

How Colorful Was Your Day? Why Questionnaires Cannot Assess Presence in Virtual Environments

Abstract

This paper argues that a scientific basis for “presence” as it’s usually understood in virtual environments research, cannot be established on the basis of postexperience presence questionnaires alone. To illustrate the point, an arbitrary mental attribute called “colorfulness of the experience” is conjured up, and a set of questions administered to 74 respondents with an online questionnaire. The results suggested that colorfulness of yesterday’s experiences was associated with the extent to which a person accomplished their tasks, and also associated with yesterday being a “good,” “pleasant,” but not frustrating day. The meaninglessness of this analysis illustrates that the equivalent methodology used by presence researchers, may, similarly, bring into being the idea of presence in the minds of VE participants. However, it is argued that there can be no evidence on this methodological basis that presence played any role in their actual mental activity or behavior at the time of the experience. It is concluded that presence researchers must move away from heavy reliance on questionnaires in order to make any progress in this area.

I Introduction

The question of “presence” arises for virtual environments (VEs) because VEs override sense impressions from the real world with those generated by computer display systems in (ideally) several sensory modalities. The body and head are tracked, so that sensory data may be delivered to participants as a function of the position and orientation of sense organs such as their eyes and ears. Since the primary sense data from which people form perceptions would have been determined in

this way, it is natural to ask how successful this has been—i.e., whether the person’s *presence* has actually been altered away from the real world and into the virtual world. The issue of presence arises whenever there is this potential choice between the formation of perceptions based on inevitable leakage of real-world sensory data into the sensory stream available to the participant, or alternatively based on the virtual sensory data (Slater & Steed, 2000).

A scientific approach to the study of presence should be concerned with ways of characterizing what exactly is meant by the notion of “overriding sense impressions by virtually generated sensory data,” with ways of quantifying and classifying virtual sensory data, and also by understanding what is meant by “successful” in this context. What would it mean to say that real sense impressions had been “successfully overridden” through generation of virtual sensory data? One approach is to concentrate all resources on asking people about their experiences—that is, subjecting them to VEs and then administering questionnaires. This paper will argue that this almost universal method of measurement, based on questionnaires, does not and cannot measure “presence” and indeed that *sole reliance* on questionnaire results cannot be used to verify that “presence” actually exists as a phenomenon. Our major argument is that after-the-event questionnaire-based measures cannot in principle rule out the possibility that the reported “presence” was called into being simply by its having been asked about. If this is the case, then it becomes at best a post hoc construction that does not relate to any specific brain activity or state that was in process during the

Mel Slater

m.slater@cs.ucl.ac.uk
Department of Computer Science
University College London

actual experience to which it was supposed to have related. Its scientific and practical relevance then becomes dubious.

The most ubiquitous component of definitions of the concept of presence in virtual VEs is the idea that it is a “sense of being there” in the situation depicted by the VE display systems (Held & Durlach, 1992; Sheridan, 1992). It is considered important by those who study it, because it is tantamount to a “common currency” for VE applications—the one result of a VE experience that may be universal independent of application and other aspects such as “task performance.” Hence, the idea of trying to discover how to “maximize presence” is thought to be a useful goal of VE research—especially since presence is likely to be associated with behavior that is appropriate to the situation. In this case it would be useful for training—where the experience in the VE should be as close as possible to the experience in the corresponding real world situation, so that whatever is learned in the VE is transferred positively to appropriate behavior in the real world.

In the next section we briefly outline some of the major ideas about presence and its measurement. In Section 3 we provide some arguments against the sole use of questionnaires in presence research. In Section 4 we discuss the method used to support the claim that we can have no evidence that “presence” exists outside of the minds of presence researchers. We make up an arbitrary concept (“colorfulness of the experience”), construct a questionnaire to “measure” this phenomenon, and show how (in Section 5) results are obtained that mirror results with presence questionnaires. We draw conclusions regarding the future of presence research in Section 6.

