Journal of Toxicology and Environmental Health, Part B
Critical Reviews

ISSN: 1093-7404 (Print) 1521-6950 (Online) Journal homepage: http://www.tandfonline.com/loi/uteb20

Attention Restoration Theory: A systematic review of the attention restoration potential of exposure to natural environments

Heather Ohly, Mathew P. White, Benedict W. Wheeler, Alison Bethel, Obioha C. Ukoumunne, Vasilis Nikolaou & Ruth Garside

To cite this article: Heather Ohly, Mathew P. White, Benedict W. Wheeler, Alison Bethel, Obioha C. Ukoumunne, Vasilis Nikolaou & Ruth Garside (2016) Attention Restoration Theory: A systematic review of the attention restoration potential of exposure to natural environments, Journal of Toxicology and Environmental Health, Part B, 19:7, 305-343, DOI: 10.1080/10937404.2016.1196155

To link to this article: https://doi.org/10.1080/10937404.2016.1196155

Published online: 26 Sep 2016.

JOURNAL OF TOXICOLOGY AND ENVIRONMENTAL HEALTH, PART B
2016, VOL. 19, NO. 7, 305–343
http://dx.doi.org/10.1080/10937404.2016.1196155

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS

Attention Restoration Theory: A systematic review of the attention restoration potential of exposure to natural environments

Heather Ohly[a], Mathew P. White[a], Benedict W. Wheeler[a], Alison Bethel[b], Obioha C. Ukoumunne[b], Vasilis Nikolaou[b], and Ruth Garside[a]

[a]European Centre for Environment and Human Health, University of Exeter Medical School, Truro Campus, and Knowledge Spa, Royal Cornwall Hospital, Truro, Cornwall, United Kingdom; [b]NIHR CLAHRC South West Peninsula, University of Exeter Medical School, South Cloisters, St Luke's Campus, Exeter, Devon, United Kingdom

ABSTRACT
Attention Restoration Theory (ART) suggests the ability to concentrate may be restored by exposure to natural environments. Although widely cited, it is unclear as to the quantity of empirical evidence that supports this. A systematic review regarding the impact of exposure to natural environments on attention was conducted. Seven electronic databases were searched. Studies were included if (1) they were natural experiments, randomized investigations, or recorded "before and after" measurements; (2) compared natural and nonnatural/other settings; and (3) used objective measures of attention. Screening of articles for inclusion, data extraction, and quality appraisal were performed by one reviewer and checked by another. Where possible, random effects meta-analysis was used to pool effect sizes. Thirty-one studies were included. Meta-analyses provided some support for ART, with significant positive effects of exposure to natural environments for three measures (Digit Span Forward, Digit Span Backward, and Trail Making Test B). The remaining 10 meta-analyses did not show marked beneficial effects. Meta-analysis was limited by small numbers of investigations, small samples, heterogeneity in reporting of study quality indicators, and heterogeneity of outcomes. This review highlights the diversity of evidence around ART in terms of populations, study design, and outcomes. There is uncertainty regarding which aspects of attention may be affected by exposure to natural environments.

There is increasing practice and policy interest in the potential for natural environments to provide positive human health and well-being benefits. Attention Restoration Theory (ART) is commonly referenced to explain how this benefit might accrue; however, it is unclear how strong the empirical evidence is that exists for this proposed mechanism. In cognitive psychology, the ability to focus on a task that requires effort is known as directed or voluntary attention (Kaplan and Kaplan 1989). This ability is finite and may become fatigued. Attention fatigue may occur when there is a need to focus on a specific stimulus or task with little or no intrinsically motivational draw, while suppressing distractions that may be inherently more interesting, with an example being filling in a tax return while your children are playing in the yard (Kaplan 1995; Kaplan and Berman 2010). Attentional fatigue is important, not least because it is associated with poorer decision making and lower levels of self-control, which in turn have been linked to a variety of health-related issues such as obesity via increasingly understood neural and behavioral pathways (Fan and Jin 2013; Hare, Camerer, and Rangel 2009; Vohs et al. 2008).

More than half the world's population lives in urban areas. From a psychological perspective, urban lifestyles impose increasing demands on our cognitive resources (Kaplan and Berman 2010). According to ART these enhanced demands on directed attention may be linked to attention fatigue (Kaplan 1995; Kaplan and Kaplan 1989). The antidote, the theory claims, is to take time out from attention-demanding tasks associated with modern life, and spend time in natural

CONTACT Ruth Garside R.Garside@exeter.ac.uk European Centre for Environment and Human Health, University of Exeter Medical School, Truro Campus, Knowledge Spa, Royal Cornwall Hospital, Truro, Cornwall, TR1 3HD, United Kingdom.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/UTEB.

environments that demand less of our cognitive resources and enable us to recover our attentional capacities.

ART proposes that individuals benefit from the chance to (1) "be away" from everyday stresses, (2) experience expansive spaces and contexts ("extent"), (3) engage in activities that are "compatible" with our intrinsic motivations, and (4) critically experience stimuli that are "softly fascinating" (Kaplan 1995). This combination of factors encourages "involuntary" or "indirect attention" and enables our "voluntary" or "directed" attention capacities to recover and restore (Kaplan 1995; Staats 2012). Relaxing settings (such as places of worship) and activities (such as sleep) may provide restorative opportunities, but ART argues that nature may be particularly useful because it has an "aesthetic advantage" (Herzog et al. 2010; Kaplan and Berman 2010; Kaplan and Kaplan 1989). It is suggested that sending time in the natural world allows individuals the opportunity for "reflection" and consideration of unresolved issues (Herzog et al. 1997; Kaplan and Berman 2010).

The original development of ART was largely descriptive, based on observations of human–nature interactions and analysis of qualitative data (Kaplan and Frey Talbot 1983). Kaplan (1995) subsequently linked ART more broadly to attention theory, for instance, associating directed attention fatigue to problems in selection and problem solving, inhibition of competing stimuli, and feelings of irritability. More general psychological research provides behavioral and neural evidence of a distinction between "top-down" directed attention and "bottom-up" involuntary attention (Fan et al. 2005). Specifically, it has been postulated that directed attention is more linked to higher order mental functions because of a greater load on working memory produced by, among other things, the need to suppress distracting stimuli or alternative attentional cues. The simpler attentional processes of alerting (becoming aware of something; Jonides et al. 2008; Fan et al. 2002) and orienting (taking actions to focus on a stimulus) should be less affected by the trials of modern living because they demand relatively few cognitive resources. These processes are thus less likely to need recovery in the same way as the executive functions of attention, such as working memory, which not only needs to hold and replay visual and auditory stimuli, but may also have to manipulate them according to rules stored in short-term memory (Jonides et al. 2008).

This discussion is important because it suggests that only relatively demanding attentional tasks should show improvements following exposure to nature. An example is the Backwards Digit Span test, which requires participants to both remember and manipulate (reverse) a series of numbers. In contrast, tasks that require relatively few cognitive resources should be less demanding and so less affected by exposure to nature. An example would be the Sustained Attention to Response Task (SART), which simply requires participants to react to the presentation of digits from 1 to 9 on a computer screen, for example, by pressing the "space bar," except when the number 3 appears. The task thus involves "inhibiting" the most frequent response (pressing a space bar) when a rare event occurs and reflects the need to "sustain attention" rather than drift into overgeneralizing a behavioral response.

Taylor, Kuo, and Sullivan (2002) explored whether children living in apartments with greater surrounding green space showed higher levels of "self-discipline," potentially due to fewer demands on attention resources. Echoing this, Kaplan and Berman (2010) and Baumeister, Vohs, and Tice (2007) proposed that directed attention fatigue may be linked to a loss in self-regulation, such as the ability to resist temptation, suggesting these two processes may share a common resource. If true, relative depletion on tasks measuring self-regulation and the ability to inhibit actions, rather than just directed attention, might also be seen following exposure to urban versus natural environments. Kaplan and Berman (2010) presented a number of studies to support these claims. However, this review did not aim to systematically collect all relevant papers in the field and thus may have missed those that came to a different conclusion.

In contrast, a recent systematic review of the evidence for general health-related benefits of exposure to nature used a more systematic search strategy (Bowler et al. 2010). Consequently, it included a broader evidence base and also included a formal appraisal of study quality in an attempt to weigh the relative importance of different investigations. From the eight studies using cognitive measures of assessment, a meta-analysis of five, focusing on postexposure measures, supported ART. However, the three better designed studies, measuring cognitive performance both before and after exposure, demonstrated no marked evidence of improvements following exposure to nature.

A number of features of the Bowler et al. (2010) systematic review are important. First, the investigations included for the cognitive aspect of the review are different from those reviewed by Kaplan and Berman (2010), suggesting the need for a systematic review of all appropriate studies. Second, investigations using subjective measures of directed attention, such as parental reports of children's ability to concentrate, were included. A more robust test of the theory would focus on those studies using objective measures. Third, and partly because of the limited number of investigations reviewed in both papers, there was little opportunity to examine whether some attention measures were more sensitive to exposure to natural environments than others. All of the measures may be considered as trying to measure the same underlying construct: directed attention capacity. As with any measurement tool, each is subject to measurement error and might be more or less effective. Moreover, since each measure may be tapping into a slightly different aspect of directed attention capacity (such as alerting, orientating, or executive functions), considering data on the effects of nature on different types of measure may shed light on the precise mechanisms by which nature may restore attentional processes (Jonides et al. 2008).

The aim of the current systematic review was to identify, select, appraise, and synthesize the evidence for ART among studies that used experimental and quasi-experimental approaches and objective measures of attention. A larger sample of studies was included (n = 31) than was the case in either of the previous reviews. The investigations were also critically appraised and meta-analyzed where possible. Such meta-analysis is consistent with Gifford's (2014) call for better evidence synthesis in environmental psychology.

Methods

This systematic review followed the general principles published by the Centre for Reviews and Dissemination (2009). The predefined protocol is available on PROSPERO (Reference CRD42013005008).

Review Question

What is the relative attention restoration potential of natural settings compared to other settings?

