



Validation Study of a Virtual Laboratory for Research in Environmental Psychology  
Marie-France Larsson  
Master's Degree Thesis  
2010 06 04  
Supervisor Richard Ahlström

*Abstract*

The aim of this study was to validate the use of a virtual environment (VE) for psychological research. A study made by Maslow and Mintz (1956) in which they found that a room's aesthetic affects how participants estimate pictures of people was replicated using a VE. A picture of a beautiful room and a picture of an ugly room were projected on an angled screen measuring 185 cm in height and 405 cm in width as to attempt to replicate the effect found in the original study. A total of 31 participants were tested in either the beautiful or the ugly environment. A t-test revealed no significant difference between the groups on the total score given on the pictures. However, a mixed-model ANOVA revealed several significant differences and interaction effects. The results of this study suggest that VE's could be valid for psychological research, in particular environmental psychological experiments, but that more investigations are needed in order to understand this instrument better.

Keywords: environment, aesthetics, laboratory testing, Maslow, Mintz, virtual environment, virtual reality

Validation Study of a Virtual Laboratory for Research in Environmental Psychology

Virtual reality (VR) and virtual environments (VE) provide almost unlimited sources of diverse environments that can be used in different situations. Of great interest in the field of psychology is the use of VE's for environmental psychology research to substitute for a real environment of interest. VR is understood as a "three-dimensional data set describing an environment based real-world or abstract objects and data" (Blade & Padgett, 2002). VR and VE are often used indistinctively, but a VE is more particularly an "artificial environment that the user interacts with" (Blade & Padgett, 2002). In this study, interaction is comprehended as a synonym of "being in contact with". The advantages the use of VR has for research in environmental psychology are many. It can for instance reduce the costs transportation of participants to a real environment would generate, or even make possible some designs that would otherwise be too costly or complicated if performed in a given real environment. Furthermore, the almost unlimited possibilities different environments allows for a great degree of variety, although all in one physical location.

However, an important issue with artificial settings is obviously that their use challenges external validity (Bell, Green, Fisher & Baum, 2001). Can a VE be used as if it was as good as an actual environment in psychological research? Are the results found in a VE true for the real milieu as well? These questions are not new, but the appearance of more and more sophisticated technology is. VR and VE are quite broad concepts and what they can concretely be may vary from a single projection of a still image where participants are mere spectators, to a more complex technology where individuals can be active and control the VE. Therefore, a main concern is to know if a given VE using a particular technology is at least good enough for a specific goal.

It seems as, even if VR has been used for a long time in psychology (Bell et al, 2001), not so much has been done to test its validity. In the field of psychology the usage of new VR technologies has been mostly documented in clinical psychology (Botella, Perpiñá, Baños & García-Palacios, 1998<sup>b</sup>), more particularly for treatment procedures (e.g. Botella, Baños, Perpiñá, Villa, Alcañiz & Rey, 1998<sup>a</sup>; North & North, 1996; Rothbaum, Hodges, Kooper, Opdyke, Williford & North, 1995). Stanney and Zyda (2002) assert that, when it comes to orientation tasks, the use of VR has been studied rather thoroughly and one can be fairly assertive that knowledge acquired in a VE can be transferred to the real environment (e.g. Darken & Banker, 1998; Ruddle & Lessels, 2009). As for environmental psychology, de Kort, Meijnders, Sponselee & Ijsselsteijn (2002) for instance conducted an experiment to observe if screen size influenced the quality of a virtual environment, and came to mixed conclusions. Others (Kjellgren & Buhrkall, 2010) compared results obtained in a natural environment with results obtained in a virtual natural

environment and discovered that the VE had similar but weaker effects compared to the natural environment.

However, none of the studies mentioned above concentrate exclusively on the validation of a given VR instrument for research in environmental psychology. Previous studies either concentrate on another field of psychology such as clinical psychology and treatments, or focus on certain aspects of the simulation, for example its efficiency in terms of image size. In order to maintain a scientific rigorousness, it is essential that the instruments used function as they are expected to. So as to gather this information, instruments themselves have to be tested. The distinctiveness of this paper was hence that it tested the validity of a given VR apparatus in order to estimate its validity when substituting a real environment in an environmental psychological experiment. The current study strived to determine to which extent the equipment described below (see under method) can replace a real environment in a research context.

The aim of this study was to validate the use of a simple VR apparatus to substitute for a real environment when the effect of the environment is the studied outcome. The intention was to explore if a VE deprived of sounds, movements, olfactory stimulus, and control from the participants could have similar effects to a real environment in laboratory testing. This was done by attempting to replicate the results of a study made in a real setting where a significant effect of the milieu was found by replacing the real environment with a VE. The chosen study, “Effects of Esthetic Surroundings: I. Initial Short-Term Effects of Three Esthetic Conditions upon Perceiving “Energy” and “Well-Being” in Faces”, was first conducted by Maslow and Mintz in 1956. During this original experiment, participants rated pictures of faces on an energy scale and a well-being scale in different environments varying in aesthetics. The authors found that faces were rated significantly more positive when participants were in a beautiful setting (Maslow & Mintz, 1956). Those results obtained by Maslow and Mintz could be explained to some extent by the affective priming effect. It is claimed that individuals who are exposed to a certain element that causes an affective reaction can later be influenced by this response in how they perceive and evaluate subsequent events or objects (Fazio, 2001). Therefore, according to this theory, individuals could be influenced by an appealing or unappealing room when perceiving another object, in Maslow and Mintz’s case, pictures of human faces. Consequently, it could be believed that whether the room that is shown is actual or virtual, the effect it has on the participant, given a virtual room can have such an effect, will influence successive appraisals.