2 Background

There are several different meanings and indeed philosophical standpoints associated with the concept of presence. A review by Draper, Kaber and Usher (1998) defines “experiential presence” as “a mental state in which a user feels physically present within the computer-mediated environment.” This is a way of expressing the common view that presence is the sense of “being

there” in the virtual environment or, similarly, the sense of being in the place depicted by the virtual reality rather than in the real physical place where the participant’s body is really located (for example, Barfield & Weghorst, 1993). A very similar idea is expressed by Lombard and Ditton (1997) who refer to presence as a “perceptual illusion of nonmediation,” so that the participant has the sense of directly operating in the world depicted by the VE technology and loses awareness of the display and interaction systems.

A fundamentally different view (Zahoric & Jenison, 1998; Flach & Holden, 1998) is that presence is “. . . tantamount to successfully supported action in the environment.” It is argued that reality is formed through action, rather than through mental filters and that “. . . the reality of experience is defined relative to functionality, rather than to appearances.” What is important in this approach is action (how things are done) and the affordances offered in the virtual environment, rather than just appearances, and that the sense of “being there” is grounded on the ability to “do” there.

The “embodied presence” approach is similar to this (Schubert, Friedmann, & Regenbrecht, 2001) and to the idea of Slater, Usoh, and Steed (1995), where it is argued that it is essential for there to be a close match between kinesthetic proprioception and the stream of sensory data: if the participant is supposed to be moving through an environment by walking, then when visual flow indicates walking, presence will be higher the more the person’s bodily movements correspond to real walking (Usoh et al., 1999).

A distinction is made between immersion and presence by many researchers (e.g., Slater & Wilbur, 1997; Draper et al., 1998). Immersion is a description of overall fidelity in relation to physical reality provided by the display and interaction systems. In this view, presence research is essentially that of carrying out experiments that manipulate the variables that make up immersion, in order to build an equation with presence on the left-hand side, and the factors of immersion on the right-hand side. Individual psychological differences between people can be also included as variables on the right-hand side. This may be a worthwhile effort in the quest to produce a statistical model of how presence and im-

mersion may be related, based on empirical data, but it does not enhance understanding of the processes involved.

In such studies, a major challenge is how to measure presence at all. The normal approach is to use questionnaires and anchored ordinal scales: the questions relating to the extent of “being there” in the virtual environment, and variations on this and other presumed aspects of presence. Schloerb (1995) has proposed an alternative psychometric approach, based on the idea of just noticeable differences between virtual and real worlds, but to our knowledge this has never been followed up—probably because there are no VEs available that approach real-world experiences (other than flight simulators). Regarding questionnaires, Freeman, Avons, Pearson, and IJsselsteijn (1999) have shown the inherent instability of questionnaire-based presence assessment methods and propose a behavioral solution, and there have been several attempts to provide behavioral measures of presence (see Freeman, Avons, Meddis, Pearson, & IJsselsteijn, 2000). These typically introduce features into the environment that would cause a bodily response (such as swaying in response to a moving visual field, or ducking in response to a flying object). The problem with such measures is that they are not ubiquitous—special features have to be introduced into an application that may be tenuously connected to the application itself. The use of physiological measures as surrogates for presence is a significant advance (Meehan et al., 2002), but is limited to situations where the physiological response is obvious (e.g., a response to a feared situation). But what should be the expected physiological response to mundane situations such as being in a virtual room that has a table and some chairs?

A fundamental problem with the vast majority of presence research is that *there is no real evidence at all for the existence of the phenomenon*. As researchers in the field we “know” that it exists through introspection of our own states and behaviors while experiencing immersive virtual environments. Indeed such experiences prompted the author to become involved in this research. (For example, the author while recently in a VE depicting a bar, was feeling tired, and walked over to

one of the virtual chairs with the intention of sitting down, only to stop himself at the last moment.) Nevertheless, to turn this introspection into a valid scientific object of inquiry is a different matter altogether.

