big, small" [8]. In short, VR enables a user to "feel" the scale in a way that even 3D content does not. The importance of the scale parameters— especially the primary importance of room scale—in these studies supports this notion, although we did not compare against non-VR media.

Both relative and absolute scale, then, seem to have greater impact on the "feeling of reality" than furniture elevation, clutter pieces, and clutter elevation, despite the fact that these parameters could cause obviously *non-physical* room states, such as objects floating in air.

We propose two possible explanations for this. One is that changing the scale parameters has the largest possible sensory impact on the user (the largest objects in the room are the room shell components—walls, floor, ceiling—followed by the furniture objects). The other is that the scale parameters are subjectively more important because they are more important from an affordance perspective. That is, the scale parameters inform action possibilities—Is this chair the right size to sit in? Is that door big enough to walk through?—in a way that the other parameters do not. (We call these the *sensory impact* and *affordance* hypotheses, respectively.)

7.4 Physical behavior is perceived only as either "plausible" or "not plausible," not as a scale

Furniture elevation and clutter elevation were different from the other parameters in this study, in that they explicitly created non-physical environments such as objects sunken into the floor, objects floating in air, or objects interpenetrating one another. These parameters were perceived as less important than both scale parameters, ranking 3rd and 5th, respectively, in both subjective rankings and transition probability in Study 3. That said, where errors in other parameters seem to be perceived as "more plausible" or "less plausible," errors in these parameters seem to be either "plausible" or "not plausible." As a specific example, the median expenditure on clutter elevation was 25 out of 25 possible points, indicating that in a majority of trials, participants chose to spend their budget to remove all error from clutter elevation. Participants generally chose to leave some error in furniture elevation, but we believe this can be at least partially explained by the fact that, as mentioned in Section 4.5.2, furniture elevation errors present at ground level, while clutter elevation errors present closer to eye level.

7.5 Lights were considered to be not impactful, or even negatively impactful

Across all metrics, across Studies 2 and 3, the lighting parameters were last and least upgraded, and subjectively considered least important. We suggest three possible explanations for this. First is that in the real world, people adapt to wildly differing and rapidly changing lighting conditions with little difficulty. Despite occasional misperceptions, such as the "dress color illusion," we are generally able to perceive colors as stable even as lighting changes. It is possible that we have learned and/or evolved in such a way as to be able to separate (and ignore) the effects of lighting in favor of a stable world model. Second is that despite great advances, real time computer-generated lighting is still unable to capture many of the effects of lighting in the real world. It may be that even the "best" lighting conditions in these user studies were perceived as low-quality by participants, so they did not bother to spend time or points upgrading the lighting. Finally, the lighting conditions in the studies did not exactly match the lighting conditions in the real room. (This is discussed further in Section 8.) If participants were trying to exactly match the conditions they experienced in the real original room, rather than trying to match a "feeling of reality" in a broader sense, it makes sense that they would not have chosen to turn lights on if lighting conditions remained "non-matching." The research presented in this paper does not enable us to determine which of these three explanations is correct; further study is required.

8 Limitations and Future Work

Despite the progress represented by the results discussed in the previous section, there are several significant limitations that must be addressed in future work in this area. First, there are simply many factors that one could imagine would affect the design of and user experience in a virtual space that were not studied in this work. Due to the logistical constraints associated with running these user studies, we always

Table 6: Ranking of subjectively most important parameters (For readability, only factors preferred by ≥ 10% of participants are shown).

Ranking	Most important factors				
1st	**RoomScale (68%)**	FurnitureScale (15%)			
2nd	RoomScale (18%)	FurnitureElev (10%)	**FurnitureScale (70%)**		
3rd	RoomScale (10%)	**FurnitureElev (40%)**	FurnitureScale (15%)	ClutterPieces (20%)	
4th	FurnitureElev (15%)	**ClutterPieces (40%)**	ClutterElev (30%)	NONE (10%)	
5th	FurnitureElev (10%)	ClutterPieces (25%)	**ClutterElev (38%)**	SunLight (15%)	NONE (10%)
6th	FurnitureElev (10%)	LampLight (15%)	**SunLight (40%)**	CeilingLight (10%)	NONE (20%)
7th	**LampLight (30%)**	SunLight (18%)	CeilingLight (18%)	**NONE (30%)**	
8th	LampLight (25%)	SunLight (13%)	**CeilingLight (43%)**	NONE (20%)	

planned to include no more than eight parameters in the studies that became Studies 2 and 3. (We ultimately included only seven parameters in these studies, based on our observations in Study 1.)