Based on the original study and theoretical backgrounds, it was hypothesized that the VE could recreate the effect found in the real environment, but based on results from previous research concerning VE, to a lower degree.

### *Method*

#### *Design*

The validity of a given VR apparatus (see under *Virtual laboratory* for a detailed description of the equipment used) was studied by replicating an experiment originally made in a real environment to study effects of various environments on participants by substituting those environments with VE's. Maslow and Mintz's (1956) study, "Effects of Esthetic Surroundings: I. Initial Short-Term Effects of Three Esthetic Conditions upon Perceiving "Energy" and "Well-Being" in Faces", was chosen for this experiment. The original study compared how participants rated negative prints of faces on an energy scale and a well-being scale in different environments varying in aesthetics. Three rooms were used initially; one beautiful, one average, and one ugly. The authors then used 2 Likert-scales, one measuring well-being, and the other measuring energy, to determine the effects of the room on the perception of 24 negative prints of faces (of which 4 were dummies), using 12 pictures for each scale. In the original study, participants were first debriefed by the authors on face stereotypy during approximately 10 minutes in order to decrease their anxiety level and consequently diminish the risk of task orientation. The original participants were thereafter sent to a naïve examiner that left each participant alone in the room for 5 minutes and then returned with the pictures as to evaluate them by talking with the participants rather than letting them fill in the scale themselves. In this manner, participants could use their own words to qualify the faces and the scales were consequently less rigid. Maslow and Mintz randomized the examiners with regards to the rooms and gender. There were 8 females and 8 males in the beautiful room, 8 males and 8 females in the ugly room, as well as 10 males in the average room. The participants in the average room were evaluated by one of the authors rather than by a naïve examiner. The results of Maslow and Mintz's study showed that participants rated the pictures significantly more positive in the beautiful room compared to the average and the ugly room, but that there was no significant difference between the average room and the ugly room in how pictures were rated.

This particular experiment was chosen for the following reasons. First of all, it was originally conducted in real environments, which is the first necessary condition to compare the results to those of a VE. Also, the effect studied in the original study was that of the environments, which was what the present study meant to explore. Finally, it also observed the short-term effects of an environment's aesthetics (the original participants were exposed to the environment approximately 12 minutes), which is compatible to the use of VR since, according to studies, the maximal necessary period of time required to attain as much presence as possible in a given VE is 15 minutes (Sadowski & Stanney, 2002).

### *Participants*

Participants were recruited from the social work, human resources and social sciences master's programs at Mid Sweden University in Östersund. They were informed about the experiment and invited to sign up during a lecture. A total of 135 students were invited, 43 signed up, and 31 completed the study. The participants' mean age was 29 (SD = 7). The sample consisted of 71% females and 29% males. Participants were thanked for their participation by being eligible to win a gift certificate (value of 100 SEK) to the local movie theater.

### *Apparatus*

*Virtual laboratory.* The experiment was conducted in the virtual laboratory at the psychology department at Mid Sweden University in Östersund (see Figure 1). The room in which the participants sat was 4 meters deep by 4 meters wide and 3 meters high. The walls were covered with matt black paint, and the ceiling and floor were in a light beige color. The room was equipped with 3 glass screens measuring



**Figure 1.** The VR laboratory showing part of the screen as well as the desk and chair where participants sat during the experiment.

135 cm wide by 185 cm high placed together to create a screen with a total area of 7.5 m<sup>2</sup>. Two of the glass screens were placed on each side of the middle screen at a 120 degrees angle to create a sense of periphery and covered the participants' vision field. Images were projected on mirrors and back to the screens with three Projectiondesign projectors equipped with a short exposure lens, all placed behind the screens and not visible to the participants. The image was sent to the projectors via a desktop computer located in another room. During the experiment, participants sat on a padded chair by a desk placed at 120 cm from the central screen in order to create a visual field of 140 degrees for the participants.

*People pictures.* The 24 pictures of the people shown on the prints were taken from a picture bank at Mid Sweden University. The printed version of each picture measured 160 mm in width and 215 mm in height. The pictures were printed on a white A4 (210 mm by 297 mm) piece of paper each, and the scale on which participants would give their impressions was placed on the same piece of paper under each picture. Since Maslow and Mintz are not clear in their study concerning which facial expression people on the pictures had, 8 pictures of happy people, 8 pictures of neutral people and 8 pictures of sad people were chosen. In order to keep the balance in all 3 expressions, dummies used by Maslow and Mintz were discarded. Twelve pictures were showing males, and the other 12 pictures showed females. People on the pictures were all wearing a light grey t-shirt and were taken from the same distance so that the face filled a similar amount of space in all pictures. Black and white prints of the pictures were used in order to minimize eventual color effects. It was chosen not to use negative prints of the pictures to neutralize them since the pictures already were similar and neutral.