The problem is that we cannot rule out the possibility that presence in a VE may seem to exist in our experimental subjects simply because questions are asked about it. As an analogy, after an experience someone may be asked to attribute a score on a “colorfulness of the experience” scale. Of course they can attribute a score, and it may mean something to them, but this doesn’t imply that *during the course of the experience* that there was anything actually happening in terms of their mental activity or behavior that had anything at all to do with “colorfulness.” Also, without careful training, people can attribute their own meaning to “colorfulness.” So with presence: Is a person’s interpretation of “being there” that of involvement, or attention, or interest, or realism, or a confounding of all these and others?

A possible way out of this is by relating questionnaire-reported presence to other phenomena such as variations in levels of immersion. But even here there is a danger of circularity, because the displays *would have had greater fidelity* (e.g., been more panoramic) and therefore participants may conclude after the event (rather than during experience) that they must have “been there” because this is what is being asked of them. Again there is no independent evidence that anything is actually happening at the level of brain activity or behavior that can be characterized as a unique psychological or behavioral process that we would call “presence.” This point, about relating presence to different levels of immersion, or “task performance” will be returned to later.

3 Against Presence Questionnaires

Questionnaires are a very useful tool in social and psychological research. They rely on respondents being able to reasonably compare a given situation with a number of other situations, and then (usually) make a quantitative assessment, at least on an ordinal scale. For

example, voting intentions can be reasonably queried, because respondents can assess how they would vote if there were actually an election in the near future. Of course, there is no guarantee that their actual behavior would conform to their present-moment intentions—but the relationship is good enough for election polls to normally be a very good indicator of the outcome of elections. Similarly, a psychological questionnaire about “anxiety” in a given situation can reasonably be employed: a respondent can compare, from past experience, the degree of anxiety engendered by similar situations, and at least rank a given situation in relation to such past experiences. Indeed anxiety can be associated with such quantitative variables as the perceived heart rate, degree of blushing, degree of sweating, stomach palpitations, and so on, that give an overall impression to the individual that quite naturally can result in a quantitative assessment. In other words, questionnaires can be useful in circumstances where there is a stock of experience against which to judge a given experience, and where comparisons can be made about a specific behavioral outcome—to vote, to be anxious, to favor a certain product, and so on.

This does not apply to “presence” for several reasons. In each of the above examples, there are known verifiable means of assessing the veracity of the questionnaire data—that go beyond just the questionnaires. For example, there will be a real election against which the voting intention questionnaire data can be judged. Stated consumer preferences can be compared to sales figures. There is an accepted paradigm about anxiety that relates it to physiological and behavioral responses that can be compared with questionnaire data, and, where correlations are known and established, questionnaires can be used to pretty well predict the likelihood of anxiety states. However, if in presence research we stay within the confines of questionnaires, there are no known external, nonquestionnaire data against which to verify even the existence of the phenomenon. This is not to say that such verifiable data can never exist, but that since almost all presence data is questionnaire data, we have no independent verifiable information against which to judge it. In other words the existence of questionnaire data in itself is insufficient to show that there

is any actual phenomenon that is associated with specific identifiable brain activity, nor is there observable behavioral data of which we could say: “This indicates presence.” The only exceptions to this are the studies that use anxiety responses as surrogates for presence in extreme circumstances (such as facing a precipice in the middle of a room—Meehan, Insko, Whitton, & Brooks, 2002).

It might be argued that we can correlate presence with task performance. However, it has been argued before (Slater & Wilbur, 1997) that there is no logical connection between presence and task performance. The latter is a function of user interface, not presence. A person can be in real life using a very poorly designed ATM for example, with very low task performance (e.g., they are unable to withdraw their cash). Clearly this problem is not a function of presence but of the user interface to the device.

Another way to introduce task performance is in relation to skill transfer from the virtual to the real world. As an example, suppose we wish to test the hypothesis that increased presence increases the chance of such skill transfer. One way to go about this is to set up a training program, and then after each person’s virtual environment training experience administer a presence questionnaire, and then examine how this correlates with their real-world task performance. We may find a positive correlation—but actually if we did, it would not be informative since we would have no idea why there was such a correlation, for in such an experiment there would be no manipulated variables. At best we could say at the end of such a process: “those people who were more inclined to report a high level of presence, also exhibited better task performance.” It is not clear what the value of such a statement would be. Is reported “presence” in such a case just another way to describe the subjects’ experiences of being at ease with a task, and it was highlighted because it was the only linguistic device available to them within the context of a questionnaire? Maybe it just means “feeling comfortable” or “finding it easy to operate the controls.” Is this presence?

