This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2025.3547241

The Role of Sensorimotor Contingencies and Eye Scanpath Entropy in Presence in Virtual Reality: A Reinforcement Learning Paradigm

Esen Küçüktütüncü, Francisco Macia-Varela, Joan Llobera, and Mel Slater, *Member*, IEEE

Abstract - Sensorimotor contingencies (SC) refer to the rules by which we use our body to perceive. It has been argued that to the extent that a virtual reality (VR) application affords natural SC so the greater likelihood that participants will experience Place Illusion (PI), the illusion of 'being there' (a component of presence) in the virtual environment. However, notwithstanding numerous studies this only has anecdotal support. Here we used a reinforcement learning (RL) paradigm where 26 participants experienced a VR scenario where the RL agent could sequentially propose changes to 5 binary factors: mono or stereo vision, 3 or 6 degrees of freedom head tracking, mono or spatialised sound, low or high display resolution, or one of two color schemes. The first 4 are SC, whereas the last is not. Participants could reject or accept each change proposed by the RL, until convergence. Participants were more likely to accept changes from low to high SC than changes to the color. Additionally, theory suggests that increased PI should be associated with lower eye scanpath entropy. Our results show that mean entropy did decrease over time and the final level of entropy was negatively correlated with a post exposure questionnaire-based assessment of PI.

Index Terms— Presence, virtual reality, sensorimotor contingencies, reinforcement learning, entropy, eye tracking, scanpath

1 Introduction

Virtual Reality systems were first invented in the 1960s, and came to public prominence in the 1990s, and then again in the past decade when new hardware and software were developed – with high performance and at a fraction of the cost of the 1990s systems. However, throughout this time some of the scientific questions have remained the same. One of the principle issues is the illusion of 'presence' first discussed in the early 1990s [1]. Presence is the illusion of being in the virtual environment - 'being there' – the scientific puzzle being that in spite of obvious differences to reality (for example, the level of visual realism, the poor resolution of displays compared to normal vision in the physical world, the relatively imprecise representations of physics if any, the non-realistic animations applied to humanoid characters, amongst others) people nevertheless tend to experience a strong illusion of presence. They feel and act as if the virtual world that is displayed to their senses is the actual world in which they are participating. So, the first question concerns understanding the causes of presence. The second issue is how to measure it. Typically, this is carried out with a questionnaire or physiological measures, for example [2-4]. While there are other methods, each has its own problems. For example, measuring behavioral and physiological responses to events in the VR [5] requires specific events to be deliberately added that may cause these responses, which may not be appropriate to the application.

Slater [6] put forward a theory for the occurrence of presence as 'being there' – referred to as 'Place Illusion' (PI) the illusion of being in the virtual place. The idea is that to the extent that the VR system affords perception through natural sensorimotor contingencies [7] so the probability is enhanced that participants will experience PI. Sensorimotor contingencies refer to the systematic relationship between body movements and sensory experiences, so that perception is not based on just passive reception of sensory stimuli but is an active process. For example, moving the upper body and head to look behind objects, reaching out, bending down to look underneath something, looking around by turning the head and the eyes, amongst others. If the VR system allows perception through the use of the body similarly to how this is carried out in physical reality, then the simplest hypothesis for the brain to adopt is that what you perceive signifies where you are. This not only refers to the act of moving the body, but how perception changes must also match expectations from reality. For example, if you turn your head to the right then your view of the world (or sounds heard) must update accordingly. You would expect to see near objects in depth, but less depth perception for objects that are further away. If you move your head close to an object and it dissolves into pixels, then this is a failure of sensorimotor contingencies.

VR devices do, to varying extents, support these types of visual and auditory sensorimotor contingencies through the 6 degrees of freedom head tracking, and higher resolution displays with wide field-of-view, and spatialized sound. However, this is a difficult theory to test, because if sensorimotor contingencies are altered for an experiment, then there is a high probability that participants will experience simulator sickness. For example, if we introduced latencies into the head tracking, then such sickness would be likely to occur, although reduced latency has been associated with higher reported presence [8].

The second problem, how to approach the measurement of presence, without reliance solely on questionnaires, or artefacts introduced for physiological measures, is an important issue for the engineering of presence-inducing systems. A method based on the impact on presence of real-time changes in various factors in the scenario (such as the level of visual realism, the field-of-view of the display, and others) was introduced by Slater, et al. [9]. This method offers the participant a number of system configurations, for example: wide field-of-view or narrow field-of-view, realistic visual rendering or simple visual rendering, vision from first- or third-person perspective, having a virtual body or not having one. The original idea of the method is that at any time the virtual environment is constituted by k such factors $F_1, F_2, \ldots, F_k$ forming a

Esen Küçüktütüncü is with the Institute of Neurosciences of the University of Barcelona, Spain. E-mail: esenkucuktutuncu@ub.edu

Francisco Macia-Varela is with the University of Barcelona, Spain. E-mail: frankmmacia@gmail.com, Current address: Full-Body Interaction Lab, Universitat Pompeu Fabra, Barcelona, Spain.

Joan Llobera is with the Artanim Foundation, Geneva, Switzerland. E-mail: joan.llobera@artanim.ch

Mel Slater is with the Institute of Neurosciences of the University of Barcelona, Spain. E-mail: melslater@ub.edu, Corresponding Author.

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 IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2025.3547241

configuration C. If each factor is binary then there are 2^k configurations. For example, F_1 is either 0 (monovision) or 1 (stereovision). The participant first enters the environment with all factors at their highest level C_H (e.g., F_{H1} is stereovision) and pays attention to the corresponding sensation of PI. They are also trained on the meaning of the various factors and how to change their levels. Then they enter the environment now with all factors at their lowest or at randomly chosen levels C_0. At various times during their exposure they are able to change the level of one of the factors (under a cost constraint) in order to attain the sensation of PI that they had during their exposure to C_H. Each such change leads to a transition $C_i \rightarrow C_j$. They continue to make transitions until they declare a *match* with their original sensation of PI while they experienced C_H. From this, over all participants, we obtain a large number of transitions from which we can compute a $2^k \times 2^k$ Markov probability matrix $\boldsymbol{P}$ with elements p_{ij}, the probability that conditional on the configuration being C_i that the next configuration would be C_j (including the possibility that $i = j$). From these data various interesting probabilities can be computed, for example the equilibrium state of the Markov chain, and conditional probabilities such as $P(match\,|C_i)$, the probability of a match being declared given that the current configuration is C_i. We refer to this method as the Multi-Modal Matching (3M) method.

2 Related Work

2.1 Adaptive Multi-Modal Matching

The 3M method was extended in [10] where instead of participants directly choosing which element in a configuration to change, changes were offered to them by a Reinforcement Learning (RL) agent. The agent initially offers configuration changes at random, but then, depending on participant choices (to accept a proposed change or not) eventually converges, meaning that it has estimated probabilities of participants accepting proposed changes, and thus is unlikely to propose changes with low probability of acceptance. Across many participants a consistent pattern of selected configurations emerges. We refer to this as Adaptive 3M (A3M).

RL was chosen because it allows the underlying agent to learn by receiving feedback based on the participant choices. The RL agent can model the decision-making process where each participant's actions (accepting or rejecting changes) provide rewards or penalties to the RL agent, resulting in the agent optimizing the presentation of factors that lead to the desired outcome of the participant. Over time, the agent learns which combinations of factors work best for each individual. Compared to other methods, like supervised learning, RL is preferable because it does not require labeled data, which is typically unavailable in scenarios where participants' preferences are dynamic and individualized. Supervised learning, for example, would require predefined examples of optimal choices, which may not exist. Evolutionary algorithms or optimization techniques could find the best factor combinations, but they do not adapt continuously based on participant interaction. The ability of RL to iteratively adapt to individual preferences, and optimize over time, makes it a suitable choice.

2.2 Relation to eye scanpath entropy

During the exposure, since participants would be selecting or rejecting changes to the configuration in order to reach a higher level of PI, the likelihood of PI occurring on the average across participants, should be increasing. It has previously been proposed, and supported by experimental evidence, that the entropy of gaze direction is inversely associated with PI. This is based on the idea that in stable perceptions the eye scanpath involves repeatedly moving between a few salient points in the environment, with perception highly influenced by top-down rather than bottom-up processing [11]. Internal top-down models actively drive vision, rather than what is in the environment. For example, when observing ambiguous figures, the eye scanpath is different depending on whether the viewer perceives one interpretation or another, even though the figure itself is fixed. This was suggested as one mechanism by which virtual reality 'works' even in very basic and simply rendered scenes – people perceive their internal models from a set of minimal cues provided by the environment [12]. Based on this, the theory was proposed that presence as PI is associated with eye scanpath entropy [13] – that the onset of PI is signified by a *reduction* in entropy.

Entropy measures the degree of disorder in a system. Note that entropy is not the same as variance – the variance of eye scanpaths might be high, but if the same set of visually salient areas are visited regularly and repeatedly then the entropy is low. Jordan and Slater [13] showed this by exposing participants to a gradually forming environment with its component triangles randomly becoming visible over time, until finally the environment was perceived by participants as their standing on top of a high column. Head-gaze movement scanpath entropy decreased at approximately the same time as skin conductance measuring arousal increased, suggesting that at the moment that the percept was formed, PI was established. Participants experienced arousal on realizing that they were standing on top of the column (higher skin conductance and lower entropy). In one control group where participants finally found themselves to be standing at ground level, the entropy decreased at approximately the same time as in the experimental condition but the skin conductance was unchanged.