In accordance with Maslow and Mintz procedure, 2 different Likert-scales with 6 answer alternatives were used to grade the pictures. One scale was measuring energy (from “very tired” to “very lively”) while the other was measuring well-being (“very unhappy” to “very happy”). In order to avoid any order effect, the pictures were randomized. So as to reduce the risk of perceptual confounds, the pictures were also randomized to the scales. Scales were calculated by adding the scores of the 24 pictures to create a total score. Subcategories were also created by adding the scores of pictures of happy females, of happy males, of neutral females, of neutral males, of sad females and of sad males. Six subcategories were thus created, using the same pictures.

*Room pictures.* For the present study, it was decided to use two VR rooms, a beautiful room and an ugly room, since the largest effect found in the original study was between those, but also because the design used when testing the average room was questionable and divergent from the initial design used for the beautiful and ugly rooms. The group used in the average room in the original study was highly different in gender from the other groups, and the procedure differed to a great extent, being much less controlled than for the other rooms (see above under *Design*, for more details, or refer to the original report).

The room pictures measured 3150 pixels by 1400 pixels so as to cover the surface of the screen. They were selected as to resemble the rooms used by Maslow and Mintz, and emphasis was put on choosing rooms which would differ greatly from each other in their aesthetics. The beautiful room for this study was similar to Maslow and Mintz’s room in the sense that it also was tastefully decorated, had beige walls and drapes around 2 openings (one a window, the other one a French door, instead of 2 windows used in the original experiment) through which light

entered the room. Comfortable furniture was placed in the room (soft chairs and a soft divan), enjoyable artworks were hanged on the walls and placed in the room, which created a pleasant and relaxed atmosphere. The main difference was that Maslow and Mintz's beautiful room was a study, while the actual research's beautiful room was a living room. As for the ugly room, that of the original study was a small repulsive and dirty room with grey walls, untidy furniture, a filthy bed and accessories to recreate a janitor's room. For the actual study, the room was also ugly and dirty, with torn newspaper on the floor, graffiti on the walls, empty and crushed cans of beer on the floor, broken benches that were placed in a messy manner, little light entering the room and liquid on the floor that gave a feeling of filthiness. The ugly room for this study was a greatly neglected waiting room.

*Questionnaire.* A questionnaire with background questions was used in order to gather information about the participants' age and gender. There also was a question using a 6 items Likert-scale (from "do not agree at all" to "completely agree") in order to control to what extent the participants felt they were "there" in the room shown on the screen. So as to control the aesthetics of the rooms, a question using a 6 items Likert-scale (from "very unaesthetic" to "very aesthetic") was used. Finally, the State-Trait Anxiety Inventory for anxiety as a state (STAI-S) questionnaire was used in order to control differences in anxiety between the groups. Since it was decided not to spend as much time with the participants to avoid influencing them, there was a higher risk of participants being more anxious. This test allowed controlling the actual anxiety level of participants, and verifying that groups did not differ on this aspect.

#### *Procedure*

At their arrival, participants were informed that their participation was voluntary and confidential. They were told that they could withdraw from the experiment at any moment if they felt the need for it, and that the data was gathered for research purposes only. Participants were instructed to rate 24 pictures of human faces on two different scales after a resting period of 5 minutes where they sat in the VR lab on their own (exactly as was done in the original study). Half the pictures were to be rated on an energy scale and the other half was to be rated on a well-being scale. Participants were told to be spontaneous and asked not to analyze the faces since their first impression was what was of interest. They were told to use synonyms of the words on the scales if preferred as to render the scales more flexible (to resemble what was done during the original study). When those pictures were rated, participants were to fill in the background questionnaire mentioned above. After participants finished rating the environment and the pictures, they were thanked and asked not to share information about the experiment to other participants as to avoid biased results. They were also informed that a follow-up of the experiment with further information on its aim would be sent later via e-mail. The exposure to the VE

lasted for approximately 15 minutes (it varied from individual to individual depending on how much time they needed to accomplish the task), which according to studies, is the maximal necessary period of time required to attain as much presence as possible in a given VE (Sadowski & Stanney, 2002). The chosen environment was randomized as to be equally distributed through gender, time of the day and scales. All communication, spoken and written, with participants was carried out in Swedish.