More likely we would set up the experiment in order to try to manipulate presence: we could take aspects of

immersion and control them in a between-groups experiment. For example, we could vary the field-of-view, or the frame-rate, or any combination of “immersion” factors. Then we could again administer the questionnaires and proceed as before. Now suppose we find positive correlations between each of, say, wider field-of-view, presence, and task performance. While this is very interesting regarding the impact of field-of-view, it is much less interesting in regard to presence. Is field-of-view “causing” both the higher reported presence and the higher task performance? Since we cannot independently manipulate presence (because it is only a questionnaire response) we cannot carry out an experiment to separate out the possible causality relations. Durlach has argued on several occasions at conference talks that presence may be at best an “intermediate variable” with no explanatory power of its own, and this view is also stated in Durlach and Mavor (1994).

In Usoh, Catena, Arman, and Slater (2000) it was argued that, given the state of VE technology today, any presence questionnaire ought to be able to distinguish experiences in the real world from experiences in the virtual world. A between-groups experiment was conducted where one group carried out a task in a real place, and another group carried out the same task in a virtual and immersive simulation of that place. Both groups were given two sets of presence questionnaires (in randomized order) after their experience. There was no significant difference in the reported questionnaire-based presence between the real world and virtual world groups. It was argued that the result could be explained by people trying to make sense of a question, even if it is absurd. To ask someone in a real-world place about their sense of being in the real world place, is ultimately absurd—because they and everyone knows that indeed they were there. So respondents reinterpret presence to mean, for example, their sense of comfort, lack of alienation, involvement, similarity to some idealized experience, and so on. It was argued that participants will tend to try to interpret the questions in ways that make sense—since after all these questions are administered by serious scientists in a laboratory setting. Another explanation of the same result, however, is that indeed the degree of presence in the two situations was the same—

simply because very minimal cues are needed to establish presence (discussed in Slater, Steed, Chrysanthou, 2001, Chapter 1). There is no way to distinguish between these two explanations from the data generated by the experiment.

If the Usoh et al. paper (2000) showed that presence questionnaires do not reliably distinguish between real and virtual experiences we have to consider whether these questionnaires are measuring anything interesting at all. In this paper we take the critique one step further by demonstrating that it is easy to make up an arbitrary concept, administer a questionnaire about it, and thereby see the “phenomenon” conjured into being by the very act of asking about it.

4 Method

A phenomenon was invented called “colorfulness of an experience.” It is left undefined—but on the other hand also deliberately chosen so that it may relate to something real such as mood. Is a “feeling of being there” any more precise or well-defined?

A questionnaire was established that attempts to assess the degree of “colorfulness” that a person will attribute to their experiences on a certain day (“yesterday”). A colorfulness score was constructed from the questionnaire responses, and then a regression analysis carried out to examine the association between other variables in the questionnaire and colorfulness. A model was established that explains some of the variation in colorfulness that could be attributed to these other variables. Thus colorfulness is established in much the same way as presence is established. In this process the degree of experienced “colorfulness” was brought into being only by asking about it—having no predictive or explanatory power, and no utility in itself. We cannot know from the questionnaire data alone whether it corresponds to anything real in the mental and behavioral activity of the subjects. Putting it yet another way, “colorfulness” was not a factor in the day of the people concerned before they answered the questionnaire—or at least, if it was, we have no way of knowing this. Similarly, we cannot know whether “presence” was involved

in any way in the experiences of the people in an experiment—before we ask about it (if asking about it is the only method we employ).

An email message was sent to the staff list in the Department of Computer Science, University College London (this list includes all members of academic staff, researchers, PhD students, technicians, and administrators). This invited them to complete an online questionnaire, without explanation.