3 Materials and Methods

3.1 Overview

In the experiment reported here we used A3M with respect to the problem of assessing factors that contribute to PI. However, all but one of the factors included relate to sensorimotor contingencies. Participants were told to make choices that would improve their sense of being there (PI) in the environment depicted by the VR. Unknown to the participants of course, the objective was to study whether the RL would converge to a configuration that is likely to include factors that involve changing sensorimotor contingencies – which were selected to minimize the risk of simulator sickness – but not include the factor that is unrelated to sensorimotor contingencies.

Moreover, we used eye tracking to obtain the eye scanpaths of participants and showed that on the average the entropy decreased with time, and that the final entropy level was linearly negatively, or possibly quadratically related with PI as assessed from a questionnaire administered after the VR exposure.

3.2 The scenario

In the main phase of the experiment participants were in a VR that depicts a string quartet playing some classical music, reusing the same scenario as in [14]. At various times during the performance, they were offered the chance to change one factor in the scenario configuration or leave it as it is. This continued for a maximum of 20 minutes. The possible changes for the participant to make were chosen by a Reinforcement Learning algorithm. When the participant had rejected 8 successive proposed changes, and provided that at least 10 minutes had elapsed, this was taken as a sign of convergence, and the scenario ended. The scenario is illustrated in Figure 1. See also Supplementary Video S1 (a higher resolution version is available on https://youtu.be/K8sXlVrXZQs).

3.3 Informed consent

Participants were recruited from the University of Barcelona campus. Thirty were recruited in total but due to technical failures there are 26 full results for the Reinforcement Learning and 29 for

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2025.3547241

the eye scanpath entropy analysis. The experiment was approved by the Comisió de Bioética de la Universitat de Barcelona (IRB00003099), and procedures were carried out in conformance with the approved procedures. Participants gave written and informed consent and were informed that they could leave the experiment at any time without giving reasons or losing benefits. They were paid 10 euros as compensation for participating to the experiment.

The experiment took place at the University of Barcelona (UB), Mundet Campus. It consisted of one session. Information about the experiment was told verbally to the participant by the experimenter and also explained in a Qualtrics Questionnaire (www.qualtrics.com). After reading the information sheet the participants were informed about their freedom to withdraw from the experiment at any time and were asked to sign the consent form.

3.4 Factors Manipulated

The RL algorithm offered changes to the following factors:

1. Vision: (0) monovision (1) 3D stereovision
2. Parallax: (0) 3 degrees of freedom head-tracking (orientation only) (1) 6 degrees of freedom head-tracking (orientation and translation through space).
3. Audio: (0) Mono audio (1) Spatialized audio.
4. Resolution: (0) 1282 × 672 (1) 3664×1920, the normal resolution of the HMD
5. Color: (0) Original color (1) Alternative color.

Items 1-4 are different aspects of sensorimotor contingencies, whereas item 5 is not. Perception of color is not dependent on body movements so it is not a sensorimotor contingency. However, in the case of stereovision head and eye movements will change how an object is perceived compared to monovision. With 6 degrees of freedom head-tracking it is possible, for example, to use head translations to look behind an object that to see another obscured object whereas that is not possible with 3 degrees of freedom head-tracking. The perception of audio may change with head movements in the case of spatialized audio but not in the case of mono audio. The appearance of an object may change as the head moves close to it in the case of lower compared to higher resolution. However, the perceived color of an environment function is not a function of body movements in the same way as the previous examples. Perception of color remains invariant under different illumination, for example, which is referred to as 'color constancy' – for a review see [15].

Therefore, it was expected that on items 1-4, changes from (0) to (1) would be far more likely to be accepted than (1) to (0), whereas in item 5 there will be an approximately equal chance of selecting (0) or (1), since these are not related to sensorimotor contingencies.

The specific sensorimotor contingencies 1-4 were chosen for two reasons. The first is that changes to these would be unlikely to cause simulator sickness. The second is that, for the problem we wished to explore in this paper, we could have chosen *any* set of 4 sensorimotor contingencies – we only wanted to see the chosen set were treated by participants differently from the one non-sensorimotor contingency (color). The 4 chosen were straightforward to implement and easy to understand.

There are 32 configurations and after each suggestion by the RL there is a potential transition from one configuration to another. The configurations are shown in Table 1 and Supplementary Table S1. A transition occurs when there is a change from one configuration in the set (0,…,31) to another in the same set. This also includes null transitions where there is no change.

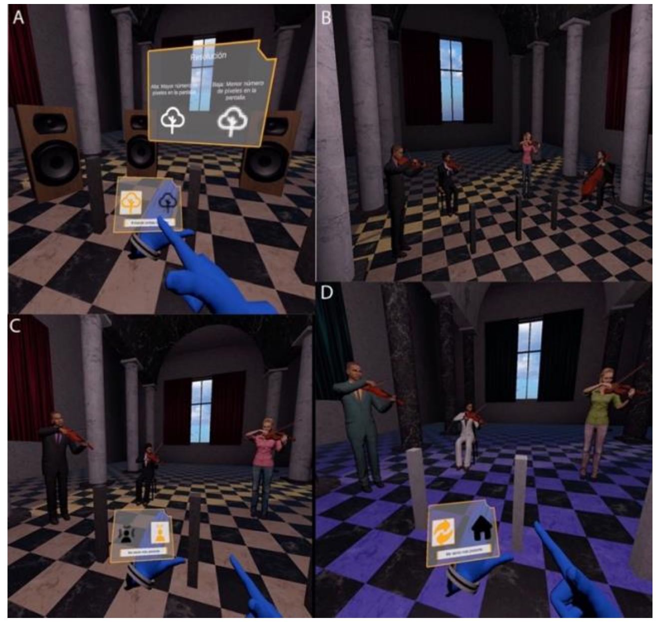


Figure 1 - The string quartet scenario. (A) The learning phase where participants were in a room showing some loudspeakers and practiced changing the settings, in the case shown, the resolution. (B) An overview of the scenario – this image has been slightly vertically stretched for alignment purposes. (C) The participant chooses whether or not to make a change to the audio. (D) The scene is shown in the alternate color scheme and the participant chooses whether or not to accept this change.

3.5 Equipment

For this experiment the Pico Neo 3 Eye head-mounted display (HMD) was used, with a resolution of 3664 × 1920, and 773 pixels per inch, and it has field of view is 98^o. The headset has a refresh rate of 90Hz and is equipped with 2 eye tracking cameras, each being 400 × 400 running at 90 Hz. The controllers of the headset were used for interaction, which have 6DoF optical positioning and linear resonant actuators. For audio equipment Skullcandy Riff Wireless Headphones were used in order to ensure comfort and immersion.

3.6 Procedures

The experimenter explained the stages of the study and gave the participant the information sheet and consent form to sign. They then were asked to complete a questionnaire. The questionnaire was divided into two parts, prior to entering the virtual environment (Pre Questionnaire in Supplementary Table S2) which was concerned with demographic and background issues, and after completing the VR experience (Post Questionnaire in Supplementary Table S2).

The questionnaire included demographic questions, level of expertise with programming in VR, and familiarity with VR experiences.

Then the experimenter explained the idea of 'Place Illusion' – the feeling to be in the virtual environment. They were told about the 5 sets of possible binary changes that that might be made, and icons were shown to the participant to indicate each change within the VR (Table 2). Then participants donned the head-mounted display and in the first phase of the experiment they were immersed in a reduced environment that did not include the virtual musicians (Figure 1A). There, each of the configurations were explained with a written description, and they were able experience each of the 2 possible choices using a pop-up panel on their left arm (Figure 1A). Once they had tried both of the options, they were asked, when ready, to press the '*I understand both options'* button on the virtual wristwatch, and then proceed to the next factor.

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2025.3547241

Table 1 – Extract from the 32 configurations. The full table can be found in Supplementary Table S1

No.	Meaning	Binary representation
0	Monovision - No Parallax - Mono Audio - Low Resolution - Original Color	00000
1	Monovision - No Parallax - Mono Audio - Low Resolution - Alternative Color	00001
2	Monovision - No Parallax - Mono Audio - High Resolution - Original Color	00010
3	Monovision - No Parallax - Mono Audio - High Resolution - Alternative Color	00011
4	Monovision - No Parallax - Spatial Audio - Low Resolution - Original Color	00100
5	Monovision - No Parallax - Spatial Audio - Low Resolution - Alternative Color	00101
6	Monovision - No Parallax - Spatial Audio - High Resolution - Original Color	00110
7	Monovision - No Parallax - Spatial Audio - High Resolution - Alternative Color	00111
8	Monovision - Parallax - Mono Audio - Low Resolution - Original Color	01000
9	Monovision - Parallax - Mono Audio - Low Resolution - Alternative Color	01001
10	Monovision - Parallax - Mono Audio - High Resolution - Original Color	01010
11	Monovision - Parallax - Mono Audio - High Resolution - Alternative Color	01011
12	Monovision - Parallax - Spatial Audio - Low Resolution - Original Color	01100
13	Monovision - Parallax - Spatial Audio - Low Resolution - Alternative Color	01101
14	Monovision - Parallax - Spatial Audio - High Resolution - Original Color	01110
15	Monovision - Parallax - Spatial Audio - High Resolution - Alternative Color	01111
16	Stereovision - No Parallax - Mono Audio - Low Resolution - Original Color	10000
…	…	…
31	Stereovision - Parallax - Spatial Audio - High Resolution - Alternative Color	11111

After this learning period was completed, the scene faded into the same environment but now with 4 musicians playing in front of them, which started the main phase. The experimental factors were randomized at the beginning of this phase, so each participant started with a randomly chosen configuration from the list in Table 1. Their task was to listen to and watch the musical performance. At uniformly random intervals between 35 to 45 seconds, they would feel the left hand controller vibrate and a corresponding sound effect, where a virtual panel would pop up on their wrist and give them the option to either change the current configuration or leave it as it is. During this task, the interface presented choices to the participants using the method presented during the pre-immersion phase, except that the button 'I understand both options' was replaced by '*I feel more present'*.