#### *Analysis*

PASW Statistics for Windows (18.0.0) was used to analyze the collected data. The alpha level for statistical tests for this paper's analysis was 0.05. To assure the groups did not differ in terms of anxiety, the STAI-S questionnaire results were analyzed with a t-test using the group the participants belonged to as the independent variable and the scores on the STAI-S questionnaire as the dependent variable. To ensure the groups did not differ in term of how realistic they thought the virtual environment was, a t-test using the group participants belonged to as the independent variable and the score from the background question asking how much the participant felt present in the environment as the dependent variable. In order to make sure the 2 environments were perceived significantly differently in terms of aesthetics, a t-test was performed using the group the participants belonged to as the independent variable and the score on the background question asking how aesthetic the participant thought the room was as the dependent variable. As to replicate Maslow and Mintz's study, a t-test was performed on the total score given by each participant as a dependent variable, and as was originally executed in the initial study, the group participants belonged to was used as the independent variable. In order to explore the data and uncover more subtle differences, a mixed-model analysis of variance (mixed-model ANOVA) was performed using the group participants belonged to along with their gender as independent variables. Since it is unclear what emotions the prints in the original study showed, the 3 different moods were analyzed separately. Given the evidence that males and females are perceived in significant different manners when it comes to facial expressions (e.g. Becker, Kenrick, Neuberg, Blackwell & Smith, 2007; Martin, 1987), the portrayed individuals' gender was also used as a variable. The different moods and the portrayed individuals' gender were therefore used as within-subject factors. Since there is no known theoretical background as to why the well-being scale and energy scale Maslow and Mintz used could be compiled (there is nothing sustaining that they measure different sides of a same concept), those scales were analyzed individually.

### Results

The t-test performed on the STAI-S results showed that the groups did not vary significantly on this variable,  $t(29) = -.97, p > .05$ . The participants in the ugly room had an average anxiety score of 1.77 ( $SE = .11$ ), while those in the beautiful room had an average anxiety score of 1.64 ( $SE = .08$ ).

The perceived degree of reality did not differ significantly between the groups according to the t-test,  $t(29) = .72, p > .05$ . The mean score for the ugly room in terms of reality was 3.27 ( $SE = .36$ ), and 3.69 for the beautiful room ( $SE = .45$ ).

The results of the t-test confirmed the rooms were significantly perceived differently in terms of aesthetics  $t(17.96) = 2.87, p = .01$ . The average score for the participants in the ugly room was 3.67 ( $SE = .48$ ), as for those in the beautiful room, the average score was 5.13 ( $SE = .18$ ).

As for the t-test replicating Maslow and Mintz's procedure, it did not find a significant difference between the groups in how they rated the pictures,  $t(29) = .99, p > .05$ , although the participants in the ugly room ( $M = 80.67, SE = 1.82$ ) rated the pictures more negatively than the participants in the beautiful room ( $M = 83.63, SE = 2.35$ ).

The mixed-model ANOVA revealed various significant effects on the two different scales. All pairwise comparisons described below used Bonferroni as adjustment for multiple comparisons, and all means mentioned in the following section are estimates means. When it comes to the well-being scale (see Table 1), significant effects of the portrayed individuals' mood,  $F(2, 52) = 281.26, p < .001, \eta^2 = .91$ , were found on how pictures were rated. Pairwise comparisons revealed that ratings of happy individuals ( $M = 9.8, SE = .23$ ) and of neutral individuals ( $M = 6.6, SE = .22$ ) were significantly higher ( $p < .001$ ) than ratings of sad individuals ( $M = 3.8, SE = .22$ ). There was also a significant difference ( $p < .001$ ) between the ratings of happy people and of neutral people on this scale.

Table 1

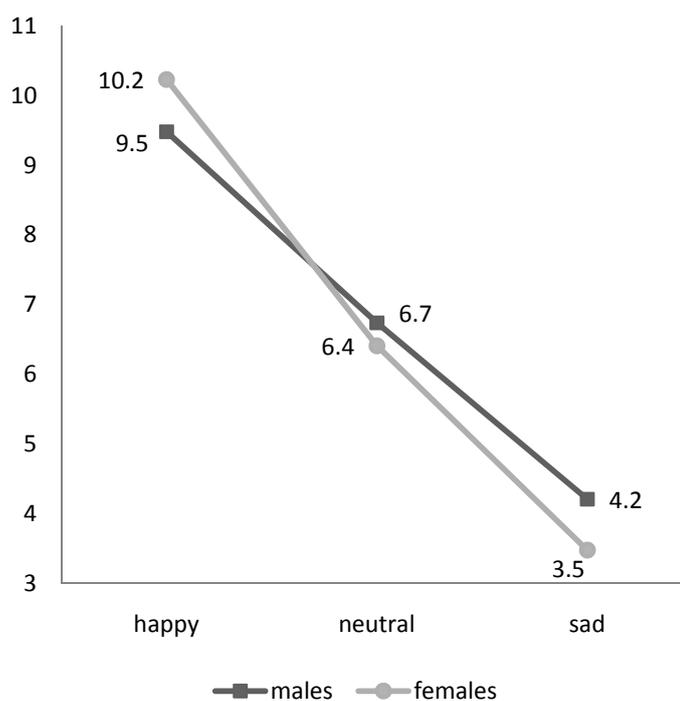
#### Significant Results on Well-Being Scale

| Variable(s)   | Df    | F         | $\eta^2$ |
|---|-------|-----------|----------|
| portrayed individuals' mood                                 | 2, 52 | 281.26*** | .91      |
| portrayed individuals' mood * participants' gender          | 2, 52 | 4.6**     | .15      |
| portrayed individuals' gender * portrayed individuals' mood | 2, 52 | 3.34*     | .11      |