The questionnaire was in three parts. Part A asked for basic information including age, gender, and job. Part B was the main questionnaire, which consisted of 13 questions on a 1 to 7 scale, where 1 = “not at all,” 7 = “a great deal.” Four of the questions directly asked about the degree of “colorfulness” experienced on the previous day. There was one metaquestion regarding the utility of the concept as a description of the day’s experiences. These questions were:

- 1) Think back to yesterday. How colorful was your day?
- 2) Were there times during the day that you would describe as having been colorful?
- 3) To what extent were there times during the day that you felt were overwhelmingly vivid?
- 4) When you review the events of yesterday in your mind’s eye, how colorful are the images?

The meta question was:

- 5) To what extent is “colorful” an appropriate metaphor to describe yesterday’s experiences for you?

The remaining questions were filler questions—simply to put gaps between the questions above, and also to obtain any other quite arbitrary information that occurred to the author in the 10 minutes devoted to constructing the questionnaire:

- 6) Yesterday, to what extent did you accomplish whatever you set out to do?
- 7) Overall would you describe your day yesterday as having been a pleasant day for you?
- 8) Were there times during the day yesterday when you wished you had chosen different clothes to wear?

- 9) Yesterday did you get up later than usual for the day of the week?
- 10) Last night did you go to bed later than normal for that day of the week?
- 11) Yesterday did you do more traveling than usual?
- 12) Was yesterday a more frustrating day than usual?
- 13) Overall, was yesterday what you would call a “good” day for you?

Part C asked respondents to write any other comments that they wished regarding the issue of the main questionnaire.

Two methods were used to construct the colorfulness score. First, following, for example, Slater et al. (1995), the number (r) of high scores (of 6 or 7) out of the ($n = 4$) questions giving as a response a count variable ($r = 0, 1, 2, 3, \text{ or } 4$) that can be used in binomial logistic regression. This strategy has been used several times, and avoids treating the ordinal response variables as if they were interval variables. Alternatively, the colorfulness scores were averaged over the four questions, assumed to be a normally distributed response to which classical linear regression can be applied (this is what most researchers would do). Explanatory variables were taken to be the demographic variables from Part A, and any of the noncolorful variables from Part B.

5 Results

5.1 Descriptive Statistics

The number of respondents was 74. Table 1 shows the means, standard deviations, and 95% confidence bounds for these variables. It is clear that the first two of these variables have means significantly higher than 4.0 (the central value), and two others are significantly lower than the central value—suggesting that people may be assigning meaning to these scales (otherwise results would cluster around the central value of 4). Table 2 shows the correlation matrix (all entries are significant at 1%) indicating consistency among the response variables, and a positive correlation of each of them with the metaquestion (5).

Table 1. Frequency Tables, Mean + SD, 95% Confidence Intervals of the Color Questionnaire Responses

| Response | QUESTION NUMBER | | | | |
|----------|-----------------|-----|-----|-----|-----|
| | 1 | 2 | 3 | 4 | 5 |
| 1 | 1 | 4 | 14 | 4 | 17 |
| 2 | 7 | 3 | 16 | 8 | 26 |
| 3 | 11 | 10 | 11 | 16 | 11 |
| 4 | 13 | 8 | 13 | 13 | 3 |
| 5 | 19 | 18 | 13 | 19 | 11 |
| 6 | 15 | 17 | 3 | 8 | 2 |
| 7 | 8 | 14 | 4 | 6 | 4 |
| Mean | 4.6 | 4.9 | 3.3 | 4.1 | 2.8 |
| SD | 1.5 | 1.7 | 1.7 | 1.6 | 1.7 |
| 95% CI: | | | | | |
| Lower | 4.3 | 4.5 | 2.9 | 3.8 | 2.4 |
| Upper | 5.0 | 5.3 | 3.7 | 4.5 | 3.2 |

Table 2. Correlation Matrix Among Colorful Questions

| Question | 1 | 2 | 3 | 4 | 5 |
|----------|------|------|------|------|------|
| 1 | 1.00 | .60 | .38 | .60 | .50 |
| 2 | | 1.00 | .40 | .59 | .46 |
| 3 | | | 1.00 | .45 | .55 |
| 4 | | | | 1.00 | .67 |
| 5 | | | | | 1.00 |

Note. $r > .29$ is significantly different from 0 at 1%; $n = 74$.