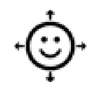
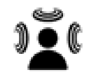
They could decide to continue with the current configuration or change it to the alternative offered. For example, if the scene was in low resolution, they could either keep it the that way by selecting the low-resolution option or change it to high resolution. The factor that was displayed was selected by the RL agent and it continued until convergence, or the 20 minute maximum time, with the application ending by fading out.

Then, they completed the post-experiment questionnaire and were asked to order the experimental factors from the most important to the least. They were also asked to leave comments regarding the experiment. Overall, the entire process lasted between 35 to 45 minutes with 20 minutes being the VR exposure.

3.7 The Reinforcement Learning Method

The RL method we used was the same as in [10]. Positive rewards were given when the RL agent proposed a configuration change that the participant accepted. Hence, the agent was designed to find the preferred configuration of the participant, using the prompts shown in Table 2. The only difference with [10] was that there were 5 different configuration factors, instead of 4. In addition, the total number of iterations per subject would different, since it was based on whether the preferences of the participant had converged (see Section 3.2).

Table 2 – Icons used to describe each possible change. Level 0 represents the lower SC and level1 the higher except for Color.

Experimental factor	Level 0	Level 1
Vision		
Head-tracking		
Audio		
Resolution		
Color		

Just as in [10] we used Q learning from Sutton and Barto [16] (Chapter 6). Each combination of factors listed in Table 1 was associated with a state of the RL algorithm. For every configuration there were 5 possible actions associated, one for each configuration change. As in [10], to account for possible intersubject differences in the configuration preferences, all values in the Q-table were initialized to 0 for each new participant.

For the RL method, we chose the same configuration values as in [10]: $l = 0.2$, for the learning rate; $\gamma = 0.15$, as the discount factor. All values in the Q-table were initialized to 0. In addition, since it was possible that different participants had different preferences, and the numbers of episodes per participant were variable, we chose to run the RL algorithm separately for each participant.

3.8 Eye Scanpath Entropy

Eye scanpaths were obtained by analyzing the raw data from the Tobii Eye-tracking system in the HMD. The system consists of illuminators placed in a ring-like structure around the HMD lenses, that create a pattern of light reflections on the participants' eyes. This reflection pattern is captured by high-resolution cameras and processed in real-time by machine learning algorithms. This process occurs at a rate of 90 Hz. This raw eye direction data is converted into a three-dimensional vector through the HMD's SDK, and can be used with the head position and rotation data from the HMD in order to calculate where a participant is looking in the virtual space at any given point.

There were two predictions with respect to eye scan paths:

1. Mean entropy over all participants would decrease over time (with increasing PI due to manipulation of the factors).

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2025.3547241

2. There would be a negative correlation between the final level of entropy per participant and their subjectively reported level of PI as measured by the questionnaire after the experience.

The entropy was calculated as:

$$e = -\sum_{i=1}^{s} p_i \log(p_i)$$

where there are s possible 'states' and p_i is the probability of being in the ith state. The states are defined quite straightforwardly as a partition of the space in front of the participant into equal angle intervals (Figure 2). Hence p_i is the proportion of times that the target of the eye tracking entered into the ith state during the time period concerned. For example, if the eyes hardly moved and stayed in one interval during the time period, then $e = 0$. If the eyes moved uniformly randomly across all segments, then the probabilities $p_i = 1/s$ and $e = \log(s)$.

The overall time is also partitioned into successive segments, and entropy calculated for each participant for each segment.

3.9 Statistical Methods

We first present results descriptively in order to provide an easy to follow overview of the outcomes. We use Bayesian methods for formal statistical analysis. Using classical null hypothesis significance tests would run into the problem that there would be many tests, which undermines the meaning of the significance level due to these multiple comparisons. Although there are *ad hoc* methods to overcome this problem (such as Bonferroni corrections) there is no such problem with Bayesian methods since there is one overall model that includes all parameters, and as many probability statements as required can be drawn from their posterior joint distribution without diminishing their validity. In order to carry out the analysis we used the Stan statistical programming language [17] in RStudio (https://posit.co/download/rstudio-desktop/) with the rstan interface (https://mc-stan.org/users/interfaces/rstan). Some of the graphs were constructed with Stata (www.stata.com) (version 16.1).

4 Results

4.1 Participant statistics

The characteristics of the participants are shown in the Pre Questionnaire section of Supplementary Table S2. There were 5 more females than males, the majority were aged between 18 and 34 and they had very little prior experience of VR. They had little technical knowledge (a low level of computer programming for VR).

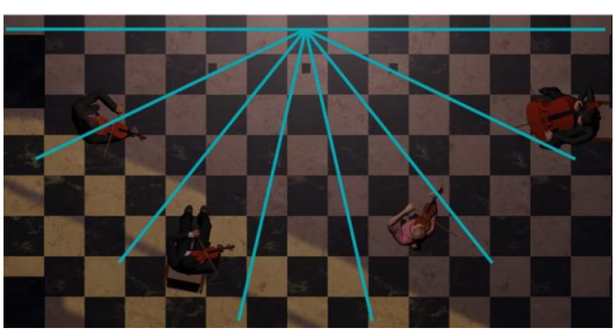


Figure 2 – Partition of the space into equal angle sectors.

4.2 Ordering of the factors

Overall participants were confident in their decisions in responses to requests for change, where the median level of confidence was 6 (out of a maximum 7) with interquartile range 2 (the variable *decisions* in Supplementary Table S2). Figure 3 shows box plots of the subjective ordering of the 5 factors with respect to their importance. It is clear that color is considered the least important, and resolution the most important, followed by vision.

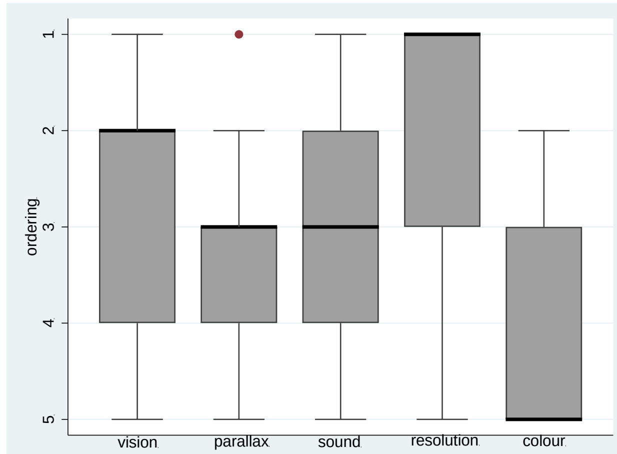


Figure 3 – Box plots of the ordering of the importance of the factors from the questionnaire. The thick black horizontal lines are the medians, the boxes are the interquartile ranges (IQR), and the whiskers range from lower quartile – max(smallest value,1.5*IQR), and upper quartile + min(max value, 1.5*IQR). Points outside these limits are shown individually.

4.3 The Place Illusion questions

The box plots for the PI questions (Supplementary Table S2) are shown in Figure 4. It can be seen that the subjective level is high, with all medians at least 5, with the highest being *virtualplace*. The variable *overall* refers to the medians across the 4 questionnaire variables. The entire IQR for this is greater than the mid-point of the scale.

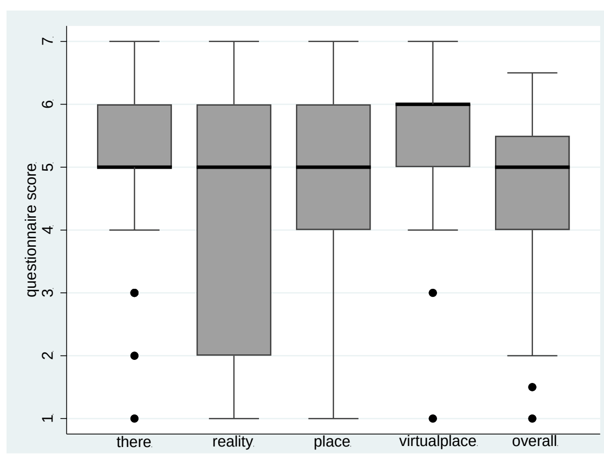


Figure 4 – Box plots for the PI questions from Supplementary Table S2. Overall refers to the median of the 4 questions.

4.4 Reinforcement Learning

The mean ± SD of the number of proposals per individual for changes was 16.7 ± 4.4. There were 433 transitions over the 26 participants. Table 3 shows the basic outcomes. For example, with respect to the Vision factor, when the current configuration included the Monovision state and a change was proposed there were 6 rejections, and 10 acceptances of the change. When it included the Stereovision state, there were 20 rejections and 2 acceptances.

Table 3 also shows the different responses to a proposal for a change by the different factors. For example, in the case of Vision, when the current configuration included Monovision 63% of the

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2025.3547241

proposals for change were accepted, but when the configuration included Stereovision only 9% of proposals for a change were accepted. It is notable that a configuration that included Monosound always transitioned to one that included Stereosound. With respect to Color, proposals for change had the lowest proportions of acceptances.

