Pictures, mock ups and animations: On the ecological validity of environmental simulations

Rainer Guski
Faculty of Psychology
Ruhr - University Bochum
Germany
1. The notion of Ecological Validity

Real scientific experiments should be completely under control and be replicable - a prerequisite that cannot be fulfilled in real environments. Environmental psychologists try to simulate and manipulate those aspects in the laboratory that are essential for their question at hand. Under which circumstances do we produce valid results? Validity generally refers to the extent to which an information indicates what it is claimed to give.

In psychology, we distinguish different forms of validity: Content validity, Face val., Factorial val., Empirical val., Predictive val., Concurrent val., Test val., and Ecological validity. Ecological validity refers to the extent to which a stimulus/response-combination indicates what it is claimed for a species. Species-typicality means that a species (e.g., a human being) does not need much learning in order to perform the experiment.

The species-typicality must be defined along 3 dimensions:

1) stimulus objects presented,

2) response measured,

3) experimental context

With 3 dimensions and all possible combinations, 8 types of experiments are possible, differing in the grade and type of ecological or non-ecological approach (Miller 1985). Ecological validity per se does not mean that the results of the lab experiment must be generalizable to real world settings, but if we want to transfer lab results to the real world we must show that the results are generalizable, i.e., independent of the specific experimental stimuli, responses and context. It seems necessary to compare the results of a systematic variation of experimental stimuli, responses and contexts; one single experiment will never be enough.
2. What are species-typical stimuli?

Typical visual stimuli used in psychological experiments are abstract and meagre: Lines, rectangles, circles, arrows, points etc. on uniform backgrounds. They provide little information about context (size, distance, affordance). Egon Brunswik (1944) wrote: Single stimulus dimensions (e.g., visual angle of an object) provide no valid information about object size; we should use multiple dimensions (e.g., visual angle of an object + visual angle of egocentric distance). James Gibson (1966) stressed: visual stimuli should be displayed on structured background in order to provide size and distance information.

Many so-called “visual illusions” rely on the fact that humans are not able to judge perceptual dimensions without a surrounding context (e.g., colour, size, speed).

One example: Depending on the visual context, the black figure looks taller or smaller than the red figure (Fig. 1)

In the absence of context information the human perceptual system uses “default assumptions” (e.g., figures are located at the same distance). Sometimes, abstract visual displays provide sufficient information for certain tasks, e.g., a horizontal wedge facilitates movements toward the direction of its tip, or point-light displays provide sufficient information about the gender of a walking person.

Real-world stimuli are usually well-known and contain semantic and emotional information. Semantic and emotional information both depend upon the cultural context, but the cultural context changes over time. Emotional and semantic information should be controlled in experiments, using pretests or statistics (e.g., covariance analysis).

Perceptual recognition depends on several factors, e.g., the uniqueness of an object; the size of the display; static vs. dynamic stimuli; the direction and field of view. Sometimes, comparisons between lab and field studies show high covariation, but still photos are at the lower end of the covariation continuum.
There is little psychological research using mock ups (design models). Pyron (1971) used plastic models of house modules and made black and white endoscopic travelling shots through the settings. The camera angle of view was 50 degrees; eye movements were recorded while subjects viewed the videos. At the end of the session, subjects were to recall the location of 4 specific houses. (Fig.2)

Results: The foveal coverage increased with increasing degree of syntactical information. The mean recognition errors were about 20%, and independent of the spatial arrangements used. The paper provides no information about reliability and validity.

Now let us turn to essential stimulus dimensions of computer-based studies. A number of dimensions have been identified: Screen size (regular (small) vs. large); screen type (translucent / opaque); motion (still pictures / animations / real movies; (sounds no sounds / .. / sounds fit to image) (cf. Hetherington et al. 1993; Becher et al. 1997; Viollon et al. 1999, Rohrmann & Bishop 2002)); degree of interaction (no choice of views / .. / subjects can choose paths, velocity etc. (cf. Bishop et al. 2001)) and the degree of immersion (presence) resp. the „elimination of mediation“ (Riva et al. 2003).

