Reviewer's report

Title: A systematic review of the evidence for direct health benefits of natural environments

Version: 1 Date: 23 April 2010

Reviewer: Terry Hartig

Reviewer's report:

This kind of research is needed to help consolidate the literature on nature and health. I appreciate the care taken by the authors in pursuing it. I do see some significant issues, but the authors should be able to address these in revisions.

- Major Compulsory Revisions

1. The authors examine data from pre- and post-activity measures, and they exclude data from measures obtained during activities in different environments. I can appreciate that data may not have been collected during the activity in comparison environments in most of the studies, but by ignoring those data that were collected from subjects while active in the environments, the authors may have neglected advances that are meaningful for the field as well as for their own results. Consider the case with the blood pressure findings. In the second study in our 1991 Environment and Behavior paper (Hartig, Mang, & Evans), we found that posttest blood pressure and heart rate differed little across comparison groups. We considered the possibility that differential effects of the environments had emerged during the environmental “treatment” but then dissipated by the time that we collected the posttest measures. Accordingly, we noted the value of ambulatory monitoring of physiology during the period spent in the comparison environments. In a subsequent field experiment, reported in the Journal of Environmental Psychology in 2003 (Hartig, Evans, Jamner, Davis, & Gärling), we followed up on the earlier observation and implemented an ambulatory monitoring protocol. We found that, as with the earlier study, there were no significant posttest differences in blood pressure, and the posttest means apparently differed little from the means obtained prior to the environmental treatment; however, the measures obtained during the treatment period enabled us to see that systolic and diastolic blood pressure did vary as a function of environment. In the present paper, it appears that the need for a common denominator in experimental design led the authors to ignore this development, and in doing so they unfortunately give a picture of the state of empirical and methodological knowledge that is so incomplete as to be inaccurate. The authors would do well to address this point.

2. The authors distinguish among different measures of emotion, but not among the different behavioral measures intended to tap the capacity to direct attention. The choice of measures over the years has reflected efforts to zero in on the cognitive inhibitory mechanism on which directed attention is thought to
depend. Since the initial use of proofreading in the studies reported in the Hartig, Mang, and Evans (1991) paper, authors have come to use measures that are linked to more specific conceptions of attention. Work by Laumann and colleagues (2003, in the Journal of Environmental Psychology), Berto (2005, in the Journal of Environmental Psychology) and Berman and colleagues (2008, in Psychological Science) exemplifies this positive development. At the same time, researchers have commented on the insensitivity of some instruments. For example, in each of two experiments reported in our 1996 paper in the Scandinavian Journal of Psychology (Hartig, Böök, Garvill, Olsson, and Gärling), we found no effects of photographic simulations of forest and city environments with a memory and search task (SMT). Recognizing that the environmental simulations may have been too weak and/or brief to have a measurable effect on directed attention capacity, we used the SMT again in our 2003 field experiment (Hartig, Evans, Jamner, Davis, & Gärling), which had a longer and presumably more powerful environmental treatment than the earlier study. We again found no reliable effect of environment with the SMT. We did however find an environment x time interaction with another measure of attention, the Necker Cube Pattern Control task, with which the inhibition of competing stimuli was more explicitly addressed. The point here is like the foregoing point about blood pressure findings. The authors could emphasize in their discussion that this is still a fairly young field of inquiry, and findings of no association may reflect on a lack of knowledge of what to measure, with what instruments, and when, as well as on a lack of potency of the experimental conditions. As it is, they only hint at this in their discussion.

3. In their Methods section (p. 6), the authors write as follows:

“Synthetic environments included non-green outdoor built environments or indoor environments (extended to include environments in which there may be a view/picture of a natural environment). ‘Exposure’ could comprise physical presence within the environment (i.e., some form a [sic] passive/sedentary activity) or the use of the environment as a setting for a form of physical activity.”

The authors’ characterization of the synthetic environment gives the impression that they have included studies that compare the natural environment (as characterized earlier in the text) with synthetic environments that sometimes included representations of nature. This is perhaps just a minor point in need of clarification, but to the extent that the results do integrate comparisons of outdoor natural environments with built environments that have distinctive natural features and representations of nature, the authors should arguably treat those as a separate category, in line with the discussions of such environments in the literature (e.g., concerning indoor plants).

- Minor Essential Revisions

1. The authors state on p. 5 that, “In a similar theory, Ulrich has proposed that nature may allow psychophysiological stress recovery by innate, adaptive responses to attributes of natural environments.” The theory is similar in some respects but quite different in others. I can accept that this is not the place to go
into the details of the theories, but the authors can at least give enough information about the restorative processes described in the two theories that readers can appreciate the differences between them.