Further details of the RL results are shown in Supplementary Table S4.

Table 3 – Frequency distribution of the responses to a proposal to change the current configuration.

Current State	Change…		Proportion accepted	%
	rejected	accepted		
Vision				
Mono	6	10	10/16	63
Stereo	20	2	2/22	9
Parallax				
3 d.f.	15	16	16/31	52
6 d.f.	45	8	8/53	15
Sound				
Mono	0	15	15/15	100
Stereo	69	2	2/71	3
Resolution				
Low	4	11	11/15	73
High	78	1	1/79	1
Color				
Original	49	24	24/73	33
Alternative	30	28	28/58	48

Table 4 – The frequency distribution of the terminating configurations in decreasing order of frequency

No.	Config.		Freq	Prop.
30	11110	Stereovision - Parallax - Spatial Audio - High Resolution - Original Color	9	0.35
31	11111	Stereovision - Parallax - Spatial Audio - High Resolution - Alternative Color	5	0.19
14	01110	Monovision - Parallax - Spatial Audio - High Resolution - Original Color	4	0.15
22	10110	Stereovision - No Parallax - Spatial Audio - High Resolution - Original Color	3	0.12
15	01111	Monovision - Parallax - Spatial Audio - High Resolution - Alternative Color	2	0.08
5	00101	Monovision - No Parallax - Spatial Audio - Low Resolution - Alternative Color	1	0.04
6	00110	Monovision - No Parallax - Spatial Audio - High Resolution - Original Color	1	0.04
23	10111	Stereovision - No Parallax - Spatial Audio - High Resolution - Alternative Color	1	0.04

4.5 Terminating Configurations

Table 4 shows the frequency distribution of the terminating configurations. The highest frequency configuration is the one that contains all of the sensorimotor factors but not the color. Note that Spatial Audio appears in all 8, High Resolution appears in 7/8 and all the rest appear in 4/8.

4.6 Markov Chain Analysis

Every proposed change results in a transition, even if it is not to accept, in which case the configuration is unchanged. From the collection of all transitions across all participants we can build a Markov Transition matrix $\boldsymbol{P}$, where the elements

$$p_{ij} = P(transition\ to\ configuration\ j\ |configuration\ i)$$

the probability of transitioning from configuration i to j. This is a 32×32 sparse matrix, given the small number of trials relative to the total number of transitions possible. Moreover, the matrix shows that some configurations were never reached, and there was never a transition from them. These eliminated configurations are shown in Table 5. This reduces the size of the matrix to 27×27.

Table 5 – Eliminated Configurations – from which there are no transitions and which were never reached from other states.

8	01000	Monovision - Parallax - Mono Audio - Low Resolution - Original Color
17	10001	Stereovision - No Parallax - Mono Audio - Low Resolution - Alternative Color
24	11000	Stereovision - Parallax - Mono Audio - Low Resolution - Original Color
25	11001	Stereovision - Parallax - Mono Audio - Low Resolution - Alternative Color
26	11010	Stereovision - Parallax - Mono Audio - High Resolution - Original Color

Suppose $\boldsymbol{\pi}$ is a $1 \times\ 27$ row-vector of probabilities of being in the configurations. Then $\boldsymbol{\pi P}$ is the probability distribution over the configurations after a single transition. In general, $\boldsymbol{P}^k$ for any positive integer k is the k-step transition matrix, i.e., the $(i, j)th$ element is the probability of being in configuration j, k transitions after starting in configuration i. Hence $\boldsymbol{\pi P}^k$ is the probability distribution over the configurations after k transitions. For increasing values of k we can consider whether $\boldsymbol{\pi P}^k$reaches a stable state, in other words an equilibrium distribution has been reached so that further transitions make no difference to this probability distribution. The equilibrium distribution can be found by solving:

$$\boldsymbol{\pi P} = \boldsymbol{\pi}$$

reflecting the idea that if $\boldsymbol{\pi}$ is the equilibrium distribution then the application of a further transition matrix does not change it.

Table 6 shows the equilibrium probability distribution over the configurations. There are 6 configurations with probabilities of at least 0.05. Spatial Audio and High Resolution appear in all 6, Stereopsis and Parallax in 4, and Alternate Color in 3. It can be seen that the sensorimotor contingency variables dominate, although Alternative Color does appear in the second place. We cannot know whether this distribution is by chance, and whether the choice or not of the color change is different from the rest of the factors, though from this table and Table 4 it would seem so. Next, we turn to a formal statistical analysis to resolve these issues.

4.7 Statistical analysis

During the course of the VR experience participants made choices to *accept* or *reject* the offer of a change. We use as a response variable *accept*, which is 0 when the proposed change is rejected, and 1 when the proposed change is accepted. These offers were in the contexts shown in Table 7.

For example, the variable $V_0 = 1$ when a change to the Vision factor is offered and the current configuration includes Monovision, and $V_0 = 0$ otherwise. Or $R_1 = 1$ when a change to resolution is

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2025.3547241

offered, and the current configuration includes High Resolution. We use these as predictor variables for $accept$. If the idea that sensorimotor contingencies are at the basis of PI is valid then we would expect that the X_0 variables would be associated with an increase in the probability of $accept = 1$ (e.g., participants would be more likely to accept a change from monoaudio to spatialaudio) and the X_1 variables would be associated with a decrease in the probability of $accept = 1$ (e.g., they would be less likely to accept a change from Stereovision to Monovision).

Table 6 –Probability distribution of the equilibrium sorted by decreasing probability. Configurations not shown have probability 0.

No.	Config	Meaning	Prob
30	11110	Stereopsis - Parallax - Spatial Audio - High Resolution - Original Color	0.476
31	11111	Stereopsis - Parallax - Spatial Audio - High Resolution - Alternative Color	0.210
22	10110	Stereopsis - No Parallax - Spatial Audio - High Resolution - Original Color	0.090
14	01110	Monocular - Parallax - Spatial Audio - High Resolution - Original Color	0.066
15	01111	Monocular - Parallax - Spatial Audio - High Resolution - Alternative Color	0.064
23	10111	Stereopsis - No Parallax - Spatial Audio - High Resolution - Alternative Color	0.053
6	00110	Monocular - No Parallax - Spatial Audio - High Resolution - Original Color	0.011
7	00111	Monocular - No Parallax - Spatial Audio - High Resolution - Alternative Color	0.008
19	10011	Stereopsis - No Parallax - Mono Audio High Resolution - Alternative Color	0.005
27	11011	Stereopsis - Parallax - Mono Audio - High Resolution - Alternative Color	0.004
21	10101	Stereopsis - No Parallax - Spatial Audio - Low Resolution - Alternative Color	0.003
29	11101	Stereopsis - Parallax - Spatial Audio - Low Resolution - Alternative Color	0.003
10	01010	Monocular - Parallax - Mono Audio - High Resolution - Original Color	0.002
18	10010	Stereopsis - No Parallax - Mono Audio - High Resolution - Original Color	0.002
20	10100	Stereopsis - No Parallax - Spatial Audio - Low Resolution Original Color	0.002
28	11100	Stereopsis - Parallax - Spatial Audio - Low Resolution - Original Color	0.001

Table 7– Predictor variables for accept

Long variable name	Short Name	Factor	Meaning
monovision	V_0	Vision	Mono
stereovision	V_1	Vision	Stereo
noparallax	P_0	Parallax	No Parallax
parallax	P_1	Parallax	Parallax
monoaudio	A_0	Audio	Mono audio
spatialaudio	A_1	Audio	Spatial audio
lowres	R_0	Resolution	Low resolution
highres	R_1	Resolution	High Resolution
origcolor	C_0	Color	Original color
alternativecolor	C_1	Color	Alternative color

Note that we cannot include all 10 variables simultaneously because of aliasing between them (knowing the values of 9 of them can predict exactly the 10th) therefore we consider the '0' and '1' variables in two equations within the same overall model.

In order to assess the influence of these variables on *accept*, we use a standard Bernoulli logit model. Let $accept_i$ be a binary response variable that indicates that the proposal was accepted and and 0 when it was rejected, over all proposals made ($i = 1,2, \dots, n = 443$).

$$P(accept_i = 1) = \theta_i$$
$$P(accept_i = 0) = 1 - \theta_i \tag{1}$$

where θ_i is the probability of the ith response being *accept*. To relate θ_i to the variables in Table 7 above we use the linear predictors:

$$\eta_{0i} = \beta_{00} + \beta_{01}V_{0i} + \beta_{02}P_{0i} + \beta_{03}A_{0i} + \beta_{04}R_{0i} + \beta_{05}C_{0i}$$

$$\eta_{1i} = \beta_{10} + \beta_{11}V_{1i} + \beta_{12}P_{1i} + \beta_{13}A_{1i} + \beta_{14}R_{1i} + \beta_{15}C_{1i} \tag{2}$$

In the Bernoulli logit model, the relationship between (1) and (2) is through the logit link, i.e.,

$$\log\left(\frac{\theta_i}{1-\theta_i}\right) = \eta_i$$

or equivalently,

$$\theta_i = \frac{1}{1 + \exp(-\eta_i)} \tag{3}$$

where η_i represents η_{0i} or η_{1i}.

This also guarantees that θ_i is in the range 0 to 1 for all possible values of η_i. The interpretation of β_j is that it is the change in log-odds of *accept* compared to *reject* for a unit change in the corresponding variable, for all else held constant.