Turning to interactive large-screen environments, I’d like to mention two of our own studies:
Generally, we use a 2.5 X 2.5 m translucent screen, with subjects sitting 1 m in front. Some studies ask for subjective security in environmental settings. Garstka (2004) performed a lab experiment with interactive 360° views of underground stations and university campus sites, he varied some stimulus variables, e.g. The brightness of the illumination, and perceivable opportunities of escape. Subjects scaled the degree of perceived subjective security. It turned out that the lightness of illumination was a stronger determinant of subjective security than opportunities of escape (Fig. 3).
The second experiment was performed by Blöbaum & Hunecke (2005) in the field (on the campus of the Ruhr-University) and asked 122 students to scale subjective security in real life settings under different degrees of illumination. It turned out that only 26% of the variance in subjective security could be explained (Fig. 4), and perceived opportunities of escape was the strongest determinant of subjective security. This stands in sharp contrast to the lab study mentioned before. It is uncertain whether our lab studies produce valid results, or whether our field studies produce valid results.

![Diagram showing the relationship between subjective security and factors such as perceived escape opportunities, lighting, and overview.](image)

**Figure 4: Results of the study of Blöbaum & Hunecke (2005)**

Turning to immersive virtual environments, I like to mention the study by Conroy 2001: She used a helmet display with a large field of view (105 x 41°), and a 3D mouse for navigation. In one experiment, she compared real observations of pedestrians in the Tate Gallery with virtual observations of ‘pedestrians’ in a virtual Tate Gallery. It should be mentioned that there were no pictures in the virtual Tate. For certain behaviour variables, r-square of 0.5 was observed. This may be a trivial result: people do not run into walls! (Fig. 5)
Further experiments by Conroy (2001) used wayfinding tasks in five virtual worlds. The floor layout and exterior design of houses was systematically varied, and the subjects’ traces on their trips through the environments were registered. Although there was a certain covariation between floor layout and exterior of houses, she concluded that subjects usually move on linear paths, following long sight-lines, with pauses in configurationally ‘integrated’ locations offering strategic visual properties, long lines of sight, and large isovist areas. The validity question cannot be answered by this study.

There has been a number of desktop virtual environments, but I’d like to mention only one by Rohrmann & Bishop (2002). The authors simulated a suburban environment and varied the quality of illumination (day/sun, day/fog, night), personal shadow (yes/no) and sound (on/off). They asked 147 subjects about perceived simulation quality, comprehension, recollection and appreciation of the simulated environment. Here are the main results: (1) the simulations were perceived as valid and acceptable; (2) the appraisals differ according to lighting and time-of-day conditions, and (3) the provision of sound enhances the perceived quality of presentations.

3. What are species-typical responses?

As mentioned in the beginning, ecological validity refers to the extent to which a stimulus/response-combination indicates what it is claimed for a species. Species-typicality means that a species (e.g., a human being) does not need much learning in order to perform the experiment, i.e., species-typical responses should not require much learning. Automatic (reflex) responses require the least amount of learning, but are sometimes of little scientific value. The choice of responses mostly involves a compromise between ease of performance and scientific usefulness.

Classical psychological experiments mainly use two response classes: (1) Overt behaviour, like verbal reports (free, bound to questionnaire items, etc.), scaling (intensity, frequency, evaluation etc.), and body movements (eye movements, navigation, button pressing etc.), (2) Covert behaviour, like reaction or decision time, physiological responses (encephalographic, electrodermal etc.).
Before starting an ecological experiment, we should first observe the typical behaviour in real-world settings, then choose forms of behaviour that are both close to typical behaviour, and recordable and analysable in an objective manner. It should be noted that even free verbal behavior can be analysed in an objective manner, e.g., by using content analysis methods.

4. The context of stimulus-response combinations

Previous psychological experiments have shown that the experimental context determines responses to a great extent. For instance, loudness judgments of speech depend on

(a) the range of objective sound levels used (Parducci 1974),

(b) the frequency of objective stimulus events used (Parducci 1954),

(c) the order of stimulus events used, and

(d) the manner of speech (whispering, normal, shouting, Gardner 1969).

It has been shown that subjects develop response strategies in the course of experiments, i.e., they learn to compare different stimulus elements of the experiment, and how to cope with them. This has been considered as a problem, and two problem solutions have been proposed: (1) single shot studies using many subjects only one time, (2) repeated measurements of few subjects in long sessions. Both solutions have their pros and cons.

Another contextual problem relates to differences between verbal reports about behaviour and observed actual behaviour: When asked about “typical behaviour” in past situations, people often report what they want to do (e.g., sleep with windows open, use public transportation often), but the observed actual behaviour may be different (e.g., people sleep with windows closed, mainly use private cars). At present, we don’t have any solution for this problem.
References


References (2)