5.2 Factors Contributing to Colorfulness

A logistic regression was carried out with the count response variable being the number of high scores out of the four questions. The first point of interest was the activities that people might do that would be associated with their reported degree of colorfulness the day before. The activities available from the questionnaire data were

- Accomplishment: question 6
- Clothes: question 8
- GetUpLater: question 9

- GoToBedLater: question 10
- Traveling: question 11

These are all activities that a person can actually carry out, rather than just how they feel.

A logistic regression found that Accomplishment and GetUpLater were significantly and positively associated with the colorfulness count. The overall deviance was 135 on 71 d.f. The deviance is approximately χ^2 with 71 d.f., so that the overall fit is not good, indicating that there are other explanatory variables that need to be included. However, removal of Accomplishment results in an increase in deviance of 18 on 1 d.f., and similarly removal of GetUpLater results in an increase of 4.5 in the deviance on 1 d.f.—hence, each of these are clearly significant at the 5% level.

The overall picture here then is that if you want to improve your chance of having a colorful day, you should get up later than usual, and accomplish what you set out to do. The clothing you wear, how late you go to bed, and how much traveling you do, probably make no difference. There are other things you should do too, but on the basis of these data, we do not know what they are.

5.3 Surrogates of Colorfulness

How can we explain what people might mean by “colorfulness”? Here there are other variables that might help to understand this:

- Pleasant: question 7
- Frustrating: question 12
- Good: question 13

Due to correlations between these three variables, no two of them can be simultaneously fitted without one becoming insignificant. However, if we fit each separately we find that a colorful day is associated with not being frustrating (increase of deviance = 13 on 1 d.f.), or being “good” (increase of deviance = 34 on 1 d.f.), or being “pleasant” (increase of deviance = 30 on 1 d.f.). If a normal regression model is used then the only change to all of the above is that GetUpLater is no longer significant.

5.4 Relationship with the Metavariable

The proportion of respondents who positively went along with the idea that colorfulness was an appropriate metaphor to describe their experiences was 23% (those who answered with a score of more than 4 on question 5). The score on this metavariable was significantly correlated with the overall “colorfulness” score ($r = 0.61$ for the count response and $r = 0.69$ for the mean color score). We ought to eliminate the impact of this variable, since it could affect the results on the main response variables. This elimination can be accomplished by regressing the left- and right-hand side variables in the regression equations on the metavariable, and then working throughout with the residuals. (In the GLIM system that we use for analysis this is accomplished simply with the “eliminate” directive.)¹ When we do this the overall model fits are better. In the binomial model for the influencing factors Accomplishment remains significant, but not GetUpLater, and in the model for the surrogates, Good remains significant. The same results hold for the normal regression model.

5.5 Reliability

As a further check on the results, the original data set was split into two—each even numbered respondent’s results were put into one file, and all the odd numbered respondents’ were put into another file, thus creating two subsamples (“even” and “odd”) each of 37 responses. The analysis was repeated on each of the subsamples. For the “even” sample, the results were the same as above. For the “odd” sample, the Accomplishment variable was just below 5% significance in the logistic regression model ($\chi^2 = 3.822$ on 1 d.f., compared with the tabulated 5% value of 3.841) and not significant on the normal regression model. However, GetUpLater is significant on both normal and logistic regression models. Frustrating is not significant, but Pleasant and Good are significant. The results seem quite consistent with those of the full sample.