Particularly, this work to date has only concerned itself with the perceived realism of static scenes. We explicitly did not attempt to measure the effects of object or character behavioral realism, which are certainly factors that impact a user's feeling of reality. In fact, existing research regarding coherence and Plausibility Illusion has focused almost entirely on behavioral coherence [4] [23] [26]. Integrating that work with our research is a very interesting avenue of future work, as the community works toward a more complete model of coherence.

Perhaps a more fundamental limitation is that many of these parameters might be inherently inseparable. (Consider the discussion of room length/room width/furniture scale possibly all representing a "relative scale" construct in Section 7.2.) A piece of furniture is neither realistic nor unrealistic in a vacuum; it receives spatial context, visual context, and use context from the room shell and other objects.

Another limitation of these studies is that, for feasibility reasons, in Study 3 we only considered "one-sided" errors. That is, all parameters were restricted to be strictly less than or equal to the veridical value. Even in this paper, it is clear that this restriction is not entirely valid; see Figure 5 for evidence that some parameters influence the sense of reality asymmetrically, such that less-than errors are more perceptually disturbing than greater-than errors, or vice versa. Allowing for both less-than and greater-than errors would significantly complicate the analysis, but this may be a necessary sacrifice to achieve improved validity.

We mentioned in Section 7.4 that the lighting conditions in the real room did not exactly match the lighting conditions available in the virtual room. When participants experienced the real room, the room was lit by the ceiling lights, as well as whatever natural light was coming through the blinds. This differs in at least three (potentially) important ways from the lighting conditions in the virtual room. First, participants never saw the real room lit with the lamps turned on. Second, the virtual ceiling lights did not actually light the room; turning on the ceiling lights in any of the studies only caused them to change color. This is because we were not able to take advantage of area lights in Unity due to our need to manipulate the models in real time. And finally, participants experienced the real room on different days, at different times of day. The virtual sun light did not change; it always appeared as bright mid-afternoon sunlight. Any or all of these may be worth revisiting as capabilities change.

Despite these limitations regarding the evaluation of lighting in our studies, it would be interesting to consider why our participants did not consider lighting to be important for their feeling of reality. In Section 7.4, we put forward three potential explanations. Evaluating these explanations is a rich area for further work.

Similarly, in Section 7.5, we proposed sensory impact and affordance hypotheses for why scale seems to be the most important percept for users to have a strong feeling of reality. Evaluating these hypotheses is also an avenue for future work.

Throughout this work, participants did not have full body representations in the virtual environment. Related work suggests that distance underestimation in virtual environments is reduced—distance perceptions are more accurate—when participants are embodied [17]. It would be interesting to see if our results regarding the importance of scale factors replicate with fully embodied users.

9 Conclusion

In this paper, we have presented the designs of and results from three user studies investigating how users perceive replicated virtual environments. Such replicas of real places are already in use in fields such as cultural heritage, real estate, and architectural design; their use will only increase as scanning and modeling technologies become more cheaper and more widely available.

These results represent a first attempt to measure which characteristics of a virtual space are most perceptually important to users, and include the fact that the scale of room components and large objects such as furniture are more important to user experience than other factors, such as lighting, which did not play a large role in users' feeling of reality in this study. (This may not be generally true, however; see our discussion in Sections 7.5 and 8.)

The investigation into users' experiences of replicated environments has only just begun, as evidenced by the discussion in the previous section. However, we are hopeful that the results included here can serve as the basis for and stimulate additional research into the specific issues faced when creating virtual replicas of real spaces.

Acknowledgments

The authors wish to thank Deborah Asabere, Shabi Mustafa, and Ahmed Salih for their considerable help in running these experiments. We would also like to thank our participants, without whom this work would not have been possible, and our anonymous reviewers, for their helpful feedback. This work was supported by a grant from Facebook.