Let the prior distributions for the β parameters be *normal(mean=0, SD=10)*, which results in prior 95% credible intervals ±20, i.e., prior to utilizing the observed data there is 0.95 probability of being within these limits. These are wide intervals corresponding to weakly informative priors [18]. The posterior distributions are summarized in Table 8. Notice that the posterior 95% credible intervals are considerably narrower than the prior intervals, reflecting the impact of the data.

It can be seen that there are strong positive associations between *accept* and the '0' variables indicating that when the configuration included these, acceptance of a change was likely. This also holds for original color. However, for the '1' variables there is a strong negative association between these variables and *accept* indicating that when a configuration included these settings a change back to the '0' setting was unlikely. For example, a change in vision would be likely to be accepted if the configuration included monovision, but rejected if it included stereovision. However, and crucially, in the case of *altcolor* there is a low level of evidence of a negative association with *accept* (the probability is 1 – 0.367 = 0.633) whereas for all other variables the probability is 1).

It is likely that when the configuration included *originalcolor* that participants may have experimented with changing it, hence the high probability of an association (0.999). However, if they had changed it, they would tend not to change it back again, hence the lower probability associated with *altcolor*. Moreover, from the posterior distributions of the β_{0j} parameters, we find that the posterior probabilities $\beta_{05} < \beta_{0j}, j = 1,2,3,4$ is at least 0.966 in every case, meaning that the *originalcolor* variable has the lowest effect size.

The factor *monoaudio* also clearly has the greatest effect size since the posterior probability that $\beta_{03} > \beta_{0j}, j = 1,2,4,5$ is at least 0.998 in each case (this can also be seen through the credible intervals). Participants were most likely to change from *monoaudio* to *spatialaudio*, which is not surprising given the nature of the scenario.

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2025.3547241

Table 8– Summary of the posterior distributions of the parameters showing the mean and standard deviations of the distribution and the 95% credible intervals. Prob > 0 are the probabilities that the parameters are positive.

Parameter	Coefficient of	Mean	SD	2.5%	97.5%	Prob > 0
β_{00}		-1.77	0.17	-2.10	-1.46	0.000
β_{01}	monovision	2.31	0.54	1.27	3.39	1.000
β_{02}	noparallax	1.83	0.40	1.05	2.62	1.000
β_{03}	monoaudio	11.11	5.10	4.38	23.41	1.000
β_{04}	lowres	2.85	0.62	1.70	4.11	1.000
β_{05}	origcolor	1.04	0.30	0.45	1.63	0.999
β_{10}		0.02	0.16	-0.29	0.34	0.560
β_{11}	stereovision	-2.56	0.85	-4.51	-1.17	0.000
β_{12}	parallax	-1.80	0.42	-2.65	-1.01	0.000
β_{13}	spatialaudio	-3.81	0.82	-5.69	-2.46	0.000
β_{14}	highres	-4.87	1.23	-7.91	-3.08	0.000
β_{15}	altcolor	-0.10	0.31	-0.70	0.51	0.367

The factor *lowres* has the second highest effect size after *monoaudio*, meaning that it was the second most likely factor to be changed. Moreover, *highres* has the largest (negative) effect size, meaning that it was the least likely to be changed to *lowres*. This also accords with the subjective data shown in Figure 4, where resolution was considered the most important of the 5 factors.

4.8 Entropy of Eye Scanpaths

Considering the space in front of the participant, suppose that the number of states $s = 18$ (each angular segment is 10^o) and that the time period is divided into 30 segments (e.g., for a 20 minute session each segment would be 40 seconds). Then we would have for each of the 29 individuals 30 entropy readings. Figure 5A shows the relationship between the mean entropy over the 29 participants for each time period by time.

It can be seen that, as predicted, there is a strong negative association between entropy and time (r = -0.66 with 95% confidence interval -0.82 to -0.39). However, this does not show that entropy is connected with PI, since entropy could have declined just because participants got used to the environment and their scanpaths became more regular.

In order to examine whether there is a relationship with subjectively reported PI we can make use of the questionnaire variables for PI (Table 2) and consider the overall measure of PI, which here we refer to as PI_median, which is the median of the 4 variables. We use the median rather than the mean since these are ordinal variables. Now we want to see whether the PI_median scores are related to entropy in the period after which participants stopped making any further changes (i.e., after the last change that they made).

Figure 5B shows the entropy by PI_median. Contrary to expectations the relationship seems to be a curvilinear one, with low PI associated with low entropy, which then increases with PI and then curves downwards (a quadratic fit to these data is highly 'significant'). However, the curvilinear aspect is entirely caused by the 3 observations that have low presence (scores of 2 or less). What happens if we eliminate these? Then the relationship is still strong with negative correlation (r = -0.36, 95% confidence interval -0.65 to 0.03).

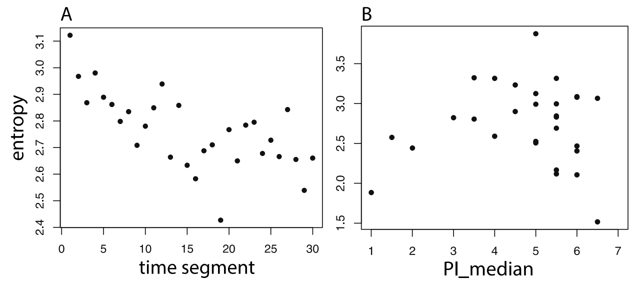


Figure 5 – Mean entropy by time and PI (A) Entropy by time segment over 30 segments (B) Mean entropy by PI_median.

A potential problem with this analysis is that we chose 18 states and 30 time periods. Perhaps the results are dependent on this choice. In order to overcome this possibility, we uniformly randomly chose between 5 and 40 time segments, and independently between 3 and 30 states, and ran the analysis for 1000 such combinations. We then computed the mean entropy over these repetitions. This results in the relationship between the entropy and subjective PI shown in Figure 6.

We see the same pattern as earlier. This is highly 'significant' as a quadratic fit. Moreover, if we eliminate the 3 with lowest presence (PI_median < 3), then there is a strong negative correlation between the entropy and the subjective PI: r = -0.41, with 95% confidence interval -0.69 to -0.02.

More formally, to analyze this we use the quadratic model:

$$\mu_i = \beta_0 + \beta_1 P_i + \beta_2 P_i^2 \tag{4}$$

$$entropy_i \sim Student\_t(\upsilon, \mu_i, \sigma)$$

where P_i is the median PI for the ith individual, corresponding to $entropy_i$. We use a Student t distribution for the likelihood (the distribution of the response conditional on the parameters) since for low degrees of freedom (υ) this distribution allows greater dispersion than the normal distribution, yet for high degrees of freedom $(\upsilon \geq 30)$ it closely approximates the normal. The median of the distribution is μ_i (and it is the mean for $\upsilon > 1$) and $\sigma > 0$ is the scale factor.

The prior distributions are $\beta_j \sim normal(0,10)$ as before and $\upsilon, \sigma \sim Gamma(shape = 2, rate = 0.1)$. The corresponding prior 95% credible intervals for these are 2.4 to 55.7. The Stan program was run with 3000 iterations and fully converged with results shown in Table 9. It is clear from the 95% credible intervals and the probabilities that the fit is very good. Also compare the very narrow posterior credible intervals with the prior intervals. As an example, the posterior probability $P(\beta_2 < 0) = 1 - 0.002 = 0.998$. Note that υ is lower than 30 (with probability 0.79) justifying the use of the Student t distribution rather than the normal.

Figure 6 shows the scatter diagram of the observed entropy plotted against the median presence, and the quadratic (4) with the parameters replaced by their means from the first column of Table 9, illustrating the close correspondence between observed and fitted data. For example, the quadratic of model (4) has maximum at $\widehat{PI} = -\beta_1/2\beta_2$. Using the means of the posterior distributions of the β parameters, $\widehat{PI} = 3.75$ which accords well with Figure 6.

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2025.3547241

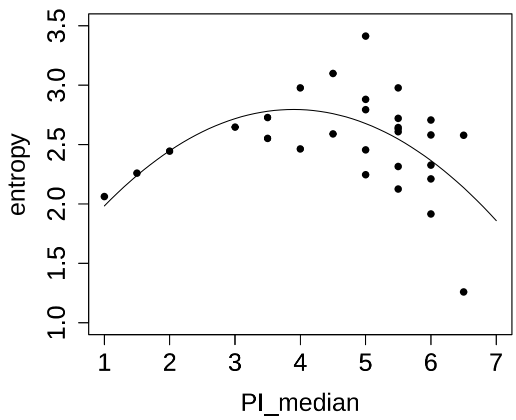


Figure 6 – Mean entropy over 1000 iterations with uniformly randomly generated time segments and states. The quadratic fitted curve is based on the means of the posterior distributions of the parameters of the quadratic model Eq (4).

Table 9 – Summary of the posterior distributions of the parameters of model (4) showing the means and standard deviations of the distribution and the 95% credible intervals. Prob > 0 are the probabilities that the parameters are positive.

Parameter	Mean	SD	2.5%	97.5%	Prob > 0
β_0	1.33	0.43	0.48	2.19	0.997
β_1	0.75	0.24	0.29	1.22	0.999
β_2	-0.10	0.03	-0.16	-0.04	0.002
σ	0.34	0.06	0.24	0.47	
υ	20.84	13.70	3.83	56.48	

Using the posterior distributions of the parameters we can generate entirely new pseudo random data on model (4) and compare with the actual values, referred to as the 'predictive posterior distributions'. For each individual i this results in a posterior distribution for predicted $entropy_i$. We can compare the means of these distributions with the original observed data. The correlation between the means of the predicted posterior distributions and the observed data is r = 0.59 (95% confidence interval 0.29 to 0.79).