This analysis was taken further with the computation of Cronbach’s Alpha, which measures the internal consistency (or reliability) of the set of items. (Instead of splitting the data set just one way, it is split in all possible ways). When all five colorful variables are used (including the metavariable) Chronbach’s Alpha is 0.84, and is 0.80 when just the 4 colorful questions are used. These values are consistent with the notion that whatever is being measured, the measure is internally reliable given the convention that the alpha score should be at least 0.80. This isn’t too bad for an arbitrary made-up phenomenon.

6 Conclusions

So if you want to have a colorful day, which is something to do with having a good, pleasant, but not frustrating day, then at the very least you should accomplish what you set out to do. Getting up later than usual might also help.

But now seriously: obviously, this construct, “colorfulness” is entirely made up. The respondents to the questionnaire have attributed consistent meaning to it, and there are some relationships with other aspects of their behavior. The important point is that this supposed colorful attribute of their mental state did not exist prior to their being asked about it—or if it did, we have no way to know this. There is no evidence that this “colorfulness” existed as an identifiable mental activity or attribute during the actual experiences that the questionnaire addressed (those of “yesterday”). Methodologically, even if there were such identifiable mental properties, there is no way that a post hoc questionnaire could identify this fact, which is the critical point of this paper.

It could be argued that this analysis is unfair, since there was in fact no experiment and no independent factor was varied in an isolated way in order to examine its impact on colorfulness. We touched on this earlier in Section 2. Now imagine that in fact we had carried out an experiment—for example, a study is carried out where n people are each separately left sitting in a room

1. http://www.nag.co.uk/stats/GDGE_soft.asp

for a day, and another n people are individually in an identical room except that they are given some task to do which a normal person is likely to complete during the day. Suppose that the two groups are matched on a number of factors such as gender, age, occupation, and so on. The null hypothesis is that the colorfulness score of the second group will be the same as for the first group. It is highly likely that this hypothesis would be rejected in favor of the alternative that the second group scored higher on the colorfulness scale than the first. But so what? Somehow, some aspect of the experience of the people involved is being transformed into the colorfulness score. The “colorful” questions *are the only way that the participants can respond to their experiences* in the context of the study! This implies nothing about the specific brain states or behavioral or psychological mechanisms involved. It merely means that the only available linguistic category was indeed used to classify an experience. Knowing this does not add to our knowledge. At best we can say that “there was some difference in the experience of the two groups,”—which is almost the experimental equivalent of a logical tautology (since obviously the different groups experienced different circumstances).

Unfortunately the same is true for the vast majority of presence research. Presence has not been established in any way, by any researcher, as an identifiable mental attribute or mental activity. A possible exception is the use of physiological measures that indirectly capture presence: the argument is that because people experienced measurable anxiety when confronting a virtual precipice, they must have been present (Meehan et al., 2002).

It could be argued that actually the “colorfulness” experiment has found something real—that in fact people are on the average interpreting “colorfulness” in a consistent way, and that it is positively related to the variables such as “achieving what you set out to do.” This statement misses the point. All we would have shown is that people are making use of certain linguistic categories in a consistent way—things that go together have been grouped together. However, there is no mechanism here, no knowledge that at the time of the experience “colorfulness” was influencing anything, or

was influenced by anything, or existed in any way. Presumably now a whole “colorfulness” discipline could open up, where people could be inventing and calibrating new questionnaires, doing consistency checks, running experiments, finding new correlations, organizing International Societies and conferences, and so on. But really—what would such an effort gain in terms of scientific understanding of human behavior and mood?

The questionnaire methodology used almost exclusively in presence research is equivalent to the one that we have used above for colorfulness. We do not know that “presence” exists in any real form as something observable that happens when a person experiences a virtual reality. As far as we know it exists only because it has been conjured up by the researchers, who call it into being through their questions.

On the optimistic side, none of the above implies that presence is not in itself a useful concept, and it may be the case that presence will in future research be associated with observable mental states and activity. The major point of this paper is to raise issues about the validity of the methodology, not the concept itself. Moreover, we believe that many of the results previously found, such as the relationship between presence and whole body engagement, will eventually be verified through an improved methodology. From this point of view the previous questionnaire-based studies might be thought of as hypothesis generators rather than as reaching conclusions.