References

[1] A. S. Azevedo, J. Jorge, and P. Campos. Combining EEG data with place and plausibility responses as an approach to measuring presence in outdoor virtual Environments. *Presence: Teleoperators and Virtual Environments*, 23(4):354–368, September 2014.

[2] R. M. Baños, C. Botella, A. García-Palacios, H. Villa, C. Perpiña, and M. Alcañiz. Presence and reality judgment in virtual environments: A unitary construct. *CyberPsychology & Behavior*, 3(3):327–335, July 2000.

[3] M. K. Bekele, R. Pierdicca, E. Frontoni, E. S. Malinverni, and J. Gain. A survey of augmented, virtual, and mixed reality for cultural heritage. *Journal on Computing and Cultural Heritage (JOCCH)*, 11(2):7:1–7:36, March 2018. doi: 10.1145/3145534

[4] I. Bergström, S. Azevedo, P. Papiotis, N. Saldanha, and M. Slater. The plausibility of a string quartet performance in virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 23(4):1352–1359, April 2017. doi: 10.1109/TVCG.2017.2657138

[5] S. Bouchard, S. Dumoulin, J. Talbot, A.-A. Ledoux, J. Phillips, J. Monthuy-Blanc, G. Labonté-Chartrand, G. Robillard, M. Cantamesse, and P. Renaud. Manipulating subjective realism and its impact on presence: Preliminary results on feasibility and neuroanatomical correlates. *Interacting with Computers*, 24(4):227–236, 2012. Special Issue on Presence and Interaction. doi: 10.1016/j.intcom.2012.04.011

[6] A. X. Chang, T. Funkhouser, L. Guibas, P. Hanrahan, Q. Huang, Z. Li, S. Savarese, M. Savva, S. Song, H. Su, J. Xiao, L. Yi, and F. Yu.

ShapeNet: An information-rich 3D model repository. Technical Report arXiv:1512.03012 [cs.GR], Stanford University — Princeton University — Toyota Technological Institute at Chicago, 2015.

[7] S. B. Gilbert. Perceived realism of virtual environments depends on authenticity. *Presence: Teleoperators and Virtual Environments*, 25(4):322–324, Fall 2017.

[8] L. Hernández, J. Taibo, A. Seoane, and A. Jaspe. Space perception in architectural visualization through immersive virtual reality. *Revista de EGA*, (18):253–261, 2011.

[9] J. S. Hvass, O. Larsen, K. B. Vendelbo, N. C. Nilsson, R. Nordahl, and S. Serafin. The effect of geometric realism on presence in a virtual reality game. In *2017 IEEE Virtual Reality (VR)*, pp. 339–340, March 2017. doi: 10.1109/VR.2017.7892315

[10] V. Interrante, B. Ries, and L. Anderson. Distance perception in immersive virtual environments, revisited. In *Virtual Reality Conference, 2006*, pp. 3–10. IEEE, 2006.

[11] V. Interrante, B. Ries, J. Lindquist, M. Kaeding, and L. Anderson. Elucidating factors that can facilitate veridical spatial perception in immersive virtual environments. *PRESENCE: Teleoperators and Virtual Environments*, 17(2):176–198, 2008.

[12] R. V. Kenyon, M. Phenany, D. Sandin, and T. Defanti. Accommodation and size-constancy of virtual objects. *Annals of biomedical engineering*, 36(2):342–348, 2008.

[13] R. V. Kenyon, D. Sandin, R. C. Smith, R. Pawlicki, and T. Defanti. Size-constancy in the cave. *Presence: Teleoperators and Virtual Environments*, 16(2):172–187, 2007.