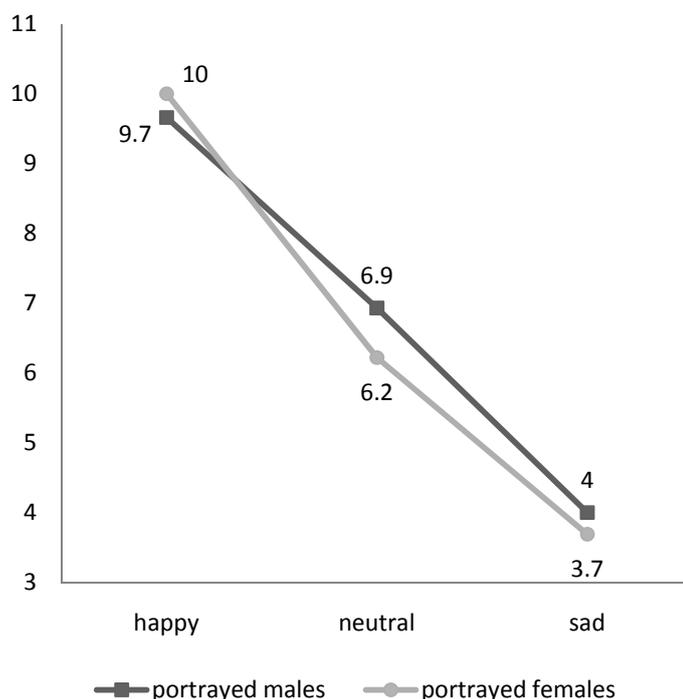
Note. \* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$

There was a significant interaction effect,  $F(2, 52) = 4.6, p < .05, \eta^2 = .15$  between the mood people were in on the pictures and the participants' gender. This indicates that the mood was rated differently depending on which gender participants had. Pairwise comparisons revealed that the differences between ratings of all moods were significant ( $p < .001$ ) for both genders, and as can be seen in figure 2, women rated pictures more extremely than men.

The last significant result found on the well-being scale revealed an interaction effect between the portrayed individuals' gender and the mood they were in,  $F(2, 52) = 3.34, p < .05, \eta^2 = .11$ . This indicates that the mood was rated differently depending on which gender the portrayed individuals had. Pairwise comparisons revealed that the differences between ratings of all moods were significant ( $p < .001$ ) for both genders, and as can be seen in figure 3, women were rated more extremely than men.



**Figure 2.** Interaction effect between the mood showed on pictures and the participant's gender on the well-being scale



**Figure 3.** Interaction effect on the well-being scale between the portrayed individuals' gender and their mood

As for the energy scale, several significant results were revealed as well (see Table 2). As with the well-being scale, significant effects of the portrayed individuals' mood were found on how pictures were rated,  $F(2, 52) = 50.75, p < .001, \eta^2 = .65$ . Pairwise comparisons revealed that ratings of happy individuals ( $M = 8.9, SE = .32$ ) were significantly higher ( $p < .001$ ) than ratings of neutral ( $M = 6.3, SE = .24$ ) and sad people ( $M = 5.5, SE = .24$ ). There was also a significant difference ( $p < .05$ ) between the ratings of neutral and sad people on this scale.

Table 2

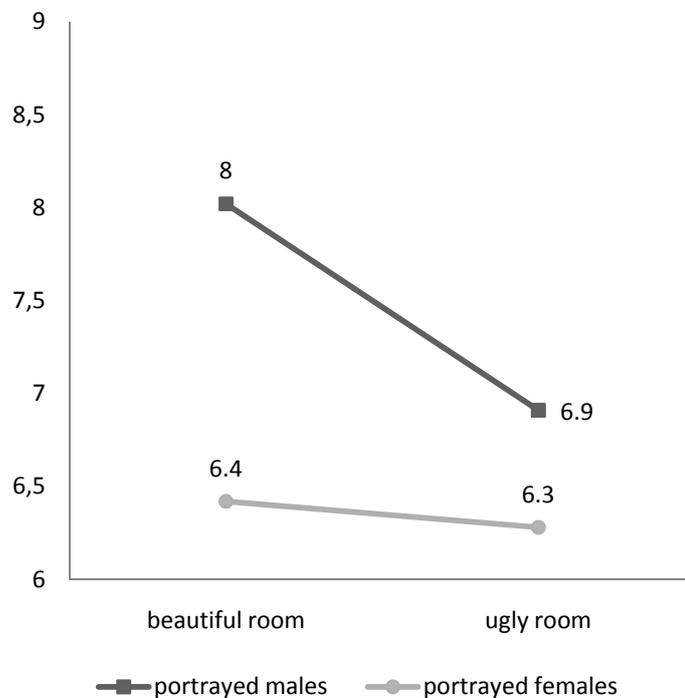
*Significant Results on Energy Scale*

| Variable(s)   | Df    | F        | $\eta^2$ |
|---|-------|----------|----------|
| portrayed individuals' mood                                 | 2, 52 | 50.75*** | .65      |
| portrayed individuals' gender                               | 1, 26 | 29.38*** | .52      |
| portrayed individuals' gender * group                       | 1, 26 | 5.5*     | .17      |
| portrayed individuals' gender * portrayed individuals' mood | 2, 52 | 12.25*** | .31      |

Note. \* $p < .05$ , \*\*\* $p < .001$

A main effect of gender,  $F(1, 26) = 29.38, p < .001, \eta^2 = .52$ , showed that males ( $M = 7.47, SE = .21$ ) were rated significantly higher than females ( $M = 6.35, SE = .19$ ) on the energy scales.