Literature Search

A search strategy was devised by the research team, led by our Information Specialist (Alison Bethel), and captured concepts of attention restoration, cognitive function, and natural versus other settings. No suitable MeSH (Medical Subject Headings) terms were identified. No methods filters were used. The master search strategy (Table 1) was adapted and run in the following electronic databases in July 2013: PsycINFO, MEDLINE, and EMBASE (using OVID); AMED, SPORT Discus, and Environment Complete (using EBSCOHost); and Web of Knowledge (on Thomson). Reference lists of included studies were scrutinized for relevant investigations. Forward citation searches were undertaken on included studies. All searches were conducted from 1989, when seminal investigations on ART were published. Citation searches were also performed in Web of Science using these key references: Kaplan (1995) and Kaplan and Kaplan (1989).

Inclusion Criteria

Studies were eligible for inclusion in the review if they met the following criteria:

Population: Any.

Intervention and comparators: Studies reporting a comparison of the effects of exposure to natural settings and other, nonnatural settings. The definition of "natural" included real settings (such as parks, forests, wilderness areas) and virtual settings (images or videos of similar settings). The definition of "nonnatural settings" included real settings (such as city centers, residential areas, parking lots) and virtual ones (images or videos of similar settings). Types of engagement with these settings included active (such as walking or running) and passive (such as looking at the view from a window).

Table 1. Master search strategy as used in OVID Medline.

1	attention restorat*.tw.
2	(theory or hypothesis).tw.
3	(attention restorat* adj1 (theory or hypothesis)).tw.
4	natur*.tw.
5	outdoor*.tw.
6	green*.tw.
7	forest*.tw.
8	condition*.tw.
9	setting*.tw.
10	4 or 5 or 6 or 7 or 8 or 9
11	environ*.tw.
12	(environ* adj2 (natur* or outdoor* or green* or forest* or condition* or setting*)).tw.
13	4 or 5 or 6 or 7
14	(setting* adj2 (natur* or outdoor* or green* or forest*)).tw.
15	12 or 14
16	restorat*.tw.
17	15 and 16
18	attent*.tw.
19	cognitive function*.tw.
20	concentrat*.tw.
21	18 or 19 or 20
22	((attent* or cognitive function* or concentrat*) adj3 ((environ* adj2 (natur* or outdoor* or green* or forest* or condition* or setting*)) or (setting* adj2 (natur* or outdoor* or green* or forest*)))).tw.
23	17 or 22
24	urban setting*.tw.
25	everyday setting*.tw.
26	garden*.tw.
27	mental fatigue*.tw.
28	4 or 5 or 6 or 7 or 24 or 25 or 26 or 27
29	16 or 18 or 19 or 20
30	((restorat* or attent* or cognitive function* or concentrat*) adj3 (natur* or outdoor* or green* or forest* or urban setting* or everyday setting* or garden* or mental fatigue*)).tw. (4872)
31	23 or 30
32	3 or 31
33	limit 32 to yr = "1989 -Current"

Outcomes: Objective measures of attention capacity, for example the Digit Span Forward or Backward.

Study design: All experimental designs including randomized controlled trials, quasi-experimental studies (nonrandomized controlled trials; randomized or nonrandomized crossover trials) and natural experiments. With the exception of natural experiments, nonrandomized studies were included if they recorded measures of attention before and after exposure to nature/non nature settings. Investigations were excluded if baseline measures were taken after exposure had commenced because it was necessary to establish baseline attentional abilities.

Other: Conference proceedings or dissertations were included if there were sufficient data to assess the risk of bias. No language restrictions were applied.

Study Selection

All references identified through the search strategy were uploaded into ENDNOTE (X7, Thomson Reuters) and duplicates were removed. Reference titles and abstracts, where available, were independently double screened against the inclusion criteria (by Heather Ohly and Ruth Garside/Alison Bethel). Studies appearing to meet these were retrieved in full text. One article was published in Chinese and this was professionally translated. Full text screening was completed independently by two reviewers (Heather Ohly and Ruth Garside) using the same criteria. Discrepancies were resolved through discussion between reviewers.

Data Extraction

A standardized, piloted data extraction sheet was developed in Excel to ensure consistency between studies and reviewers. Data extracted for each study included study design, sample characteristics, setting characteristics (natural and nonnatural), type of exposure and engagement, duration of exposure, measures of attention, and duration of follow-up. Authors were contacted to clarify or supply missing data where necessary. Data were independently extracted by one reviewer (Heather Ohly) and checked by a second (Ruth Garside/Ben Wheeler/Mathew White). Disagreements were resolved by discussion, including the full team where necessary.

Quality Appraisal

The overall quality of the included studies was assessed using a combination of resources and guidelines: quality indicators from the Centre for Reviews and Dissemination (2009); critical appraisal checklists from the Critical Appraisal Skills Programme (2013); and quality assessment tool for quantitative studies from the Effective Public Health Practice Project (2013). These tools aim to assist reviewers in identifying potential sources of bias and make a considered judgment how robust the evidence may be.

A bespoke quality appraisal rating system was developed using 20 standard indicators of robust study conduct considered relevant in a review context. The rationale and method of applying some of these indicators is provided in Table 2. Each quality indicator used the following ratings: yes = 2; partial = 1; no = 0; unclear = 0. To accommodate the fact that not all questions were applicable for all study designs, for example, questions about appropriate randomization, all studies were rated overall using a percentage score based only on "applicable" criteria. Overall quality assessment was given as low (0–33%), moderate (34–66%), or high (67–100%).

Given changing standards relating to methods reporting and limited word counts for many journals, first authors of included papers were contacted where an indicator had initially been scored "unclear" (n = 24). They were asked to provide more information regarding the study, and these "unclear" indicators in particular, to aid a more informed assessment of the papers. Of the 24 requests, 9 responses with authors were received providing further details of study design. Whether or not authors responded to our request is recorded in Table 4 (shown later).

Data Synthesis

Random effects meta-analysis models were fitted using standard methods, in which inverse variance is used to weight individual study results, to pool the effect estimates across investigations. The meta-analyses compared attention outcomes at follow-up between groups exposed to natural settings (intervention) and groups exposed to non-natural settings (control) (Sutton et al. 2000). Summary data, the mean and standard deviation of the outcome, and sample size in each group for each study were used in the meta-analysis, with pooled results reported as mean differences (intervention minus control). Data were pooled from investigations that used the same measures of attention and use the same outcome (some attention measures have multiple associated outcomes). Care was taken not to double count participants, for example, in crossover trials when participants completed the same walk twice, alone and with a friend (Johansson, Hartig, and Staats 2011).

One article included three separate studies each with independent samples (Berto 2005). The first investigation compared two groups, one viewing natural images and the other viewing urban images. The second study had only one group, viewing geometric images, which were compared with data from the first investigation. As the urban group from the first study and the geometric group from the second were independent, their results were combined before being pooled with the third investigation in the meta-analysis.

None of the studies that measured outcomes at baseline reported using the correct method to adjust for baseline imbalance, analysis of covariance (ANCOVA) (Vickers and Altman 2001). Consequently, all meta-analyses were carried out using data at follow-up only. The main meta-analyses included all studies regardless of the level of baseline imbalance. Investigations for which the mean difference between the groups in outcome score at baseline was greater than one-tenth of the standard deviation in the control arm (i.e., "effect size" of 0.1) were considered to have imbalance at baseline. Cohen (1992), in his summary of effect sizes in behavioral science studies, used a threshold of 0.2 to define a small effect size. Sensitivity analyses were also conducted in which only those investigations that had low levels of baseline imbalance were included, to illustrate the potential impact of imbalanced experiments on results of the meta-analyses.

Data analysis was carried out using Stata 13 and Review Manager (RevMan 5.2).

Where studies could not be meta-analyzed, as they contained outcomes not shared with other investigations, or where insufficient data were supplied, results are described narratively.

Results

Search Results

Searches identified 10,979 unique records. Twenty-four articles met the inclusion criteria (Figure 1). Some of these articles reported results from more than one investigation, so 31 separate studies were included. Given the complexity of the tables and figures in this section, references are presented using only the first author's name and the date, to simplify data presentation.

Table 2. Quality appraisal rating system for selected indicators.

Quality indicators	Rationale	Yes	Partial	No	Unclear
Random allocation to groups/condition order	Based on the scientific definition of randomization for the purpose of allocation to environment groups.	Random allocation to groups (or condition order in studies with within-subject designs).	NA	Nonrandom allocation to groups. In natural experiments, where participants had previously been "randomly" allocated to housing, this was not classed as random allocation to environment groups because it was not within the experiment's control.	Method of allocation not clearly stated.
Randomization procedure appropriate	Failure to properly randomize has been shown to be associated with increased levels of effect in controlled clinical trials (Jüni, Altman et al. 2001).	Appropriate method/tools of randomization used.	NA	Inappropriate methods/tools of randomization used.	Method/tools of randomization not clearly stated.
Groups balanced at baseline (attention scores)	Effect sizes were calculated to show whether groups were balanced or imbalanced at baseline in terms of attention scores (or in some cases study authors provided this info).	An effect size <0.1 implies that groups were balanced at baseline (i.e., similar attention scores). Therefore both groups have similar potential to improve their attention scores.	Multiple attention outcomes reported, of which some were balanced and others were imbalanced at baseline.	An effect size >0.1 implies that groups were imbalanced at baseline (i.e., dissimilar attention scores).	Data needed to calculate effect size (baseline means and/or standard deviations) not reported.
Participants blind to research question	If participants were aware of the purpose of the study or hypothesis, this may have been a source of bias in favor of the intervention group through attracting a self-selected sample of those for whom the question had resonance or influencing their experience of intervention and control exposures.	Participants were not aware of the research question, and/or may have been given vague information about the nature of the study.	NA	Participants were aware of the research question and specifically its focus on attention outcomes.	Awareness or blindness to research question not clearly stated.

(*Continued*)

Table 2. (Continued).