[14] C. Lee, G. A. Rincon, G. Meyer, T. Höllerer, and D. A. Bowman. The effects of visual realism on search tasks in mixed reality simulation. *IEEE Transactions on Visualization and Computer Graphics*, 19(4):547–556, April 2013. doi: 10.1109/TVCG.2013.41

[15] M. Naseer, S. Khan, and F. Porikli. Indoor scene understanding in 2.5/3d for autonomous agents: A survey. *IEEE Access*, 7:1859–1887, 2019. doi: 10.1109/ACCESS.2018.2886133

[16] T. D. Nguyen, C. J. Ziemer, T. Grechkin, B. Chihak, J. M. Plumert, J. F. Cremer, and J. K. Kearney. Effects of scale change on distance perception in virtual environments. *ACM Trans. Appl. Percept.*, 8(4):26:1–26:18, 2008. doi: 10.1145/2043603.2043608

[17] L. Phillips, B. Ries, M. Kaeding, and V. Interrante. Avatar self-embodiment enhances distance perception accuracy in non-photorealistic immersive virtual environments. In *2010 IEEE Virtual Reality Conference (VR)*, pp. 115–118, 2010.

[18] M. Portman, A. Natapov, and D. Fisher-Gewirtzman. To go where no man has gone before: Virtual reality in architecture, landscape architecture and environmental planning. *Computers, Environment and Urban Systems*, 54:376 – 384, 2015. doi: 10.1016/j.compenvurbsys.2015.05.001

[19] E. D. Ragan, D. A. Bowman, R. Kopper, C. Stinson, S. Scerbo, and R. P. McMahan. Effects of field of view and visual complexity on virtual reality training effectiveness for a visual scanning task. *IEEE Transactions on Visualization and Computer Graphics*, 21(7):794–807, July 2015. doi: 10.1109/TVCG.2015.2403312

[20] R. S. Renner, B. M. Velichkovsky, and J. R. Helmert. The perception of egocentric distances in virtual environments—A review. *ACM Computing Surveys*, 46(2):23:1–23:40, December 2013. doi: 10.1145/2543581.2543590

[21] N. Saleeb. Effects of the differences between virtual and physical perception of space on building information modelling. *WIT Transactions on the Built Environment*, 149:21–32, 2015.

[22] R. Skarbez, F. P. Brooks, Jr., and M. C. Whitton. A survey of presence and related concepts. *ACM Computing Surveys*, 50(6):96:1–96:39, November 2017. doi: 10.1145/3134301

[23] R. Skarbez, S. Neyret, J. Frederick P. Brooks, M. Slater, and M. C. Whitton. A psychophysical experiment regarding components of the Plausibility Illusion. *IEEE Transactions on Visualization and Computer Graphics*, 23(4):1369–1378, April 2017. doi: 10.1109/TVCG.2017.2657158

[24] M. Slater. Place illusion and plausibility can lead to realistic behavior in immersive virtual environments. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 364:3549–3557, 2009.

[25] M. Slater, P. Khanna, J. Mortensen, and I. Yu. Visual realism enhances realistic response in an immersive virtual environment. *IEEE Computer Graphics and Applications*, 29:76–84, 2009. doi: 10.1109/MCG.2009.55

[26] M. Slater, B. Spanlang, and D. Corominas. Simulating virtual environments within virtual environments as the basis for a psychophysics of presence. *ACM Transactions on Graphics*, 29(4):92:1–92:9, July 2010. doi: 10.1145/1778765.1778829

[27] W. B. Thompson, P. Willemsen, A. A. Gooch, S. H. Creem-Regehr, J. M. Loomis, and A. C. Beall. Does the quality of the computer graphics matter when judging distances in visually immersive environments? *Presence: Teleoperators and Virtual Environments*, 13(5):560–571, 2017/07/12 2004. doi: 10.1162/1054746042545292

[28] R. B. Welch, T. T. Blackmon, A. Liu, B. A. Mellers, and L. W. Stark. The effects of pictorial realism, delay of visual feedback, and observer interactivity on the subjective sense of presence. *Presence: Teleoperators and Virtual Environments*, 5(3):263–273, Summer 1996. doi: 10.1162/pres.1996.5.3.263

[29] J. Yoon, E. Byun, and N. S. Chung. Comparison of space perception between a real environment and a virtual environment. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 44(5):515–518, 2020/08/23 2000. doi: 10.1177/154193120004400508

[30] I. Yu, J. Mortensen, P. Khanna, B. Spanlang, and M. Slater. Visual realism enhances realistic response in an immersive virtual environment - part 2. *IEEE Computer Graphics and Applications*, 32(6):36–45, Nov 2012. doi: 10.1109/MCG.2012.121