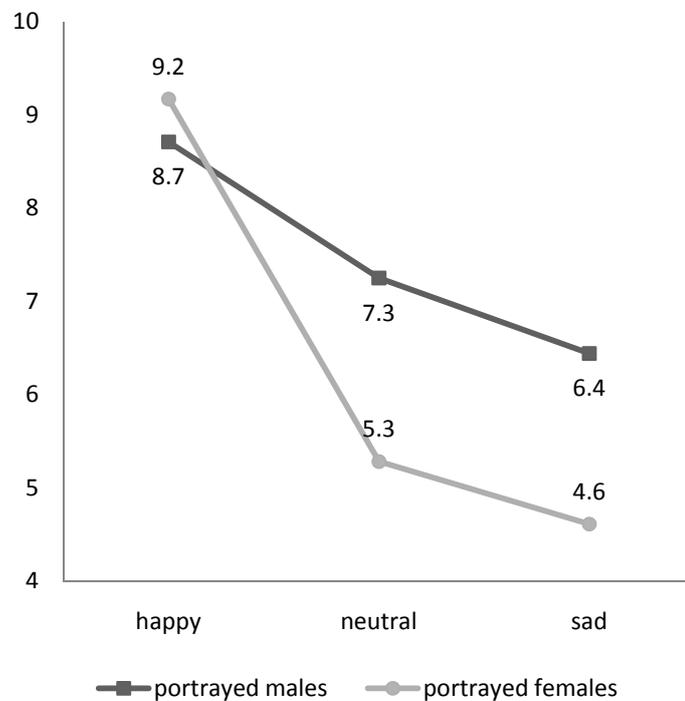
A significant interaction effect was found between the portrayed individuals' gender and the group participants belonged to (beautiful room or ugly room),  $F(1, 26) = 5.5, p < .05, \eta^2 = .17$ . Pairwise comparisons revealed that males were rated significantly higher when participants were in the beautiful room ( $M = 8.02, SE = .28$ ) than when participants were in the ugly room ( $M = 6.91, SE = .3$ ). Looking at the interaction graph (see Figure 4), this effect reflects that portrayed males were rated significantly lower in the ugly room, but that portrayed females' ratings remained almost unaffected by the environment ( $p > .05$ ).



**Figure 4.** Interaction effect on the energy scale between the group participants were in (beautiful room or ugly room) and the gender of the individuals being rated

Finally, a significant interaction effect was found between the portrayed individuals' gender and their mood,  $F(2, 52) = 12.25, p < .001, \eta^2 = .31$ . This indicates that the mood was rated differently depending on which gender the individual on the picture had. Pairwise comparisons revealed that the difference between happy and neutral

females, as well as between happy and sad females, was significant ( $p < .001$ ). The pairwise comparisons also showed that the difference between happy and neutral males, as well as between happy and sad males was significant ( $p < .01$ ). The difference between the neutral and sad individuals was not significant for both genders ( $p > .05$ ). The difference between how both neutral and sad males and females were rated was also significant ( $p < .001$ ). Happy males and females were however not rated in a significantly different manner ( $p > .05$ ). Looking at the interaction graph (see Figure 5), these effects reflect that neutral females ( $M = 5.28$ ,  $SE = .34$ ) were rated significantly lower than neutral males ( $M = 7.25$ ,  $SE = .26$ ). It also shows that sad females ( $M = 4.61$ ,  $SE = .24$ ) were rated significantly lower than sad males ( $M = 6.44$ ,  $SE = .39$ ).



**Figure 5.** Interaction effect between the portrayed individuals' gender and the mood they were in on the energy scale

### *Discussion*

The results of this experiment show that Maslow and Mintz's experiment could not entirely be replicated using VEs. This could imply VEs cannot be used to replace a real environment. However, the significantly lower rating of pictures portraying males while participants were in the ugly room suggests that more research on the subject is needed in order to understand how VEs influence people. One could interpret these results as a partially successful replication of the initial study, but to a lower degree. As to why the ugly environment had a larger effect on pictures showing males than on pictures showing females, the cause would need to be explored further.

The significant differences found between the different moods on each scale are along the lines of what could logically be anticipated; sad individuals should receive lower ratings than happy or even neutral individuals. As to why the difference between the neutral and sad individuals on the energy scale was significant at a lower level, one could presume that sad individuals still exude some level of energy, albeit negative, while well-being does not emanate from sad individuals.