Quality indicators	Rationale	Yes	Partial	No	Unclear
Demonstrated need for attention restoration	Normal or high attention scores at baseline would limit the potential for an exposure to nature to improve attention (i.e., ceiling effect).	Study populations with diagnosed medical conditions that affect attention were considered to have demonstrated need for attention restoration, for example, children with ADHD. Some studies used a mental loading task to induce attention fatigue prior to exposure to natural/other settings. This was only considered to have fully demonstrated need for attention restoration if measures of attention were taken before and after the task (i.e., T1—loading—T2—exposure—T3), thereby showing the extent of attention depletion from baseline.	Studies that measured attention after the mental loading task but not before (i.e., loading—T1—exposure—T2) were assigned "partial" because, although this sequence does not show the effectiveness of the mental loading task, it does show the impact of the intervention.	Studies that measured attention before the mental loading task but not after (i.e., T1—loading—exposure—T2) were assigned "no" because this sequence does not show the effectiveness of the mental loading task and the impact of the intervention on depleted attention is unclear.	Study populations may or may not have depleted levels of attention (e.g., schizophrenia patients) and this was not clearly stated. Some study authors may have considered it plausible that subjects were depleted but this was not demonstrated.
Consistency of intervention (within and between groups)	Between-group consistency is important because some of the experimental studies involved participants completing the same activity in two or more different settings, i.e., "active control" rather than "placebo control." Any differences between groups (other than the setting) may induce performance bias and lead to false results. For studies that involved viewing images of natural and other settings, aspects assessed for consistency included number of images, duration of exposure, location and conditions.	Consistency of intervention clearly described. Taking natural vs. urban walking studies as example, within-group consistency might include the walking route, day of the week and time of day, so that traffic volume is similar for each participant completing the urban walk. Between-group consistency might include the distance, speed, number of walkers, talking and other conditions of the walks. Some studies used a guide to ensure that a defined route and pace were adhered to and that conversation/ distractions were kept to a minimum.	Some attempt to ensure consistency. For example, some natural vs. urban walking studies attempted to ensure compliance by sending participants out with maps or global positioning system (GPS); however, in some cases it was not clear whether compliance was actually checked.	Lack of consistency within and/or between groups. For example, some interventions were home-based and self-directed, where participants chose from a range of possible activities. In the natural experiment studies, participants may have lived in their current home for varying periods of time, and may have had different levels of exposure to nature outside the home.	Consistency of intervention not clearly described.

(*Continued*)

Table 2. (Continued).

Quality indicators	Rationale	Yes	Partial	No	Unclear
Outcome assessors blind to group allocation	It has been repeatedly shown in controlled trials that lack of assessor blinding may lead to biased assessment and inflated effect sizes (Schulz et al. 1995; Jüni, Altman et al. 2001). However, this kind of detection bias may be less problematic where objective outcome measures (such as attention tests) are used.	Assessor blindness clearly stated.	NA	Assessors were clearly not blind to group allocation, for example, if one researcher led the group activities and administered the attention tests.	Assessor blindness not clearly stated.
Consistency of data collection	This includes consistency within groups and between groups.	Consistency of data collection clearly described. Within-group consistency might include the pre–post assessment interval, particularly in studies where participants were recruited at different times across a wider study period and self-managed the intervention phase at home (such as cancer patients or pregnant women). Between-group consistency might include the location, conditions, and timing of data collection. All of these factors may influence the outcome of the attention tests and the extent to which any changes over time (and differences between groups) are observed.	Some attempt to ensure consistency. For example, there may have been some variation in the pre–post assessment interval, but the testing was done in the same room with all other factors consistent.	Lack of consistency within and/or between groups.	Consistency of data collection not clearly described. Lack of detail about data collection methods generally.
Statistical analysis methods appropriate for study design	In order to demonstrate the effect of nature on attention capacity, statistical comparison of the mean change in attention capacity between groups was expected (i.e., ANOVA or ANCOVA time × environment). For natural experiments, baseline measures were not collected and this type of analysis would not be possible; therefore comparison between groups after the exposure (e.g., *t*-test) was considered appropriate.	Appropriate statistical analysis methods used.	Some doubt as to the appropriateness of the statistical analysis methods used, such as correct choice of test but incorrect use of data (e.g., collapsing three time points).	Inappropriate statistical analysis methods used. However, we recognise that accepted norms of statistical analysis may have changed since the time of publication.	Statistical analysis methods not clearly stated.

(*Continued*)

Table 2. (Continued).

Quality indicators	Rationale	Yes	Partial	No	Unclear
Sample representative of target population	Studies were assessed for selection bias (or sampling bias) including self-selection, prescreening or recruitment from groups likely to show a better response. If study participants are not representative of the target population, this may introduce bias in favor of the intervention group.	Sample considered representative of target population; recruitment strategy included measures to reduce selection bias.	NA	Sample not considered representative of target population. For example, students enrolled on environmental courses such as forestry would be expected to show preference for natural environments, which may increase the likelihood of experiencing restoration effects in natural environments relative to other types of environments.	Insufficient information about the sample and how it was recruited to make a judgment about representativeness.

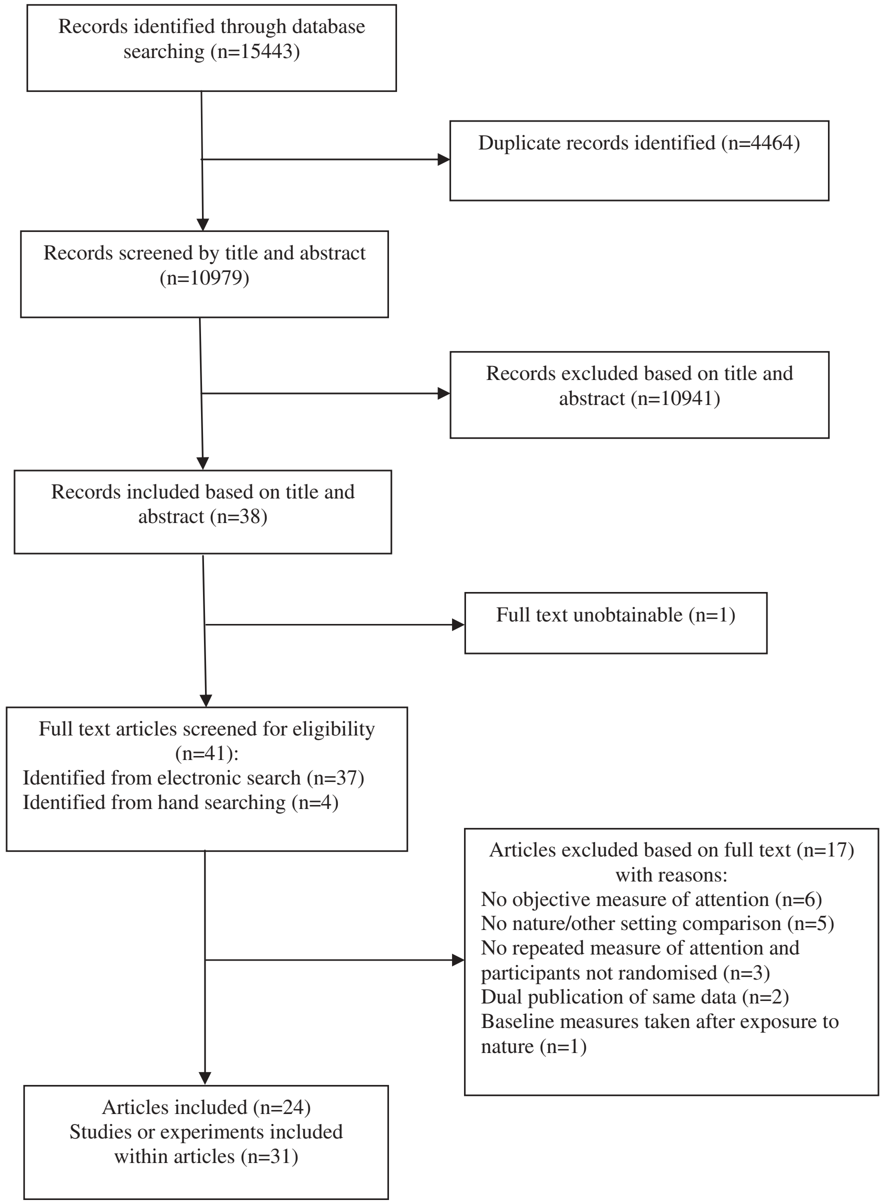

Figure 1. PRISMA flow chart.

Study Characteristics

The 31 included studies were from Europe, the United States, and Asia, and varied in terms of experimental design, sample characteristics, and study duration, as well as the type and extent of exposure to nature (Tables 3A to 3D).

Study designs included 16 randomized controlled trials (RCT) from 12 articles (Berto 2005; Chen, Lai, and Wu 2011; Cimprich and Ronis 2003; Hartig et al. 1996; 2003; Hartig, Mang, and Evans 1991; Laumann, Gärling, and Stormark 2003; Mayer et al. 2009; Perkins, Searight, and Ratwik 2011; Rich 2008; Stark 2003; van den Berg, Koole, and van der Wulp 2003); 7 randomized crossover trials from 6 articles (Berman et al. 2008; 2012; Bodin and Hartig 2003; Johansson, Hartig, and Staats 2011; Shin et al. 2011; Taylor and Kuo 2009); 3 natural experiments (Kuo 2001; Taylor, Kuo, and Sullivan 2002; Tennessen and Cimprich 1995); 3 nonrandomized controlled trials (Berto 2005; Hartig, Mang, and Evans 1991; Wu et al. 2008); and 2 nonrandomized crossover trials (Ottosson and Grahn 2005; van den Berg and van den Berg 2011).

Study populations included children, "students" and adults. Some samples were of individuals with

Table 3A. Characteristics of included studies—RCT design; actual exposures (or mixed actual/virtual).