This connection between the subjective data from the questionnaires and the entropy of the scanpath eye tracking data is fairly remarkable, since these two different scores in principle have nothing to do with one another but are nevertheless related, and in a way partially predicted by the theory. There is insufficient data to know whether the relationship actually has a quadratic shape or whether the three very low PI scores are just anomalies. Looking at the brief comments written by the participants, the one with the lowest median PI score (of 1) was not able to work the wrist buttons and therefore unable to change any of the factors: "I have not been able to change any aspect of the environment." The participant with the median PI score of 1.5 only commented: "When the resolution was low, the sensation of dizziness was greater. The two colors made me feel equally present." The one with the median score of 2 also mentioned a technical difficulty: "The buttons were too close, I've accidentally pressed the wrong option two times." No other participant mentioned such technical interface difficulties. The important point is that discounting these three data points there is the negative association as predicted by the theory.

Another way to consider this is that the underlying theory predicts that there must be some level P of PI such that for PI $\geq P$ there is a negative relationship between PI and entropy. We constructed an alternative statistical model where this P is estimated from the data. The model assigns weights such that the influence of the linear model $\beta_0 + \beta_1 PI_i$ on entropy is reduced for PI $< P$. Hence P acts as a discriminator between low and high values of PI, and adjusts the linear model accordingly. The details of this model are given in Supplementary Text S1 and Table S3 . The posterior distribution of P has 95% credible interval 1.42 to 6.66, with mean 4.48. The linear model between PI and entropy has negative slope with 95% credible interval -2.39 to -0.21 with mean 1.20. The posterior probability that the slope is negative is 0.997. The value $P = 4.48$ accords well with what can be seen in Figure 6. Just the fact that such a P can be estimated from these data is a further demonstration of the underlying theory – it supports the notion of a negative relationship between entropy and PI above some level of PI.

In all the above analysis we use confidence intervals for the correlation coefficient r as an indication of the strength of the relationships, rather than as formal statistical tests.

5 Discussion

5.1 Overview

The theory presented in [6] postulates that the illusion of 'being there' (Place Illusion, PI) in the environment depicted by the virtual reality displays is based on affordances for perception in VR via sensorimotor contingencies (SC) that match those of physical reality. It is a difficult theory to test, since there are many SC rules, and breaking some of them would be likely to result in simulator sickness. Here we presented a method whereby 4 different SC factors could be manipulated (switched) by participants, and the same for another factor that was not a sensorimotor contingency (color). We found that participants were more likely to choose to make transitions between configurations that enhanced SC factors than the change to color. This provides direct evidence for the theory, though of course, only with respect to these 4 factors. On the other hand, had participants made changes to color to the same extent that they made changes to the 4 SC factors then this would have been evidence against the theory.

5.2 Sensorimotor contingencies and Place Illusion

PI as a direct result of the perception through everyday SC makes sense, since if we perceive in VR using our bodies in the same way as in physical reality, and each act of perception results in predicted changes to perception based on prior experience, then the simplest hypothesis for the brain to make is 'this is where I am'. When we turn our head and eyes to look to one side, we expect to see what is to the side of us, and for sound to be modulated accordingly. If the wind was blowing in our right ear and we turn our head to the right we would be surprised if the wind still impacted our right ear rather than the front of our face (this would be an example of a breakdown in SC). When these same types of SC occur in VR this is strong evidence to the brain about where we are located (PI). This is not to say that PI cannot sometimes break [19, 20], for example, when the participant is moving around and accidentally collides with a real wall, or when the participant looks closely at an object and the pixels become very obvious. However, after a break, PI is likely to form again, since the SC remain the same [21].

5.3 Scanpath entropy

The scanpath theory of visual perception [11] argues that when individuals view environments similar to ones previously seen, they tend to follow similar scanpaths. The scanpaths are sequences of saccades (rapid eye movements) and fixations (periods when the eyes are relatively stationary and focused on a specific point). Such scanpaths have low entropy because they follow a regular pattern as the eyes move between a set of fixation points. The setup of our experiment was such that through making changes, participants were attempting to transition closer to a higher level of PI. The results show that from the neutral level (4) on the subjective scale (the maximum of the quadratic occurs at 3.75 in Figure 6) greater PI was associated on the average with lower eye scanpath entropy. Although the raw data suggests a possible quadratic relationship, this is likely to be the result of anomalies given that it is too much of a coincidence that the participants who reported technical difficulties are those where the PI score is low. Moreover, we have presented an alternative model which indicates that there is a discriminatory level of PI above which there is the clear negative correlations with entropy.

Jordan and Slater [13] argued on the basis of the earlier work in [11, 12] that this is because visual perception is largely top-down driven, i.e., that from minimal bottom-up cues in VR, participants rapidly infer the type of environment that they are in and then operate largely out of their own internal models based on prior perception of environments with similar features. Hence ‘being there’ in the VR is like being there in reality – minimal cues provide evidence about the type of environment, and then perception is driven by internal models.

5.4 Components of presence

In earlier work [6] ‘presence’ was decomposed into two orthogonal elements: PI the illusion of ‘being there’ (a perceptual illusion) and Plausibility (Psi), the illusion that the events that are happening there are really happening (a cognitive illusion), in both cases even though the participant knows for sure that these are not true. The entropy finding raises the possibility of an interesting additional dimension in the meaning of presence. There could also be an important non-conscious cognitive element that involves recognition of the type of place displayed and of having been in that type of place before. In the case of the environment of this experiment, most people would have been before in a large hall with pillars, and also would have previously witnessed a musical band playing live in front of them. Putting these two elements together also creates a recognizable environment. A prediction that follows from this that can be experimentally tested is that if the environment were completely bizarre, a type of structure that no one could have ever experienced in reality, although SC would still lead to the illusion of being there, recognition would not be possible, and therefore scanpath entropy would not decrease over time and would be unrelated to subjective assessment of PI. To some extent this was shown in one of the control conditions of the experiment of Jordan and Slater [13] which rendered the same environment (the participant on top of the pillar) but where all the triangles that formed the scene were randomly rotated. The recognition component would be most important when VR is being used for simulations of situations that could happen in reality, for example for training purposes. In such cases we do not only want the illusion of being there, but also the sense that ‘I’ve been in this type of place before’.

5.5 Uses of the 3M method

The 3M method was introduced in [9] in order to examine the impact of four factors on PI or Psi. In each transition they could change the illumination model or the field-of-view or the display type (first or third person perspective) or their virtual body representation. The method was also used to show that PI and Psi relied on different configurations. Azevedo, et al. [22] examined the impact of 4 factors (vision, hearing, haptics and olfaction) on PI and Psi in the context of virtual environments that depicted outdoor scenes. They also combined the method with the use of EEG monitoring. Bergström, et al. [14] used the method to examine the impact on Psi of the same string quartet used in the current experiment of the factors gaze (whether or not the musicians sometimes looked towards the participant), sound spatialization (Mono, Stereo, Spatial), auralization (no sound reflections, incorrect or correct sound reflections in relation to the virtual room), and environment (no outside sounds, or sound that corresponded to the setting). Skarbez, et al. [23], in the context of participant interaction with virtual human characters, also examined the impact on a variation of Psi referred to as ‘coherence’ of the behaviors of virtual human characters and other objects, as well as that of their own virtual body. Debarba, et al. [24] examined the impact on Psi in relation to a virtual human and of a self-avatar of various aspects of their animation - the face, hands, upper body and lower body of the character.

The method can be used for any qualitative response to a VR experience not just presence. For example, Murcia-López, et al. [25] required participants to change factors in the configuration to optimize the quality of their experience – where they observed a virtual human giving a presentation and could manipulate the eye gaze, blinking, mouth animation, and micro expressions of the virtual speaker. Gao, et al. [26] examined the impact on the level of believability of a virtual rock climbing environment of the visual appearance of the rocks, the overall visual scene, the environmental sound and dynamic behavioral factors. Fribourg, et al. [27] used the method to discover preferences for embodiment in a self-avatar of type of its appearance, control over it and perspective position. Gonçalves, et al. [28] required participants to optimize the feeling of being in a virtual room that was a replica of a corresponding real room that they had experienced, where they could switch amongst illumination rendering methods (global illumination, ambient occlusion, screen space reflections and direct shadows) finding that global illumination was the most effective. In an Extended Reality environment Lim and Ji [29] considered presence as the illusion of being co-located with surrounding objects and the impact of four factors on this (force feedback, occlusion, lighting, and material properties).

Combining 3M with Reinforcement Learning, leading to what we have called A3M, was first introduced by Llobera, et al. [10]. An advantage of A3M is that the process of participant choice selection amongst configurations is automated. They do not have to remember which possible changes might be made, since at each change-point the specific alternatives offered can be demonstrated and their impacts can be experienced before the choice is made. This is more important the greater the number of factors. However, the disadvantage is that a large number of trials might be needed for convergence, though this has not occurred in our applications to date, albeit with a small number of possible factor changes. Although 3M results in a Markov Chain which is a probabilistic model of participant choices, the A3M method produces a further probabilistic model which is the *policy* associated with the RL. This is the set of probabilities associated with which transitions will be offered in the context of each configuration that maximize the long-term reward to the RL. Although we have not made use of this in the current paper, it may be useful in other contexts.