2. The authors state on p. 5 that “the evidence for the existence of general and direct benefits of nature on health has not been rigorously assessed.” It may be true that there is no other quantitative synthesis in English like that of the authors, but careful assessment work has been done. I refer in particular to the report commissioned by the Health Council of the Netherlands, readily available in English on the internet. Here is the citation: Health Council of the Netherlands (2004). Nature and health. The influence of nature on social, psychological and physical well-being (Publication no. 2004/09). The Hague: Health Council of the Netherlands and Dutch Advisory Council for Research on Spatial Planning, Environment and Nature.

3. Together with the allowance for views/pictures, the allowance for passive activity in their characterization of the exposure had me thinking that the authors would include in their review studies in which subjects looked at videos or slides of natural versus urban/built environments. As I went further in the text, however, I saw that several such studies were not included in the review (e.g., Ulrich, 1981, in Environment and Behavior; Ulrich et al., 1991, in Journal of Environmental Psychology; Hartig et al., 1996, Studies 1 and 2, in Scandinavian Journal of Psychology; Hartig et al., 1999, Studies 1 and 2, in Journal of Environmental Psychology; Laumann et al., 2003, in Journal of Environmental Psychology; Berman et al., 2008, Study 2, in Psychological Science). The authors could in their methods section clarify that they did not include such studies (they do so now at the top of p. 21 – see comment below).

4. On p. 7, line 21, the authors write that four studies were not eligible (i.e., those reported in articles 24-27). Those four studies appear, however, in the description of the studies included in the review given on p. 10 and in Table 1. It seems some clarification is needed.

5. On p. 18, the authors note the potential for self-selection bias with samples of volunteers (an issue especially with the quasi-experiments), and they note the possibility that effects may differ for different subsets of the population. They could at this point add that effects may be systematically underestimated in the true experimental studies, in that they would not have captured aspects of leisure experiences relevant to the respective comparison conditions (e.g., freedom of choice regarding time, place, activity, and solitude versus company; intrinsic motivation to engage in the activity).

6. The authors explain on p. 21 that they “did not include in our review studies which compared the effect of pictures or more synthetic environments, however, such studies have been conducted.” They cite Staats and Hartig (2004) on this point. That study did use pictures of different environments, but it did so to solicit environmental evaluations, not to measure effects of the different environments on specific emotional states, attention, or physiology. The authors can more accurately cite here the studies mentioned in point 3 above. Alternatively, they
can explicitly treat liking/preference for an environment as an emotional response like those in the studies that they have reviewed. This would however mean that they had excluded a very large number of relevant studies from their review. Stamps has however already done some meta-analytic work in this regard (in Journal of Planning Literature, 1999).

7. On pp. 21-22, the authors separate the restoration and physical activity explanations for health benefits of nature experiences. This separation is problematic; see the Lancet, 372, 1614-1615, 2008 [an invited comment on the study by Mitchell and Popham]. Alternatively, see the chapter on restorative environments in the book edited by Ward Thompson and Travlou, Open space: People space (pp. 163-179; Taylor & Francis).

8. At the bottom of p. 22, the authors describe the generalisability of study findings as “questionable.” This is of course true, just as it is true in some sense of most if not all findings from social and behavioral research. I am concerned however that the word “questionable” carries negative connotations (e.g., doubtful, unlikely) that diminish the value of the work reviewed. I think that it is both more constructive and more accurate to frame the issue of generalisability in empirical terms; that is, it remains for future research to show that comparable effects can be found with samples from different populations of people, environments, and circumstances.

9. The authors should I suppose check the details given in Table 1. I note that they have given the wrong sample size for Hartig et al., 1991, Study 2; it should be 102, not 105.

- Discretionary Revisions

1. I realize that the authors have to conclude their search for relevant articles at some point, but it appears that they have missed some relevant studies. Three that come to mind are the 1995 study by Tennessen and Cimprich in Journal of Environmental Psychology, Study 1 in the 2008 paper by Berman et al. in Psychological Science, and Study 3 in the 1999 Hartig et al. paper in Journal of Environmental Psychology. Results from these studies may have a bearing on the emotional and attentional conclusions.

2. The authors excluded some studies from their quantitative synthesis for some outcomes because they used a measure used too seldom in other studies. As I indicated above, the authors treat different behavioral measures of attention as equivalent. Although problematic, this has allowed for the inclusion of more studies in the quantitative synthesis. In contrast, the differentiated treatment of measures of emotion means that some effect estimates have been excluded. The authors could refer to the extensive literature on the circumplex structure of affect to justify a simpler, complementary approach to the integration of results from different measures of emotion. That is, they could group emotion measures with regard to their location on each of two dimensions, activation and pleasantness. For example, the studies that used energy/vigour, fatigue/tiredness, or both could be combined for one assessment relevant to
activation. I am not saying that this is preferable to integration of results for specific emotions (for a discussion of the dimensional versus specific affects approaches to measurement, see the 1997 chapter by Stone in the book Measuring stress: A guide for health and social scientists), but I am saying that it could enable the authors to include results obtained with measures that have been used less frequently.

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I declare that I have no competing interests.