Author, year ()[a]	Study design	Country and setting	*n*	Sample characteristics: gender, mean age, population, ethnicity, and s/e status if reported	Intervention characteristics: activity, setting (each group) and duration of exposure	Attention measures (objective)
Cimprich, 2003	RCT	United States University medical center	185	100% female 53.8 years Patients with newly diagnosed breast cancer (with surgery as primary treatment plan) 86% White	Home-based, patient-led program of nature activities Control group logged relaxation time 120 min per week (nature activities) Baseline—follow-up period approx. 36 days (pre- and postsurgery)	Digit Span Backward Digit Span Forward Necker Cube Pattern Control Trail Making Tests A Trail Making Test B
Hartig, 2003	RCT	United States University and local area	112	50% male 20.8 years Students	Sitting, natural view; then walking, natural (nature reserve) Sitting, no view; then walking, urban (city streets) 1 hour (10 min passive; 50 min active)	Necker Cube Pattern Control Search and Memory Test
Hartig, 1991 (2)	RCT	United States University and local area	102	50% male 20 years Students	Walking, natural (regional park) Walking, urban (city centre) Reading magazines, comfortable laboratory setting 40 min	Proofreading Task
Mayer, 2009 (1)	RCT	United States University	76	29% male Mean age not reported Students	Walking, natural (woods/creek) Walking, urban (downtown) 10 min	Memory Loaded Search Task (similar to Search and Memory Task)
Mayer, 2009 (2)	RCT	United States University	92	30% male Mean age not reported Students	Walking, natural (woods) Watching video, natural (woods) Watching video, urban (busy streets) 10 min	Memory Loaded Search Task (similar to Search and Memory Task)
Perkins, 2011	RCT	United States University	26	27% male Age range 19–24 years Mean age not reported Students	Walking, natural (woods) Walking, urban (residential/business) Walking, urban (parking lot) 20 min	Digit Span Backward Digit Span Forward Logical Memory
Stark, 2003	Cluster RCT	United States Prenatal classes	57	100% female 29.1 years Pregnant women in the third trimester 94.7% White	Outdoor "restorative" activities Alternative session on the discomfort of pregnancy 120 min per week (outdoor activities) Baseline—follow up period varied 13–64 days	Category Matching Digit Span Backward Digit Span Forward Errors Scale Trail Making Tests A Trail Making Test B

Note. Outcome measures listed in alphabetical order, not the order in which they were administered.
[a]Experiment number in parentheses; each experiment has distinct sample; s/e = socioeconomic.

Table 3B. Characteristics of included studies—RCT design; virtual exposures.

Author, year ()[a]	Study design	Country and setting	n	Sample characteristics: gender, mean age, population	Intervention characteristics: activity, setting (each group) and duration of exposure	Attention measures (objective)
Berto, 2005 (1)	RCT	Italy University	32	50% male 23 years Students	Viewing images, natural Viewing images, urban 25 images × 15 sec each	Sustained Attention to Response Test
Berto, 2005 (3)	RCT	Italy University	32	50% male 22 years Students	Viewing images, natural Viewing images, urban 25 images × duration of their choice	Sustained Attention to Response Test
Chen, 2011 (1)	RCT	China Senior secondary school	48	42% male Mean age not reported Students	Viewing images, natural Viewing images, city Viewing images, urban nightscape Viewing images, sports 10 images × 15 sec each	Colored number pictures
Hartig, 1996 (1)	RCT	Sweden University and high schools	102	38% male 21.4 years Students	Watching simulated walk, natural (trees) Watching simulated walk, urban (city) No simulated walk (control) 80 slides × 10 sec each (13.5 min)	Search and Memory Task
Hartig, 1996 (2)	RCT	Sweden University	18	50% male 27.4 years Students	Watching simulated walk, natural (trees) Watching simulated walk, urban (city) 80 slides manually (12 min)	Search and Memory Task
Laumann, 2003	RCT	Norway University	28	100% female Age range 18–24 years Mean age not reported Students	Watching video, natural (island waterside) Watching video, urban (city streets) 80 scenes × 15 sec each	Posner's Attention Orienting Task (note: no raw data and comparisons focus on different types of stimuli; therefore not reported in Table 5)
Rich, 2008 (1)	RCT	United States University	145	17% male Mean age not reported Students	Looking at view, natural (forest) Looking at view, urban (buildings) No view 1 min	Vigilance Task Stroop Colour-Word Test
Rich, 2008 (2)	RCT	United States University	36	42% male Age range 18–21 years Mean age not reported Students	Reading magazines, room with plants Reading magazines, room with other objects 10 min	Digit Span Backward
van den Berg, 2003	RCT	The Netherlands University	114	32% male (after exclusions for n = 106) 21.9 years Students	Watching simulated walk, natural (forest with or without water) Watching simulated walk, urban (city with or without water) 7 min	D2 mental concentration test

Note. Outcome measures listed in alphabetical order, not the order in which they were administered.
[a]Experiment number in parentheses; each experiment has distinct sample; s/e = socioeconomic

Table 3C. Characteristics of included studies—Other designs; actual exposures.

Author, year ()[a]	Study design	Country and setting	*n*	Sample characteristics: gender, mean age, population, ethnicity, and s/e status if reported	Intervention characteristics: activity, setting (each group), duration of exposure, and washout period (crossover only)	Attention measures (objective)
Berman, 2008 (1)	Randomized crossover trial	United States University	38	39% male 22.6 years Students	Walking, natural (park) Walking, urban (downtown) 50–55 min Two walks, 1 week apart	Digit Span Backward
Berman, 2008 (2)	Randomized crossover trial	United States University	12	33% male 24.3 years Students	Viewing images, natural (Nova Scotia) Viewing images, urban (downtown) 50 images in 10 min Two sessions, 1 week apart	Attention Network Test Digit Span Backward
Berman, 2012	Randomized crossover trial	United States University and local area	20	40% male 26 years Adults diagnosed with major depressive disorder (MDD)	Walking, natural (park) Walking, urban (downtown) 50–55 min Two walks, 1 week apart	Digit Span Backward
Bodin, 2003	Randomized crossover trial	Sweden Running club	12	50% male 39.7 years (males) 37.0 years (females) Runners	Running, natural (park) Running, urban (city streets) 60 min Two runs, 1 week apart	Combined Digit Span Backward and Forward Symbol Digit Modalities Test
Johansson, 2011	Randomized crossover trial	Sweden University	20	50% male 24.2 years (males) 22.4 years (females) Students	Walking, natural (park) Walking, urban (streets) 40 min Four walks, 1 week apart (natural with friend; urban with friend; natural alone; urban alone)	Symbol Substitution Test
Shin, 2011	Randomized crossover trial	South Korea University	60	58% male 23.3 years Students	Walking, natural (park) Walking, urban (city streets) 50–55 min Two walks, 1 week apart	Trail Making Test B
Taylor, 2009	Randomized crossover trial	United States	25	88% male (after exclusions for n = 17) 9.2 years Children diagnosed with ADHD	Walking, natural (urban park) Walking, urban (downtown) Walking, urban (neighborhood) 20 min Three walks, 1 week apart	Digit Span Backward Stroop Colour-Word Test Symbol Digit Modalities Test Vigilance Task (Note: Only DSB reported)
Hartig, 1991 (1)	Nonrandomized controlled trial	United States Trailheads and local clubs	68	62% male 35.9 years (G1) 29.2 years (G2) 31.6 years (G3) Experienced backpackers	Wilderness backpacking vacation Nonwilderness vacation No vacation 4–7 days (vacation groups)	Proofreading Task
Wu, 2008 (1)	Nonrandomized controlled trial	Taiwan Public psychiatric centre	23	72% male Mean age not reported Schizophrenia patients	Horticulture activities (indoors and outdoors) Regular hospital activities like watching movies, singing, drawing, cooking (indoors) 90 min per week × 15 classes	Chu's Attention Test

(*Continued*)

Table 3C. (Continued).

Author, year ()[a]	Study design	Country and setting	n	Sample characteristics: gender, mean age, population, ethnicity, and s/e status if reported	Intervention characteristics: activity, setting (each group), duration of exposure, and washout period (crossover only)	Attention measures (objective)
Ottosson, 2005	Nonrandomized crossover trial	Sweden Residential care home	17	87% female (after exclusions for n = 15) 86 years Elderly residents of the care home	Leisure time outside (terrace and gardens) Leisure time inside (own room and shared space) 1 h Two sessions, 14 days apart	Digit Span Backward Digit Span Forward Necker Cube Pattern Control Symbol Digit Modalities Test
van den Berg, 2011	Nonrandomized crossover trial	The Netherlands Two care farms	12	83% boys 12.8 years Children diagnosed with ADHD	Building a cabin, natural (woodland) Walking “expedition,” urban (quiet neighborhood) 1 hour Two activities, 1 day apart	Test of Everyday Attention for Children
Kuo, 2001	Natural experiment	United States Inner city community	145	100% female 34 years Heads of household; African American residents of inner city housing development	Living near high levels of vegetation (“green”) Living near low levels of vegetation (“barren”)	Digit Span Backward
Taylor, 2002	Natural experiment	United States Inner city community	169	54% boys 9.6 years Children; African American residents of inner city housing development	High level of near-home nature (“green” view from apartment) Low level of near-home nature (“barren” view from apartment) At least 1 year living in current location	Alphabet Backward Category Matching Delayed Gratification Task Digit Span Backward Matching Familiar Figures Test Necker Cube Pattern Control Symbol Digit Modalities Test Stroop Colour-Word Test
Tennessen, 1995	Natural experiment	United States University	72	42% male 20 years Students	All natural view from dormitory Mostly natural view from dormitory Mostly built view from dormitory All built view from dormitory	Digit Span Backward Digit Span Forward Necker Cube Pattern Control Symbol Digit Modalities Test

Note. Outcome measures listed in alphabetical order, not the order in which they were administered.
[a]Experiment number in parentheses; each experiment has distinct sample; s/e = socioeconomic.

Table 3D. Characteristics of included studies—Other designs; virtual exposures.

Author, year ()[a]	Study design	Country and setting	*n*	Sample characteristics: gender, mean age, population, ethnicity, and s/e status if reported	Intervention characteristics: activity, setting (each group), duration of exposure, and washout period (crossover only)	Attention measures (objective)
Berto, 2005 (2)	Nonrandomized controlled trial	Italy University	64	50% male 23 years Students	Viewing images, natural Viewing images, urban Viewing geometric images Note: Two groups from Berto 2005 (1) 25 images × 15 sec each	Sustained Attention to Response Test

Note. Outcome measures listed in alphabetical order, not the order in which they were administered.
[a]Experiment number in parentheses; each experiment has distinct sample; s/e = socioeconomic.

psychological conditions such as attention deficit hyperactivity disorder (ADHD), depression, or schizophrenia (Berman et al. 2012; Taylor and Kuo 2009; van den Berg and van den Berg 2011; Wu et al. 2008), lower income groups such as African American residents of an inner city housing development (Kuo 2001; Taylor, Kuo, and Sullivan 2002); or participants experiencing other circumstances that, it is suggested, might influence their attention capacity, such as pregnancy or breast cancer (Cimprich and Ronis 2003; Stark 2003). The remainder of the samples had "normal" cognitive function.