Research by Johnson and Shulman (1988) seems to be consistent with the interaction effect between the portrayed individuals' mood and the pictured individuals' gender found on the well-being scale. Johnson and Shulman (1988) found that women are usually rated as experiencing both positive and negative feelings more intensively than men, which is consistent with the more extreme ratings that women received on the well-being scale. As for the interaction effect between the portrayed individuals' gender and their mood on the energy scale, some previous research also is consistent with this result. For instance, a study by Plant, Hyde, Keltner and Devine (2000) showed women were rated significantly sadder than men, whether the facial expression was ambiguous or not. In this study, women were rated significantly lower than men on neutral and sad facial expressions.

This same study by Plant et al (2000) could also partially explain why the males' average rating was higher than females' average rating on the energy scale. Since a sad woman's picture is usually rated more negatively than a picture of a sad man, this could explain why the average rating, even if all 3 moods were included, was significantly lower. One could assume that the lower ratings on both neutral and sad facial expressions brought down the average rating for women even when the happy facial expression was included, hence the significant difference between males and females.

As for the differences in how male and female participants rated the pictures, the result could be in line with Thayer and Johnsen's study where they found that women could recognize and classify emotions more accurately than men. Although Thayer and Johnsen's study observed accuracy in recognizing emotions, one could

speculate that women's ratings, which differed from men's rating, were also more accurate, which consequently led to the interaction effect.

#### *Limitations*

The lack of complete replication of the results could be due to various factors. Various aspects of Maslow and Mintz's study were not replicated with the exact parameters of the original experiment. For instance, the rooms shown on the screen were not identical to those used in the original study. Although they were proven to differ significantly in terms of aesthetics, they were not pictures of the actual rooms used originally. Although the beautiful room had a similar design (two sources of natural light, beige walls, comfortable furniture, drapes, paintings, and so on), the ugly room differed more noticeably (no janitor equipment, no dirty bed, and so on). Surprisingly, the ugly room received a quite high score on the aesthetics scale (3.67) which is, if rounded up, on the positive part of the scale. The fact that the room was not perceived as ugly on an average could explain why the result could not be completely replicated.

Another difference from the original study is the set of pictures of people used. This study used black and white prints rather than negative prints for the reasons mentioned above in the method. However, this could have influenced the way participants evaluated the pictures. Furthermore, due to the lack of information about the original pictures (facial expression, ethnicity, size, and more), it is quite difficult to know to what extent the pictures used in this experiment were similar to the original ones. A final point about the pictures is moreover that no dummy pictures were used for reasons mentioned earlier in the method, which could also have influenced the final scores.

Other rather important dissimilarities in the experiment's design are the much shorter introductory stage, as well as the use of a questionnaire participants filled in themselves rather than the use of a naïve examiner interviewing the participants. It was decided to have a shorter introductory stage in order for the examiner to influence the participants as little as possible. One can argue for obvious reasons that Maslow and Mintz could have influenced their participants and found significant differences due to this questionable design. Although they meant to limit their influence on the participants during the experiment by using a naïve examiner, it is not guaranteed that this individual did not influence the participants either. Maslow and Mintz raise the question themselves concerning the possibility of participants to have been biased by the examiner based on their longer exposure time to the environment. This and other factors caused by the examiner could have created the significant differences found in the original study.

However, although the design chosen for this experiment presumably reduced the risks for participants being influenced by the examiner, it also may perhaps have resulted in a higher anxiety level compared to the original study. The

reasons why Maslow and Mintz spent time with the participants prior to the actual experiment were to calm them down in order to avoid task orientation and anxiety. The anxiety level was controlled in this study with the STAI-S and did not differ significantly between the groups. Furthermore, it was on the lower part of the scale, which could indicate a low level of anxiety. Alternatively, it is not possible to know if the participants of the actual study had an anxiety rate similar to that of the original participants. As for task orientation, even if participants of the present study were asked to be spontaneous and encouraged to be relaxed, it is not possible to know how task oriented they were, and how much they differed from the original participants.

In the original study, it was strongly suggested to the participants to choose their own terminology as to evaluate the pictures. Although participants in the actual study were told to think with synonyms or their own terms of the words used on the printed scales, it did not leave as much freedom of choice and could have been too restrictive compared to the original study. Finally, although both samples constituted of students, 50% of the original participants were females and 50% were males. However, 71% of the participants for this study were females, which also could be a factor influencing the incongruent results.

With regards to those limitations, the significant interaction effect found between the room and the pictured individuals' gender on the energy scale points towards the need for more research in the area in order to understand more clearly what the appropriate usages of a VE can be. It is possible that the results of the original study could not be replicated due to either the inefficacy of the original experiment or due to a lack of effect of the VE. It is however possible that only a small part of a real effect was found in the interaction effect due to the design of the actual experiment itself and even due to the small size of the sample, for instance. In other words, a different and improved design may possibly find stronger significant differences given that there is an actual effect to be measured.

The fact that a weaker effect could be detected in this study is in agreement with other studies. For example, Kjellgren & Buhrkall (2010) showed that although a virtual natural environment could be used for attention restoration, its effects were not as strong as those of a real natural environment. Ideally, in order to understand the VR apparatus used in the experiment better, two exactly identical measurements should be performed using a real environment and, in the second condition, an exact representation of that same environment. This would allow for a clearer comparison.