The 3M method has some similarities with conjoint analysis that has been used for half a century in market research [30, 31]. It addresses the same problem as 3M – assessing how a number of factors influence an outcome such as consumer preference for a product. It is a survey method whereby respondents are presented with a set of alternatives, where each alternative is a combination (configuration) of various attributes (factors) with varying levels. Survey respondents rate these configurations, and conjoint analysis then decodes their preferences, uncovering the relative importance of each attribute. Agarwal, et al. [32] refer to it as “one of the most celebrated research tools in marketing and consumer research” and it has been applied thousands of times in different applications. Conjoint analysis has also been combined with eye tracking [33] where in addition to subjective consumer choices the amount of time that a consumer looks towards a particular item can also be exploited. Conjoint analysis has even been used in VR for analysis of design alternatives [34]. However, its more recent uses have been also for preference analysis where VR is used as a presentation mode rather than to understand issues relating to VR in itself – for example, consumers preferences for product packaging [35]. A major difference between conjoint analysis and 3M is that 3M involves real-time manipulation of the factors that then lead to

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2025.3547241

transitions across the configurations, and in this sense, it is bottom-up, whereas conjoint analysis is top-down since it presents consumers with different configurations and then attempts to separate out the relative importance of the constituent factors. However, in future uses of 3M the statistical methods that have been developed over the past 50 years for conjoint analysis should be taken into account.

5.6 Constituents of presence

Presence in VR has been studied since the 1990s originating from the notion of telepresence with respect to teleoperator systems where people had the sense of being at the location of the remote robot that they were manipulating, and was originally discussed in a famous paper by Minsky [36]. Ideas from telepresence were introduced into the VR field, early influential papers being [1, 37-40]. This led to a huge amount of research covering the definition of presence, its measurement and to some extent theories about how it occurs and the idea that presence leads to realistic behavior [41]. Explanations have mainly focused on properties of the display and interactive systems, essentially a list of factors that need to be addressed, for example: field of view [42], latency [8], visual display [43-45], framerate [46], and many others. See [47] for a comprehensive and systematic review. Instead of a list our theory provides an organization by considering the extent to which each factor contributes towards the VR participant experiencing PI. It is important to note that by PI we strictly mean the illusion of 'being there' even though the participant knows for sure that they are not 'there' (and this feeling is itself part of the quale, the subjective sensation of PI in this case). PI does not include issues such as attention, interest, involvement, enjoyment, or any the other possible responses to experiencing and interacting in a virtual environment [3, 40]. For example, an individual can be in a real environment, yet be bored, not pay attention to the events around them, but in terms of SC everything that they perceive is in the environment around them, and unless they lose (or enter an altered state of) consciousness through drugs or hypnosis, they will not have any doubt as to the environment that they are in. The theory postulates that to the extent that this is the same in VR, so the greater the likelihood that presence will be engendered.

Examination of eye scanpath entropy is not only of interest from the point of view of measurement, but its ramifications have also led to an interesting new component of presence alongside PI and Psi, which is that of Recognition: to what extent does the environment lead to (probably nonconscious) recall of other similar environments that participants have experienced? This is orthogonal to PI in the sense that based only on SC participants may experience PI without Recognition, and they may have the sense of Recognition in an environment where there is no PI. For example, viewing an environment without head-tracking may still spark Recognition, but PI will fail as soon as the participants move their head. As mentioned, we therefore put forward the hypothesis that immersion in an extremely unusual, otherworldly environment, with no familiar points of reference (for example, no floor or walls) is likely to lead to PI if SC are sufficiently supported, but are unlikely to lead to entropy changes or negative correlation of entropy with PI.

5.7 Limitation

The sample size is relatively small and it could be argued that this limits the possible generalizability of the findings. However, the Bayesian method we have used for statistical analysis of the RL transitions puts this in a different light. As can be seen we started with very wide prior distributions on the parameters which have then become quite narrow after the introduction of the data. This can be seen by looking at the 95% credible intervals for the parameters. For example, in Table 9, consider the coefficient of 'monovision' – the prior 95% credible interval was -20 to 20 and the posterior credible interval is 1.27 to 3.39. It is the same with the other parameters. The Bayesian method does allow for smaller sample sizes, and one can see at a glance the impact that the data has had on the prior distributions in the transformation to posteriors. While it is possible that with a larger sample size the results may change, it is very unlikely that there will be such shifts in the posterior distributions that the findings into question. However, of course there need to be further replications of this experiment as is normal in science.

6 Conclusions

There are three main conclusions to the work described in this paper. First, we have provided empirical evidence in support of the theory that the PI component of presence is influenced by the extent to which the VR system supports sensorimotor contingencies that match those of perception in physical reality. Second, we have provided a replication of the result in [13] that eye scanpath entropy is inversely correlated with PI. However, more evidence is needed to explore the possible finding that the relationship between scanpath entropy and PI is curvilinear, with low levels of PI also associated with low entropy. Our supplementary analysis indicates that the quadratic relationship may be spurious, but more work is needed on this. Finally, we have introduced a potentially new component of presence. Alongside the illusions of 'being there' (PI) and that events are really occurring (Psi), there is the sense of Recognition ('I have been in this type of place before'). We predict that if PI is strong but Recognition is absent then there will not a relationship between PI and scanpath entropy. We are currently designing a study to explore this. Our ultimate prediction is that when PI, Psi and Recognition are all strong people will behave realistically in VR: they are there, these events are happening, they've been in this type of place before, therefore they respond accordingly. It would be interesting to study how behavior varies with independent variation of these three presence components.

Acknowledgments

We thank the company Pico for providing the Pico Neo 3 head-mounted displays with the eye tracking capability. This research was initially supported by the European Research Council Advanced Grant MoTIVE (#742989), and ES is supported by the GuestXR project which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101017884. ES is also partially supported by a PhD studentship from Facebook. This work (MS and JL) is also supported by the European project Presence XR, a Horizon Europe Innovation Project co-financed by the EC under Grant Agreement ID: 101135025.

Data Accessibility

All data and the statistical analysis programs are available on the Kaggle system at these addresses:

For the parsing of the raw data, overview of the reinforcement learning scores and questionnaire responses, as well as building the transition table:

https://www.kaggle.com/code/joanllobera/suplementary-material-for-k-kt-t-nc-et-al

For the analysis of transitions:

For the analysis of transitions:

https://www.kaggle.com/code/melslater/transitions-analysis

For the analysis of the scanpath data:

https://www.kaggle.com/code/melslater/scanpath

The programs can be executed by entering edit mode.

Supplementary Material

Supplementary Video S1: An illustration of the scenario from the viewpoint of a participant.

References

[1] T. B. Sheridan, "Musings on Telepresence and Virtual Presence," *Presence: Teleoperators and Virtual Environments,* vol. 1, pp. 120-126, 1992.

[2] J. Lessiter, J. Freeman, E. Keogh, and J. Davidoff, "A cross-media presence questionnaire: The ITC-Sense of Presence Inventory," *Presence: Teleoperators and Virtual Environments,* vol. 10, pp. 282-298, 2001.

[3] B. G. Witmer and M. J. Singer, "Measuring presence in virtual environments: a presence questionnaire," *Presence: Teleoperators and Virtual Environments,* vol. 7, pp. 225-240, 1998.

[4] S. Grassini and K. Laumann, "Questionnaire measures and physiological correlates of presence: A systematic review," *Frontiers in psychology,* vol. 11, p. 349, 2020, doi: 10.3389/fpsyg.2020.00349.

[5] M. Meehan, B. Insko, M. Whitton, and F. P. Brooks Jr, "Physiological measures of presence in stressful virtual environments," *Acm transactions on graphics (tog),* vol. 21, no. 3, pp. 645-652, 2002, doi: 10.1145/566654.566630.

[6] M. Slater, "Place Illusion and Plausibility can lead to realistic behaviour in immersive virtual environments," *Philos Trans R Soc Lond,* vol. 364, pp. 3549-3557, 2009, doi: 10.3389/frobt.2016.00074.

[7] J. K. O'Regan and A. Noë, "A sensorimotor account of vision and visual consciousness," *Behav Brain Sci,* vol. 24, pp. 939-1031, 2001.

[8] M. Meehan, S. Razzaque, M. C. Whitton, and F. P. Brooks Jr, "Effect of latency on presence in stressful virtual environments," *Virtual Reality, 2003. Proceedings. IEEE,* pp. 141-148, 2003.

[9] 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,* vol. 29, p. Paper: 92, 2010, doi: 10.1145/1833349.1778829.

[10] J. Llobera, A. Beacco, R. Oliva, G. Şenel, D. Banakou, and M. Slater, "Evaluating participant responses to a virtual reality experience using reinforcement learning," *R. Soc. Open Sci.,* vol. 8, p. 210537., 2021, doi: https://doi.org/10.1098/rsos.210537.

[11] L. Stark and S. R. Ellis, "Scanpaths revisited: Cognitive models, direct active looking," in *Eye movements: Cognition and visual perception*, D. F. Fisher, R. A. Monty, and J. W. Senders Eds.: Routledge Taylor & Francis Group, 1981, pp. 193-226.

[12] L. W. Stark, "How Virtual Reality Works! The Illusions of Vision in Real and Virtual Environments.," *Proc SPIE: Symposium on Electronic Imaging: Science and Technology,* pp. 5-10, 1995.