Study duration and intensity varied, from less than an hour of exposure in controlled conditions, to multiple days or weeks of exposure in real-life settings. The longest exposures were seen in the natural experiments, where participants had been exposed to their surroundings for months or years (Kuo 2001; Taylor, Kuo, and Sullivan 2002; Tennessen and Cimprich 1995). Investigations used various cognitive tests.

Some studies involved actual exposure to nature: either through active engagement (walking, running, or other activities)(Berman et al. 2008; 2012; Bodin and Hartig 2003; Cimprich and Ronis 2003; Hartig et al. 1996; Hartig, Mang, and Evans 1991; Johansson, Hartig, and Staats 2011; Mayer et al. 2009; Perkins, Searight, and Ratwik 2011; Shin et al. 2011; Stark 2003; Taylor and Kuo 2009; van den Berg and van den Berg 2011; Wu et al. 2008) or passive engagement (resting outside or living with a view) (Kuo 2001; Ottosson and Grahn 2005; Rich 2008; Taylor, Kuo, and Sullivan 2002; Tennessen and Cimprich 1995). Other investigations involved virtual exposure to nature; this was exclusively passive engagement (watching video or viewing images) (Berman et al. 2008; Berto 2005; Chen, Lai, and Wu 2011; Hartig et al. 1996; Laumann, Gärling, and Stormark 2003; Rich 2008; van den Berg, Koole, and van der Wulp 2003). Most studies had a comparison group that involved equivalent exposure to a nonnatural (urban or indoor) setting. Four studies used a placebo control setting involving relaxation time or usual activities (Cimprich and Ronis 2003; Hartig et al. 1996; Hartig, Mang, and Evans 1991; Rich 2008).

Quality scores varied from 22.5 to 75% (Table 4). Seven of the 31 included studies were classified as "high" quality (scoring 67–100%), while 22 were classified as "moderate" (scoring 34–66%) and 2 were classified as "low" quality (scoring 0–33%). The quality indicators reflected overall experimental quality and also how well the study answered our review question, which may not have been the main focus of the individual studies. Indicators that few investigations reported clearly were power calculation, randomization procedure, whether participants were blind to the research question, demonstrated need for attention restoration, whether outcome assessors were blind to group allocation, and whether the sample was representative of the target population.

Evidence for Effects of Nature on Attention Capacity

Some investigations reported results for subgroups, such as men/women (Bodin and Hartig 2003), task/no task (Hartig et al. 1996; 2003), and alone/with friend (Johansson, Hartig, and Staats 2011). One study compared one natural group (walking in woods) and two nonnatural groups (walking in neighborhood and walking in parking lot), but only presented attention scores for the sample as a whole (Perkins, Searight, and Ratwik 2011).

Table 4. Indicators of quality of included studies.

	RCT design; real exposures							RCT design; virtual exposures								
Quality indicators	Cimprich, 2003	Hartig, 2003	Hartig, 1991 (2)	Mayer, 2009 (1)	Mayer, 2009 (2)	Perkins, 2011	Stark 2003	Berto 2005 (1)	Berto 2005 (3)	Chen, 2011 (1)	Hartig, 1996 (1)	Hartig, 1996 (2)	Laumann, 2003	Rich 2008 (1)	Rich 2008 (2)	van den Berg, 2003
Study design																
Power calculation reported	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No
Inclusion criteria reported	Yes	Yes	**Pa.**	No	No	No	Yes	No	No	No	Yes	No	No	No	No	No
Individual level allocation	Yes	Yes	Yes	Yes	Pa.	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Random allocation to groups/ condition order	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Randomization procedure appropriate	Un.	**Yes**	**Yes**	Un.	Un.	Yes	Un.	Un.	Un.	Yes	**Un.**	**Un.**	Un.	Un.	Un.	Un.
Confounders																
Groups similar (sociodemographic)*	Pa.	**Yes**	**Yes**	Un.	Un.	**Yes**	Yes	Un.	Un.	Un.	**Yes**	**Yes**	Un.	Un.	Un.	Un.
Groups balanced at baseline (attention scores)	No	**Yes**	**Pa.**	Un.	Un.	**No**	Pa.	Pa.	Pa.	No	**No**	**No**	Un.	Un.	Un.	Un.
Participants blind to research question	No	**Yes**	**Yes**	Un.	Yes	**Yes**	No	Un.	Un.	Un.	Yes	**Un.**	Un.	Un.	Un.	Un.
Intervention integrity																
Demonstrated need for attention restoration	No	**Pa.**	**Un.**	No	No	No	No	No	No	Yes	**Un.**	**Un.**	Pa.	No	No	No
Clear description of intervention and control	Yes	Yes	Yes	Yes	Yes	Pa.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Pa.	Yes	Yes
Consistency of intervention (within and between groups)	No	Yes	Pa.	Pa.	Yes	Pa.	No	Yes	Yes	Yes	Yes	Pa.	Yes	Un.	Yes	Yes
Data collection methods																
Outcome assessors blind to group allocation	Un.	**No**	**No**	Un.	Un.	**No**	Un.	Un.	Un.	Un.	**No**	**No**	Un.	No	No	Un.
Baseline attention measures taken before the exposure	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes	No	No	Yes	No	No	No
Consistency of data collection	Pa.	Yes	Pa.	Un.	Un.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Un.	Yes	Yes
Analyses																
All attention outcomes reported (means and SD/SE)	Yes	**Pa.**†	**Pa.**	Yes	Yes	No	No†	Yes	Yes	Yes	Yes	Yes	No	No	No	No
All participants accounted for (i.e., losses/exclusions)	Yes	Yes	**Yes**	Un.	Yes	Yes	Yes	Yes	Yes	Yes	**No**	**Yes**	Yes	Un.	Un.	Yes
ITT analysis conducted (all data included after allocation)	No	No	**No**	Un.	No	No	Yes	Yes	Yes	No	**No**	**Yes**	No	Un.	Un.	No
Individual level analysis	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Statistical analysis methods appropriate for study design	Yes	Yes	**No**	Yes	Yes	**Pa.**	Yes	No	No	Pa.	Yes	**Un.**	Yes	Yes	Yes	Yes
External validity																
Sample representative of target population	No	**Un.**	**Un.**	Un.	Un.	**Un.**	No	Un.	Un.	Yes	**Un.**	**Un.**	Un.	Un.	Un.	Un.

(*Continued*)

Table 4. (Continued).

Quality indicators	RCT design; real exposures							RCT design; virtual exposures								
	Cimprich, 2003	Hartig, 2003	Hartig, 1991 (2)	Mayer, 2009 (1)	Mayer, 2009 (2)	Perkins, 2011	Stark 2003	Berto 2005 (1)	Berto 2005 (3)	Chen, 2011 (1)	Hartig, 1996 (1)	Hartig, 1996 (2)	Laumann, 2003	Rich 2008 (1)	Rich 2008 (2)	van den Berg, 2003
Overall quality score																
Total number of points (out of possible 40)	20	30	23	13	17	21	21	21	21	23	20	19	19	9	14	16
Quality rating as percent	50	75	57.5	32.5	42.5	52.5	52.5	52.5	52.5	57.5	50	47.5	47.5	22.5	35	40
Responded to query about "uncertain" ratings	No	Yes	Yes	No	No	Yes	No	No	No	No	Yes	Yes	No	No	No	No

Quality indicators	Other designs; real exposures													Other; virtual	
	Berman, 2008 (1)	Berman, 2012	Bodin, 2003	Johansson, 2011	Shin, 2011	Taylor, 2009	Hartig, 1991 (1)	Wu, 2008 (1)	Ottosson, 2005	van den Berg, 2011	Kuo 2001	Taylor, 2002	Tennessen, 1995	Berman, 2008 (2)	Berto 2005 (2)
Study design															
Power calculation reported	No	Yes	No	No	No	No	No	No	No	No	No	No	No	No	No
Inclusion criteria reported	No	Yes	Yes	Yes	No	Yes	Yes	Un.	Yes	Yes	Yes	Yes	Yes	No	No
Individual level allocation	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Random allocation to groups/condition order	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No	NA	NA	NA	**Yes**	No
Randomization procedure appropriate	**Yes**	**Yes**	**Yes**	**Un.**	Un.	**Yes**	NA	NA	NA	NA	NA	NA	NA	**Yes.**	NA
Confounders															
Groups similar (sociodemographic)*	**Yes**	**Yes**	**Yes**	**Yes**	Un.	**Yes**	**Yes**	Yes	Un.	Yes	Un.	Un.	Pa.	**Yes**	Un.
Groups balanced at baseline (attention scores)	Yes	No	Pa.	No	Yes	Un.	**Yes**	Un.	Un.	Un.	NA	NA	NA	Pa.	Pa.
Participants blind to research question	**Yes**	**Yes**	**Yes**	**Yes**	Un.	**Yes**	**Un.**	Un.	Un.	Yes	Yes	Yes	Un.	**Yes**	Un.
Intervention integrity															
Demonstrated need for attention restoration	No	No	**Un.**	No	No	Yes	No	Un.	No	Yes	NA	NA	NA	No	No
Clear description of intervention and control	Yes	Yes	Yes	Yes	Yes	Pa.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Consistency of intervention (within and between groups)	**Yes**	**Yes**	**Yes**	**Pa.**	Pa.	Yes	No	Yes	Pa.	No	No	No	No	Yes	Yes
Data collection methods															
Outcome assessors blind to group allocation	Un.	Un.	**No**	**No**	Un.	Yes	No	Un.	Un.	No	Un.	Yes	Un.	Un.	Un.
Baseline attention measures taken before the exposure	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	NA	NA	NA	Yes	Yes
Consistency of data collection	Yes	Yes	Yes	Yes	Yes	Yes	Pa.	Un.	Pa.	Un.	Pa.	Pa.	Pa.	Yes	Yes
Analyses															
All attention outcomes reported (means and SD/SE)	Yes	Yes	Yes	Yes	Yes	No	**Yes**	No	No	No	Yes	No	Yes	Yes	Yes
All participants accounted for (i.e., losses/ exclusions)	Yes	Yes	Yes	Yes	Un.	Yes	**Un.**	Yes	Yes	Un.	Un.	Un.	Yes	Yes	Yes

(*Continued*)

Table 4. (Continued).