The growing group of researchers interested in studying the concept called presence might find a way to abandon the easy but ultimately useless employment of questionnaires, and search for a better way to capture this elusive concept. To return to the opening paragraph, we take the subject of presence as being about verifying the “success” of replacing real sense data with virtually generated sense data. There are many responses to a VE experience—gross behavioral, eye movements, measurable physiological responses, what people say in interviews, how people respond to questionnaires. Why elevate just one of these to be the preeminent position that it now has? Is it only because it is the easiest approach to take?

References

- Barfield, W., & Weghorst, S. (1993). The sense of presence within virtual environments: A conceptual framework. In G. Salvendy & M. Smith (Eds.), *Human-computer interaction: Software and hardware interfaces* (Vol. B) (pp. 699–704). Elsevier.
- Draper, J. V., Kaber, D. B., & Usher, J. M. (1998). Telepresence. *Human Factors*, *40*(3), 354–375.
- Durlach, N. I. & Mavor, A. S. (Eds.) (1994). *Virtual reality: Scientific and technological challenges*. Washington, DC: National Academy Press.
- Flach, J. M., & Holden, J. G. (1998). The reality of experience: Gibson's way. *Presence: Teleoperators and Virtual Environments*, *7*(1), 90–95.
- Freeman, J., Avons, S. E., Meddis, R., Pearson, D. E., & IJsselsteijn, W. A. (2000). Using behavioral realism to estimate presence: A study of the utility of postural responses to motion-stimuli. *Presence: Teleoperators and Virtual Environments*, *9*(2), 149–164.
- Freeman, J., Avons, S. E., Pearson, D. E., & IJsselsteijn, W. A. (1999). Effects of sensory information and prior experience on direct subjective ratings of presence. *Presence: Teleoperators and Virtual Environments*, *8*(1), 1–13.
- Held, R. M., & Durlach, N. I. (1992). Telepresence. *Presence: Teleoperators and Virtual Environments*, *1*(1), 109–112.
- Lombard, M., & Ditton, T. B. (1997). At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, *3*(2). Available from <http://www.ascusc.org/jcmc/vol3/issue2/lombard.html>
- Meehan, M., Insko, B., Whitton, M., Brooks, F. P. (2002). Physiological measures of presence in stressful virtual environments. *ACM Transactions on Graphics, Proceedings of ACM SIGGRAPH 2002*, *21*(3), 645–653.
- Schloerb, D. W. (1995). A quantitative measure of telepresence. *Presence: Teleoperators and Virtual Environments*, *4*(1), 64–80.
- Sheridan, T. B. (1992). Musings on telepresence and virtual presence. *Presence: Teleoperators and Virtual Environments*, *1*(1), 120–126.
- Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence: Teleoperators and Virtual Environments*, *10*(3) 282–297.
- Slater, M., Usoh, M., Steed, A. (1995). Taking steps: The influence of a walking metaphor on presence in virtual reality. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *2*(3) S, 201–219.
- Slater, M., & Steed, A. J. (2000). A virtual presence counter. *Presence: Teleoperators and Virtual Environments*, *9*(5), 413–434.
- Slater, M., Steed, A. J., & Chrysanthou, Y. (2001). *Computer graphics and virtual environments: From realism to real-time*. Reading, MA: Addison-Wesley.
- Slater, M., & Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, *6*(6), 603–616.
- Usoh, M., Arthur, K., Whitton, M., Bastos, R., Steed, A., Slater, M., et al. (1999). Walking > Walking-in-place > Flying, in virtual environments. *Proceedings of SIGGRAPH '99. Computer Graphics Proceedings, Annual Conference Series, ACM SIGGRAPH*, 359–364.
- Usoh, M., Catena, E., Arman, S., Slater, M. (2000) Presence questionnaires in reality. *Presence: Teleoperators and Virtual Environments*, *9*(5), 497–503.
- Zahoric, P., & Jenison, R. L. (1998). Presence and being-in-the-world. *Presence: Teleoperators and Virtual Environments*, *7*(1), 78–89.