[13] J. Jordan and M. Slater, "An Analysis of Eye Scan Path Entropy in a Progressively Forming Virtual Environment," *Presence: Teleoper. Virtual Environ.*, vol. 18, pp. 185-199, 2009.

[14] 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,* vol. 23, no. 4, pp. 1352-1359, 2017, doi: 10.1109/TVCG.2017.2657138.

[15] D. H. Foster, "Color constancy," *Vision research,* vol. 51, no. 7, pp. 674-700, 2011.

[16] R. S. Sutton and A. G. Barto, "Reinforcement Learning: An Introduction," 1998.

[17] B. Carpenter *et al.*, "Stan: A probabilistic programming language," *Journal of statistical software,* vol. 76, no. 1, 2017.

[18] N. P. Lemoine, "Moving beyond noninformative priors: why and how to choose weakly informative priors in Bayesian analyses," *Oikos,* vol. 128, no. 7, pp. 912-928, 2019.

[19] M. Slater and A. Steed, "A Virtual Presence Counter," *Presence: Teleoperators and Virtual Environments,* vol. 9, no. 5, pp. 413-434, 2000.

[20] B. Liebold, M. Brill, D. Pietschmann, F. Schwab, and P. Ohler, "Continuous measurement of breaks in presence: Psychophysiology and orienting responses," *Media Psychology,* vol. 20, no. 3, pp. 477-501, 2017.

[21] M. Garau, D. Friedman, H. R. Widenfeld, A. Antley, A. Brogni, and M. Slater, "Temporal and Spatial Variations in Presence: Qualitative Analysis of Interviews from an Experiment on Breaks in Presence," *Presence: Teleoperators and Virtual Environments,* vol. 17, pp. 293-309, 2008.

[22] 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,* vol. 23, no. 4, pp. 354-368, 2014.

[23] R. Skarbez, S. Neyret, F. P. Brooks, M. Slater, and M. C. Whitton, "A psychophysical experiment regarding components of the plausibility illusion," *IEEE transactions on visualization and computer graphics,* vol. 23, no. 4, pp. 1369-1378, 2017.

[24] H. G. Debarba, S. Chagué, and C. Charbonnier, "On the plausibility of virtual body animation features in virtual reality," *IEEE Transactions on Visualization and Computer Graphics,* vol. 28, no. 4, pp. 1880-1893, 2020, doi: 10.1109/TVCG.2020.3025175.

[25] M. Murcia-López, T. Collingwoode-Williams, W. Steptoe, R. Schwartz, T. J. Loving, and M. Slater, "Evaluating Virtual Reality Experiences Through Participant Choices," in *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2020: IEEE, pp. 747-755.

[26] B. Gao, J.-I. Kim, and H. Kim, "Sensory and Perceptual Consistency for Believable Response in Action Feedback Loop," in *Proceedings of Computer Graphics International 2018,* 2018: ACM, pp. 201-210.

[27] R. Fribourg, F. Argelaguet, A. Lécuyer, and L. Hoyet, "Avatar and Sense of Embodiment: Studying the Relative Preference Between Appearance, Control and Point of View," *IEEE Transactions on Visualization and Computer Graphics,* vol. 26, no. 5, pp. 2062-2072, 2020.

[28] G. Gonçalves, M. Melo, P. Monteiro, H. Coelho, and M. Bessa, "The role of different light settings on the perception of realism in virtual replicas in immersive Virtual Reality," *Computers & Graphics*, vol. 117, pp. 172-182, 2023, doi: doi.org/10.1016/j.cag.2023.10.021.

[29] C. Lim and Y. G. Ji, "The Effects of Physical Coherence Factors on Presence in Extended Reality (XR)," *International Journal of Human-Computer Studies,* vol. 172, p. 102994, 2023, doi: 10.1016/j.ijhcs.2022.102994.

[30] P. E. Green and V. Srinivasan, "Conjoint analysis in consumer research: issues and outlook," *Journal of consumer research,* vol. 5, no. 2, pp. 103-123, 1978, doi: 10.1086/208721.

[31] P. E. Green and V. R. Rao, "Conjoint measurement-for quantifying judgmental data," *Journal of Marketing research,* vol. 8, no. 3, pp. 355-363, 1971, doi: 10.1177/002224377100800031.

[32] J. Agarwal, W. S. DeSarbo, N. K. Malhotra, and V. R. Rao, "An interdisciplinary review of research in conjoint analysis: Recent developments and directions for future research," *Customer Needs and Solutions*, vol. 2, pp. 19-40, 2015, doi: 10.1007/s40547-014-0029-5.

[33] M. Meißner and R. Decker, "Eye-tracking information processing in choice-based conjoint analysis," *International Journal of Market Research,* vol. 52, no. 5, pp. 593-612, 2010, doi: 10.2501/S147078531020151X.

[34] J. Dijkstra and H. J. Timmermans, "Conjoint analysis and virtual reality: A review," in *4th Design & Decision Support Systems (DDSS 1998) in Architecture and Urban Planning Conference, July 26-29, 1998, Maastricht, The Netherlands*, 1998: Eindhoven University of Technology.

[35] G. Branca, R. Resciniti, and S. M. C. Loureiro, "Virtual is so real! Consumers' evaluation of product packaging in virtual reality," *Psychology & Marketing,* vol. 40, no. 3, pp. 596-609, 2023, doi: 10.1002/mar.21743.

[36] M. Minsky, "Telepresence," *Omni,* pp. 45-52, 1980.

[37] T. B. Sheridan, "Further musings on the psychophysics of presence," *Presence: Teleoperators and Virtual Environments,* vol. 5, pp. 241-246, 1996.

[38] R. M. Held and N. I. Durlach, "Telepresence," *Presence: Teleoperators and Virtual Environments,* vol. 1, pp. 109-112, 1992.

[39] D. Zeltzer, "Autonomy, interaction, and presence," *Presence: Teleoperators and Virtual Environments,* vol. 1, no. 1, pp. 127-132, 1992.

[40] T. D. Matthew Lombard, "At the Heart of It All: The Concept of Presence," *Journal of Computer-Mediated Communication,* vol. 3, 1997.

[41] M. V. Sanchez-Vives and M. Slater, "From Presence to Consciousness Through Virtual Reality," *Nature Reviews Neuroscience,* vol. 6, pp. 332-339, 2005.

[42] J.-W. Lin, H. B.-L. Duh, D. E. Parker, H. Abi-Rached, and T. A. Furness, "Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment," *Proceedings ieee virtual reality 2002,* pp. 164-171, 2002, doi: 10.1109/VR.2002.996519.

[43] M. Slater, P. Khanna, J. Mortensen, and I. Yu, "Visual realism enhances realistic response in an immersive virtual environment.," *IEEE computer graphics and applications,* vol. 29, pp. 76-84, 2009, doi: 10.1109/MCG.2009.55.

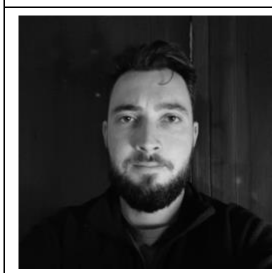
[44] I. Yu, J. Mortensen, P. Khanna, B. Spanlang, and M. Slater, "Visual realism enhances realistic response in an immersive virtual environment - Part 2," *IEEE Computer Graphics and Applications,* vol. 32, pp. 36-45, 2012.

[45] C. Hendrix and W. Barfield, "Presence within Virtual Environments as a Function of Visual Display Parameters," *Presence-Teleoperators and Virtual Environments,* vol. 5, pp. 274-289, 1996.

[46] W. Barfield and C. Hendrix, "The Effect of Update Rate on the Sense of Presence within Virtual Environments," *Virtual Reality: The Journal of the Virtual Reality Society,* vol. 1, pp. 3-16, 1995.

[47] C. S. Oh, J. N. Bailenson, and G. F. Welch, "A systematic review of social presence: Definition, antecedents, and implications," *Frontiers in Robotics and AI,* p. 114, 2018, doi: 10.3389/frobt.2018.00114.

Esen Küçüktütüncü obtained her Masters degree in Cognitive Systems and Interactive Media at Universitat Pompeu Fabra. She is currently pursuing a PhD degree with the Event Lab, Institute of Neurosciences, Universitat de Barcelona, Spain. Her research focuses on emergent social dynamics within social virtual reality and development of virtual agents to ensure social harmony in immersive shared environments.



Francisco Macia-Varela is an XR Developer and Researcher based in Barcelona with a MSc in Serious Games and Virtual Reality. He has worked on several industry and research projects for VR focusing on the areas of healthcare, experimental psychology, and digital art installations. He is currently a Senior Unity Engineer working on an experimental projective AR device.

Joan Llobera obtained his PhD at the Event Lab, Universitat de Barcelona, Spain and did a postdoc at the Cognitive Neuroscience Lab at Ecole Polytechnique Federale de Lausanne. His research combines software engineering and behavioral experiments to study different aspects of virtual reality. He currently works as a senior researcher at the Artanim Foundation working on bringing Physics-based Character Animation techniques to Virtual Reality.

Mel Slater is a Distinguished Investigator at the University of Barcelona in the Institute of Neurosciences, and co-Director of the Event Lab (Experimental Virtual Environments for Neuroscience and Technology). He was previously Professor of Virtual Environments at University College London in the Department of Computer Science. He has been involved in research in virtual reality since the early 1990s, and has been first supervisor of 40 PhDs in graphics and virtual reality since 1989. He is a Research Award Winner of the Alexander von Humboldt Foundation in 2021, and was elected to the IEEE VGTC Virtual Reality Academy in 2022.