	Other designs; real exposures													Other; virtual	
Quality indicators	Berman, 2008 (1)	Berman, 2012	Bodin, 2003	Johansson, 2011	Shin, 2011	Taylor, 2009	Hartig, 1991 (1)	Wu, 2008 (1)	Ottosson, 2005	van den Berg, 2011	Kuo 2001	Taylor, 2002	Tennessen, 1995	Berman, 2008 (2)	Berto 2005 (2)
ITT analysis conducted (all data included after allocation)	Yes[†]	No	Yes	Yes	Un.	No	**Un.**	No	No	Un.	Un.	Un.	Yes	Yes	Yes
Individual level analysis	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Statistical analysis methods appropriate for study design	Yes	Yes	Pa.	Yes	No	Yes	**Yes**	Un.	Yes	Un.	Yes	Yes	Yes	Yes	No
External validity															
Sample representative of target population	Un.	**Un.**	Un.	Un.	No	Un.	**Un.**	Un.	No	Un.	Yes	Yes	Un.	Un.	Un.
Overall quality score															
Total number of points (out of possible 40, or fewer where criteria are NA)	30	30	30	27	17	27	21/38	14/38	16/38	14/38	17/30	17/30	18/30	29	19/38
Quality rating as %	75	75	75	67.5	42.5	67.5	55.3	36.8	42.1	36.8	56.7	56.7	60	72.5	50
Responded to query about "uncertain" ratings	Yes	Yes	Yes	Yes	No	Yes	Yes	No	No	No	No	Yes	No	Yes	No

Note. Yes = 2; Partial (Pa.) = 1; No = 0; Unclear (Un.) = 0; NA = criterion not applicable to this study design. Results in boldface reflect changes to "unclear" categorization after further information was provided by authors.

*Significance level for differences between groups was $p < .05$.

[†]Additional data were provided by the author for our meta-analyses. In some cases additional information was provided by study authors for the quality appraisal, and therefore this table would not be entirely replicable by other reviewers. Any changes made after consultations are highlighted in boldface.

Various cognitive tests were used to measure attention capacity. Evidence for the effects of nature on attention capacity is presented below for each attention measure separately although, as noted in the preceding text, it is not clear which of these measures is the most appropriate in the context of ART. Where data could be pooled, forest plots were produced and are reproduced here if three or more studies were included. Meta-analyses pooling only two studies are described in the text, and forest plots are presented in the Supporting Information. Data relating to two outcome measures could not be pooled and these data are reported narratively (Proof Reading Task and Symbol Substitution Test). Measures of attention unique to a single study are described narratively. Full details of all study outcomes are presented in Tables 5A to 5L.

Digit Span

Participants are presented with a series of digits (e.g., "8, 3, 4") and need to immediately repeat them back. If this is done successfully, they are given a longer list of digits (e.g., "9, 2, 4, 0"). The length of the list is increased until the participant fails to accurately recall a list of that length on two subsequent occasions. The length of the longest list a subject can remember is that subject's digit span. In the Digit Span Forward (DSF), participants have to recall the digits in the same order they are presented. In the Digit Span Backward (DSB), participants have to reverse the order with which they are presented.

Digit Span Forward (DSF)

Five studies reported DSF scores (Table 5A) (Cimprich and Ronis 2003; Ottosson and Grahn 2005; Perkins, Searight, and Ratwik 2011; Stark 2003; Tennessen and Cimprich 1995). The meta-analysis included data from three experiments, none of which were balanced at baseline (Cimprich and Ronis 2003; Stark 2003; Tennessen and Cimprich 1995). The natural exposure groups performed significantly better than controls (Figure 2).

Digit Span Backward (DSB)

Eleven studies, reported in 10 articles, reported DSB scores (Table 5B) (Berman et al. 2008, 2012; Cimprich and Ronis 2003; Kuo 2001; Ottosson and Grahn 2005; Perkins, Searight, and Ratwik 2011; Rich 2008; Stark 2003; Taylor and Kuo 2009; Tennessen and Cimprich 1995). The meta-analysis included data from eight investigations reported in seven articles(Berman et al. 2008; 2012; Cimprich and Ronis 2003; Kuo 2001; Stark 2003; Taylor and Kuo 2009; Tennessen and Cimprich 1995). The natural exposure groups performed significantly better than controls (Figure 3). Sensitivity analysis, including only the two studies that were balanced at baseline (Berman et al. 2008, 2012), also indicated better DSB performance in the intervention groups (natural).

Combined Digit Span Backward/Forward (DSB/DSF)

One study reported combined DSB/DSF scores, obtained by summing the two scores (Bodin and Hartig 2003) (Table 5C). The meta-analysis included data from two independent groups (men and women) from one experiment, which were not balanced at baseline (Bodin and Hartig 2003). There was little evidence of a marked difference between groups at follow-up (Figure 4).

Proofreading Task (PR)

The participant is asked to find simple misspellings, typographical errors, and grammatical errors in a five-page passage of text. The score is percent of errors detected from the total present at the point in the text reached after 10 min. Higher scores indicate better performance. Due to the length of the task, it measures attentional vigilance, a key aspect of which is the inhibition of distractions.

One article, containing two studies, reported proofreading scores (Table 5D) (Hartig, Mang, and Evans 1991). The first study found that proofreading scores improved in the natural group (wilderness backpacking) and declined in the two nonnatural groups (nonwilderness vacation and no vacation); the difference in change between groups was not statistically significant. The second experiment reported proofreading scores at follow-up only, which were significantly higher in the natural group (nature walk) compared to the two nonnatural groups

Table 5A. Results of included studies—Digit Span Forward (DSF): Length of sequence repeated correctly.

		Natural settings		Nonnatural settings					
Study design	Author, year, and sample (n)	Baseline mean (SD) and n	Follow-up mean (SD) and n	Baseline mean (SD) and n	Follow-up mean (SD) and n	Baseline mean (SD) and n	Follow-up mean (SD) and n	Difference between groups at follow-up	Difference in change between groups
RCTs	Cimprich, 2003 $n = 185$	6.71 (SE 0.14)	6.98 (SE 0.15)	6.54 (SE 0.13)	6.53 (SE 0.16)	–	–	$p = .04$	NR
	Perkins, 2011 $n = 26$	Woods NR	Woods NR	Neighborhood NR	Neighborhood NR	Parking lot NR	Parking lot NR	NR	No significant difference
	Stark, 2003 $n = 57$	6.8 (1.2)	7.1 (1.4)	6.6 (1.0)	6.7 (1.1)	–	–	NR	NR
Nonrandomized crossover trial	Ottosson, 2005 $n = 17$	NR	NR	NR	NR	–	–	NR	$p < .001$

		Natural settings		Nonnatural settings		
Study design	Author, year, and sample (n)	Follow-up mean (SD) and n	Follow-up mean (SD) and n	Follow-up mean (SD) and n	Follow-up mean (SD) and n	Difference between groups at follow-up
Natural experiment	Tennessen, 1995 $n = 72$	All natural 7.50 (1.08)	Mostly natural 7.40 (1.35)	Mostly built 7.38 (0.90)	All built 6.96 (1.11)	No significant difference

Note. SD, standard deviation; SE, standard error; NR, not reported. There were no baseline values for this study; therefore, no pre–post exposure (baseline to follow-up) change values.

Table 5B. Results of included studies—Digit Span Backward (DSB): Length of sequence reversed correctly.

		Natural settings		Nonnatural settings					
Study design	Author, year ()a and sample (n)	Baseline mean (SD) and n	Follow-up mean (SD) and n	Baseline mean (SD) and n	Follow-up mean (SD) and n	Baseline mean (SD) and n	Follow-up mean (SD) and n	Difference between groups at follow-up	Difference in change between groups
RCTs	Cimprich, 2003 $n = 185$	4.99 (SE 0.15)	5.20 (SE 0.14)	4.51 (SE 0.13)	4.58 (SE 0.14)	—	—	$p = .002$	NR
	Perkins, 2011 $n = 26$	NR	NR	NR	NR	—	—	NR	No significant difference
	Rich, 2008 (2) $n = 36$	NR	NR	NR	NR	—	—	F(1, 35) = 2.43; p = ns	NR
	Stark, 2003 $n = 57$	5.1 (1.3)	5.4 (1.6)	4.7 (1.1)	5.0 (1.6)	—	—	NR	NR
Randomized crossover trials	Berman, 2008 (1) $n = 38$	7.90 (SE 0.37)	9.4 (SE 0.41)	7.90 (SE 0.30)	8.4 (SE 0.33)	—	—	NR	$F(1, 36) = 6.055$; Prep = 0.95
	Berman, 2008 (2) n = 12	7.92 (SE 0.96)	9.33 (SE 0.86)	7.83 (SE 1.04)	8.83 (SE 0.90)	—	—	NR	$F(1, 10) = 0.486$; Prep = 0.68
	Berman, 2012 $n = 20$	7.42 (3.00)	8.63 (2.87)	8.26 (2.51)	7.84 (2.24)	—	—	NR	$F(1, 18) = 20.5$; $p < .001$
	Taylor, 2009 $n = 25$	NR	Urban park 4.41 (1.18)	NR	Downtown 3.82 (1.07)	NR	Neighborhood 3.71 (1.21)	F(2,16) = 4.72; $p = .02$	NR
Nonrandomized crossover trial	Ottosson, 2005 $n = 17$	NR	NR	NR	NR	—	—	NR	$p < .001$

[a]Experiment number in parentheses; each experiment has distinct sample.