#### *Further research*

The results of this study suggest that until VR is better understood, research using this instrument should be done carefully with much attention given to the validation of the instrument used. More research in the area is obviously needed in order to validate different VEs as reliable tools for research in psychology. For instance,

within-group designs comparing the scores of participants in a real environment with the scores obtained in a virtual representation of that same environment could be an efficient way to test VR's validity further. Moreover, VR instruments are diverse and numerous, which of course requires validation for each and every one of them. It is clear that validation studies are required for various combinations of different VR technologies combined with diverse VEs. One could also question which role the represented environment has; could some environments be easier to represent virtually than others, for instance? Research possibilities in the field are still many and it is necessary to explore them more thoroughly.

## References

- Becker, D.V., Kenrick, D.T., Neuberg, S.L., Blackwell, K.C., & Smith, D.M. (2007). The Confounded Nature of Angry Men and Happy Women. *Journal of Personality and Social Psychology*, 92 (2), 179-190.
- Bell, P.A., Green, T.C., Fisher, J.D., & Baum, A. (2001). *Environmental Psychology* (5<sup>th</sup> ed.). Fort Worth, TX: Arcourt College Publishers.
- Blade, R.A., & Padgett, M.L. (2002). Virtual Environments Standards and Terminology. In K.M. Stanney (Ed.), *Handbook of Virtual Environments: Design, Implementation, and Applications* (pp. 15-27). Mahwah, NJ: Lawrence Erlbaum Associates.
- Botella, C., Baños, R., Perpiñá, C., Villa, H., Alcañiz, M. & Rey, A. (1998<sup>a</sup>). Virtual Reality Treatment of Claustrophobia : A Case Report. *Behaviour Research and Therapy*, 36 (2), 239-246.
- Botella, C., Perpiñá, C., Baños, R.M., & García-Palacios, A. (1998<sup>b</sup>). Virtual reality : a new clinical setting lab. *Studies in Health Technology and Informatics*, 58, 73-81.
- Darken, R.P., & Banker, W.P. (1998). Navigation in Natural Environments : Training Transfer Study. In *Proceedings of VRAIS -98*, IEEE Computer Society Press, 12-19.
- De Kort, Y.A.W., Meijnders, A.L., Sponselee, A.A.G., & IJsselsteijn, W.A. (2006). What's wrong with virtual trees? Restoring from stress in a mediated environment. *Journal of Environmental Psychology*, 26, 309-320.
- Fazio, R.H. (2001). On the automatic activation of associated evaluations: An overview. *Cognition and Emotion*, 15 (2), 115-141.
- Johnson, J.T., & Shulman, G.A. (1988). More Alike than Meets the Eye: Perceived Gender Differences in Subjective Experience and Its Display. *Sex Roles*, 19, 67-79.
- Kjellgren, A., & Buhrkall, H. (2010). A comparison of the restorative effect of a natural environment with a simulated natural environment. *Journal of Environmental Psychology*, doi:10.1016/j.jenvp.2010.01.011
- Martin, C.L. (1987). A Ratio Measure of Sex Stereotyping. *Journal of Personality and Social Psychology*, 52 (3), 489-499.
- Maslow, A.H., & Mintz, N.L. (1965). Effects of Esthetic Surroundings: I. Initial Short-Term Effects of Three Esthetic Conditions upon Perceiving "Energy" and "Well-Being" in Faces. *Journal of Psychology*, 41, 247-254.
- North, M. & North, S. (1996). Virtual Reality Psychotherapy. *The Journal of Medicine and Virtual Reality*, 1, 28-32.
- Plant, E.A., Hyde, J.S., Keltner, D., & Devine, P.G. (2000). The Gender Stereotyping of Emotions. *Psychology of Women Quarterly*, 24, 81-82.

- Rothbaum, B., Hodges, L., Kooper, R., Opdyke, D., Williford, M. & North, M. (1995). Effectiveness of Computer-Generated (Virtual Reality) Graded Exposure in the Treatment of Acrophobia. *American Journal of Psychiatry*, 152, 626-628.
- Ruddle, R.A., & Lessels, S. (2009). The Benefits of Using a Walking Interface to Navigate Virtual Environments. *ACM Transactions on Computer-Human Interaction*, 16 (1), 5:1-5:18.
- Sadowski, W., & Stanney, K. (2002) Presence in Virtual Environments. In K.M. Stanney (Ed.), *Handbook of Virtual Environments: Design, Implementation, and Applications* (pp. 791-806). Mahwah, NJ: Lawrence Erlbaum Associates.
- Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). *Manual for the state-trait anxiety inventory*. Palo Alto, CA: Consulting Psychologists Press.
- Stanney, K.M., & Zyda, M. (2002) Virtual Environments in the 21<sup>st</sup> Century. In K.M. Stanney (Ed.), *Handbook of Virtual Environments: Design, Implementation, and Applications* (pp. 1-14). Mahwah, NJ: Lawrence Erlbaum Associates.
- Thayer, J.F., & Johnsen, B.H. (2000). Sex differences in judgement of facial affect: A multivariate analysis of recognition errors. *Scandinavian Journal of Psychology*, 41, 243-246.