		Natural settings		Nonnatural settings		
Study design	Author, year, and sample (n)	Follow-up mean (SD) and n	Follow-up mean (SD) and n	Follow-up mean (SD) and n	Follow-up mean (SD) and n	Difference between groups at follow-up
Natural experiments	Kuo, 2001 $n = 145$	Green 4.96 (1.0)	—	Barren 4.64 (1.2)	—	$t = -1.74$; $p = .05$
	Tennessen, 1995 $n = 72$	All natural 5.60 (1.35)	Mostly natural 5.11 (1.62)	Mostly built 5.04 (1.18)	All built 5.00 (1.13)	No significant difference

Note. There were no baseline values for these studies; therefore, no pre–post exposure (baseline to follow-up) change values.

Table 5C. Results of included studies—Combined Digit Span Backward and Forward (DSB/DSF): Sum of scores.

			Natural settings		Nonnatural settings			
Study design	Author, year, and sample (*n*)	Subgroups within sample	Baseline mean (SD) and *n*	Follow-up mean (SD) and *n*	Baseline mean (SD) and *n*	Follow-up mean (SD) and *n*	Difference between groups at follow-up	Difference in change between groups
Randomized crossover trial	Bodin, 2003 $n = 12$	Men	10.83 (2.32)	11.92 (2.31)	11.92 (2.76)	12.17 (2.89)	NR	Men and women
		Women	11.67 (3.82)	10.58 (3.68)	10.67 ± 2.89	11.25 (3.22)	NR	$F(1, 10) = 0.92$; $p = .36$

Table 5D. Results of included studies—Proofreading Task (PR): Percent of errors detected.

		Natural settings		Nonnatural settings					
Study design	Author, year ()a, and sample (*n*)	Baseline mean (SD) and*n*	Follow-up mean (SD) and *n*	Baseline mean (SD) and *n*	Follow-up mean (SD) and *n*	Baseline mean (SD) and *n*	Follow-up mean (SD) and *n*	Difference between groups at follow-up	Difference in change between groups
RCT	Hartig, 1991 (2) $n = 102$	Natural NR	Natural 56.5	Urban NR	Urban 49.5	Relaxation NR	Relaxation 48.0	$t(94) = 2.45$; $p < .01$	NR
Nonrandomized controlled trial	Hartig, 1991 (1) $n = 68$	Wilderness 51.12	Wilderness 53.92	Nonwilderness 56.77	Nonwilderness 53.78	No vacation 60.16	No vacation 56.52	NR	$F(2, 65) = 2.39$; $p < .09$

[a]Experiment number in parentheses; each experiment has distinct sample.

Table 5E. Results of included studies—Necker Cube Pattern Control (NCPC) (three separate outcomes).

Study design	Author, year, and sample (*n*)	Attention outcome; subgroups within sample	Natural settings: Baseline mean (SD) and *n*	Natural settings: Follow-up mean (SD) and *n*	Nonnatural settings: Baseline mean (SD) and *n*	Nonnatural settings: Follow-up mean (SD) and *n*	Difference between groups at follow-up	Difference in change between groups
RCTs	Cimprich, 2003 n = 185	Percent reduction in reversals (30 sec period)	10.29 (SE 6.61)	19.48 (SE 4.15)	17.87 (SE 7.73)	13.87 (SE 7.31)	No significant difference	NR
	Hartig, 2003 n = 112	Number of reversals (average 2 x 30 s periods); Task	T1 Baseline 4.37 (1.92) T2 During walk 3.96 (1.93)	T3 Follow-up 3.98 (2.44)	T1 3.87 (1.85) T2 4.76 (2.19)	T3 4.67 (2.63)	NR	Task and no task T1-T2 $F(1,98) = 13.15$; $p < 0.001$ T1-T3 $F(1,100) = 5.59$; $p = 0.02$
		Number of reversals (average 2 × 30 sec periods): No task	T1 Baseline 4.17 (1.92) T2 During walk 4.06 (1.90)	T3 Follow-up 3.88 (2.21)	T1 3.70 (1.35) T2 4.43 (1.67)	T3 4.09 (1.72)	NR	
Nonrandomized crossover trial	Ottosson, 2005 n = 17	Difference between baseline and controlled (60 s period)	NR	NR	NR	NR	NR	$p < 0.05$

Note. T1, T2, and T3 used where attention measures were taken at three time points (as described in nature group cells). Underscore: inverse outcome, therefore lower score indicates better performance.

Table 5E. Continued—natural experiment.

Study design	Author, year, and sample (*n*)	Attention outcome	Natural settings: Follow-up mean (SD) and *n*	Natural settings: Follow-up mean (SD) and *n*	Nonnatural settings: Follow-up mean (SD) and *n*	Nonnatural settings: Follow-up mean (SD) and *n*	Difference between groups at follow-up
Natural experiment	Tennessen, 1995 *n* = 72	Percent reduction in reversals (30 sec period)	All natural 60.53 (13.61)	Mostly natural 64.18 (22.48)	Mostly built 35.38 (38.14)	All built 36.35 (39.78)	No significant difference

Note. There were no baseline values for this study, and therefore no pre–post exposure (baseline to follow-up) change values.

Table 5F. Search and Memory Task/Memory Loaded Search Task (SMT).

			Natural settings				Nonnatural settings				
Study design	Author, year ()a sample (*n*)	Attention outcome; subgroups within sample	Baselinemean (SD) and *n*	Follow-up mean (SD) and *n*	Baseline mean (SD) and *n*	Follow-up mean (SD) and *n*	Baseline mean (SD) and *n*	Follow-up mean (SD) and *n*	Baseline mean (SD) and *n*	Follow-up mean (SD) and *n*	Difference between groups at follow-up
RCTs	Hartig, 1996 (1) *n* = 102	Percent error (accuracy) Task	NR	Natural slides Block 1 34.0 (18.3) Block 2 43.4 (16.0)	–	–	NR	Urban slides Block 1 33.0 (15.4) Block 2 44.7 (14.6)	NR	No slides Block 1 33.6 (15.1) Block 2 42.5 (15.5)	No significant difference
		% error (accuracy) No task	NR	Natural slides Block 1 40.6 (18.6) Block 2 37.5 (17.2)	–	–	NR	Urban slides Block 1 33.5 (16.3) Block 2 39.9 (18.7)	NR	No slides Block 1 33.4 (14.5) Block 2 47.1 (18.0)	No significant difference
		Number of letters (speed) Task	NR	Natural slides Block 1 611.4 (156.9) Block 2 679.5 (196.5)	–	–	NR	Urban slides Block 1 688.2 (135.5) Block 2 772.5 (218.2)	NR	No slides Block 1 697.6 (252.0) Block 2 774.7 (222.2)	No significant difference
		Number of letters (speed) No task	NR	Natural slides Block 1 547.2 (155.4) Block 2 573.7 (153.3)	–	–	NR	Urban slides Block 1 591.4 (161.5) Block 2 652.4 (200.2)	NR	No slides Block 1 589.1 (180.4) Block 2 622.1 (207.2)	No significant difference
	Hartig, 1996 (2) *n* = 18	Percent error (accuracy)	NR	Block 1 33.18 (14.20) Block 2 40.21 (13.56)	–	–	NR	Block 1 32.30 (15.55) Block 2 42.97 (15.55)	–	–	No significant difference
		Number of letters (speed)	NR	Block 1 620.90 (139.76) Block 2 656.99 (145.58)	–	–	NR	Block 1 658.39 (177.77) Block 2 718.22 (243.01)	–	–	No significant difference
	Hartig, 2003 *n* = 112	Accuracy x speed	NR	NR	–	–	NR	NR	–	–	NR
	Mayer, 2009 (1) *n* = 76	Number of errors per line completed	NR	1.18 (0.47)	–	–	NR	1.60 (0.74)	–	–	$F(1,67) = 8.49$; $p < 0.01$
	Mayer, 2009 (2) *n* = 92	Number of errors per line completed	NR	Actual nature 1.00 (0.47)	NR	Virtual nature 1.15 (0.59)	NR	Virtual urban 1.41 (0.49)	–	–	$F(1, 56) = 1.31$; $p = 0.28$

Note. There were no baseline values for these studies, therefore no pre–post exposure (baseline to follow-up) change values. The exception was Hartig (2003), which did measure SMT at baseline but did not report this datum; there was no significant difference pre–post exposure.
[a]Experiment number in parentheses; each experiment has distinct sample. Underscore: inverse outcome, therefore lower score indicates better performance.

Table 5G. Results of included studies—Sustained Attention to Response Test (SART) (four separate outcomes).

Study design	Author, year ()a, and sample (n)	Attention outcome	Natural settings		Nonnatural settings				Difference between groups at follow-up	Difference in change between groups
			Baseline mean (SD) and n	Follow-up mean (SD) and n	Baseline mean (SD) and n	Follow-up mean (SD) and n	Baseline mean (SD) and n	Follow-up mean (SD) and n		
RCTs	Berto, 2005 (1) n = 32	Reaction time (msec)	Natural 313.71 (38.36)	Natural 267.38 (73.78)	Urban 319.59 (70.98)	Urban 299.61 (41.43)	—	—	$t(30) = -2.19$; $p = .03$	NR
	Berto, 2005 (1) n = 32	Number of correct responses	Natural 11.68 (5.28)	Natural 13.62 (5.37)	Urban 13.25 (5.09)	Urban 13.00 (5.4)	—	—	$t(30) = 0.32$; $p = .74$	NR
	Berto, 2005 (1) n = 32	Number of incorrect responses	Natural 1.81 (3.83)	Natural 2.06 (4.79)	Urban 3.25 (6.22)	Urban 1.62 (4.96)	—	—	$t(30) = 0.25$; $p = .80$	NR
	Berto, 2005 (1) n = 32	d-prime or sensitivity	Natural 1.40 (0.71)	Natural 1.86 (0.89)	Urban 1.97 (0.96)	Urban 2.00 (0.95)	—	—	$t(30) = -0.40$; $p = .68$	NR
	Berto, 2005 (3) n = 32	Reaction time (msec)	Natural 311.27 (35.6)	Natural 302.22 (32.09)	Urban 306.21 (78.79)	Urban 297.91 (52.98)	—	—	No significant difference	NR
	Berto, 2005 (3) n = 32	Number of correct responses	Natural 14.81 (5.55)	Natural 17.12 (4.09)	Urban 12.5 (6.36)	Urban 14.56 (5.95)	—	—	No significant difference	NR
	Berto, 2005 (3) n = 32	Number of incorrect responses	Natural 1.62 (2.5)	Natural 0.81 (1.42)	Urban 1.68 (2.62)	Urban 0.75 (1.52)	—	—	No significant difference	NR
	Berto, 2005 (3) n = 32	d-prime or sensitivity	Natural 2.12 (1.08)	Natural 2.47 (1.04)	Urban 1.79 (1.24)	Urban 2.03 (0.93)	—	—	No significant difference	NR
Nonrandomized controlled trial	Berto, 2005 (2) n = 64	Reaction time (msec)	No new data; used data from Berto 2005 (1)		No new data; used data from Berto 2005 (1)		Geometric 310.24 (43.89)	Geometric 289.46 (55.2)	$F(2, 61) = 5.57$; $p = .00$; rsq = 0.01	NR
	Berto, 2005 (2) n = 64	Number of correct responses	No new data; used data from Berto 2005 (1)		No new data; used data from Berto 2005 (1)		Geometric 13.90 (5.15)	Geometric 13.59 (5.66)	$F(2, 61) = 4.60$; $p = .01$; rsq = 0.10	NR
	Berto, 2005 (2) n = 64	Number of incorrect responses	No new data; used data from Berto 2005 (1)		No new data; used data from Berto 2005 (1)		Geometric 1.5 (2.59)	Geometric 1.71 (3.12)	No significant difference	NR
	Berto, 2005 (2) n = 64	d-prime or sensitivity	No new data; used data from Berto 2005 (1)		No new data; used data from Berto 2005 (1)		Geometric 1.98 (0.92)	Geometric 1.95 (1.01)	No significant difference	NR

[a]Experiment number in (); each experiment has distinct sample. Inverse outcome therefore lower score indicates better performance.

Table 5H. Results of included studies—Symbol Digit Modalities Test (SDMT): Number of correct symbol/digit pairs (90 sec period).

Study design	Author, Year Sample (n)	Subgroups within sample	Natural settings: Baseline mean (SD) and *n*	Natural settings: Follow-up mean (SD) and *n*	Nonnatural settings: Baseline mean (SD) and *n*	Nonnatural settings: Follow-up mean (SD) and *n*	Difference between groups at follow-up	Difference in change between groups
Randomized crossover trial	Bodin, 2003 *n* = 12	Men Women	51.17 (7.17) 56.33 (11.33)	48.83 (4.22) 52.50 (10.04)	51.50 (3.67) 56.00 (9.36)	49.17 (5.38) 52.83 (9.56)	NR NR	Men and women $F(1, 10) = 0.02$; $p = .90$
Nonrandomized crossover trial	Ottosson, 2005 *n* = 17	Not applicable	NR	NR	NR	NR	NR	$p < .001$

Table 5H. Continued—natural experiment.

Study design	Author, year, and sample (*n*)	Natural settings: Follow-up mean (SD) and *n*	Natural settings: Follow-up mean (SD) and *n*	Nonnatural settings: Follow-up mean (SD) and *n*	Nonnatural settings: Follow-up mean (SD) and *n*	Difference between groups at follow-up
Natural experiment	Tennessen, 1995 *n* = 72	All natural 74.00 (9.87)	Mostly natural 64.40 (10.76)	Mostly built 61.50 (10.36)	All built 63.08 (8.74)	$F(3, 67) = 3.78$; $p < .05$

Note. There were no baseline values for this study, and therefore no pre–post exposure (baseline to follow-up) change values.

Table 5I. Results of included studies—Symbol Substitution Test (SST); number of correct assignments (60 sec period).

Study design	Author, year, and sample (*n*)	Subgroups within sample	Natural settings: Baseline mean (SD) and *n*	Natural settings: Follow-up mean (SD) and *n*	Nonnatural settings: Baseline mean (SD) and *n*	Nonnatural settings: Follow-up mean (SD) and *n*	Difference between groups at follow-up	Difference in change between groups
Randomized crossover trial	Johansson, 2011 *n* = 20	Alone With friend	38.65 (5.28) 40.00 (6.78)	37.85 (5.10) 36.35 (5.09)	37.85 (5.21) 37.70 (4.78)	37.80 (4.87) 36.85 (4.79)	NR NR	Alone and with friend $F(1, 18) = 5.99$; $p = .025$, $n^2 = 0.250$

Table 5J. Results of included studies—Trail Making Test A (TMTA); completion time (seconds).

Study design	Author, year sample (*n*)	Natural settings: Baseline mean (SD) and *n*	Natural settings: Follow-up mean (SD) and *n*	Nonnatural settings: Baseline mean (SD) and *n*	Nonnatural settings: Follow-up mean (SD) and *n*	Difference between groups at follow-up	Difference in change between groups
RCTs	Cimprich, 2003 *n* = 185	30.23 (SE 1.3)	25.65 (SE 1.03)	37.08 (SE 2.3)	31.21 (SE 1.34)	$p = .001$	NR
	Stark, 2003 *n* = 57	22.06 (7.09)	19.84 (4.79)	21.67 (5.46)	19.60 (5.33)	NR	NR

Note. Inverse outcome, therefore lower score indicates better performance.

Table 5K. Results of included studies—Trail Making Test B (TMTB); completion time (seconds).

Study design	Author, year, and sample (*n*)	Subgroups within sample	Natural settings: Baseline mean (SD) and *n*	Natural settings: Follow-up mean (SD) and *n*	Nonnatural settings: Baseline mean (SD) and *n*	Nonnatural settings: Follow-up mean (SD) and *n*	Difference between groups at follow-up	Difference in change between groups
RCTs	Cimprich, 2003 *n* = 185	Not applicable	65.01 (SE 3.4)	56.51 (SE 2.92)	77.09 (SE 4.9)	68.01 (SE 4.06)	$p = .02$	NR
	Stark, 2003 *n* = 57	Not applicable	46.55 (11.92)	38.89 (10.75)	48.15 (10.42)	40.40 (9.30)	NR	NR
Randomized crossover trial	Shin, 2011 *n* = 60	First walk Second walk	37.03 (6.81) 37.04 (6.90)	29.48 (6.82) 29.45 (6.72)	37.03 (6.81) 37.04 (6.90)	39.24 (21.23) 39.17 (21.23)	NR NR	NR NR

Note. Inverse outcome, therefore lower score indicates better performance.

Table 5L. Results of included studies—Other measures of attention (each used only in one study).

Study design	Author, year ()*, and sample (*n*)	Attention test and outcome	Natural settings: Baseline mean (SD) and *n*	Natural settings: Follow-up mean (SD) and *n*	Nonnatural settings: Baseline mean (SD) and *n*	Nonnatural settings: Follow-up mean (SD) and *n*	Nonnatural settings: Baseline mean (SD) and *n*	Nonnatural settings: Follow-up mean (SD) and *n*	Difference between groups at follow-up	Difference in change between groups
RCTs	Chen, 2011 (1) $n = 48$	Coloured Number Pictures: reaction time (msec)	Nature T1 Baseline 587.0 (9.5) T2 After loading task 777.7 (6.9)	Nature T3 Follow-up 557.2 (7.7)	City T1 652.8 (11.2) T2 836.9 (12.3)	City T3 749.8 (10.2)	Urban night T1 615.9 (10.3) T2 709.9 (11.1) Sports T1 600.5 (10.4) T2 808.8 (11.0)	Urban night T3 534.8 (7.3) Sports T3 605.1 (9.4)	$p < .001$	T1–T2–T3 $F = 8.27; p < .001$
	Cimprich, 2003 $n = 185$	Total attention: combined scores (DSB, DSF, TMA, TMB, NCPC)	0.56 (SE 0.34)	1.75 (SE 0.27)	−0.56 (SE 0.37)	0.04 (SE 0.34)	—	—	$p < .001$	$p < .001$
	Perkins, 2011 $n = 26$	Logical Memory: number of segments correctly recalled	Woods NR	Woods NR	Neighborhood NR	Neighborhood NR	Parking lot NR	Parking lot NR	NR	No significant difference
	Rich, 2008 (1) $n = 145$	Stroop Colour-Word: % errors	NR	Forest view 1.05 (0.976)	NR	Buildings 3.53 (0.937)	NR	No view 1.43 (0.950)	$F(2, 129) = 2.61$; $p = .08$	NR
		Vigilance task: outcome unclear	NR	NR	NR	NR	NR	NR	No significant difference	NR
	Stark, 2003 $n = 57$	Category Matching: number of correct pairs circled	41.48 (9.13)	NR	43.82 (9.97)	NR	—	—	NR	NR
		Errors Scale: sum of uncorrected errors (TMA, TMB, CM)	2.6 (2.8)	NR	3.3 (2.3)	NR	—	—	NR	Standardized beta = −0.68; $t = -2.98; p = .001$

Study design	Author, year ()*, and sample (*n*)	Attention test and outcome	Natural settings: Baseline mean (SD) and *n*	Natural settings: Follow-up mean (SD) and *n*	Nonnatural settings: Baseline mean (SD) and *n*	Nonnatural settings: Follow-up mean (SD) and *n*	Difference between groups at follow-up	Difference in change between groups
RCT	van den Berg, 2003 $n = 114$	D2 mental concentration test: concentration index	NR	399.84	NR	379.30	$F(1, 102) = 2.79; p = .098$	NR
		D2 mental concentration test: speed index	NR	421.18	NR	399.92	$F(1, 102) = 2.74; p = .10$	NR
		D2 mental concentration test: Accuracy index	NR	NR	NR	NR	No significant difference ($p = .86$)	NR

(*Continued*)

Table 5L. (Continued).

Study design	Author, year ()*, and sample (n)	Attention test and outcome	Natural settings: Baseline mean (SD) and n	Natural settings: Follow-up mean (SD) and n	Nonnatural settings: Baseline mean (SD) and n	Nonnatural settings: Follow-up mean (SD) and n	Difference between groups at follow-up	Difference in change between groups
Randomized crossover trial	Berman, 2008 (2) n = 12	Attention Network Test: Executive: Response time (msec)	86 (SE 11.30)	67 (SE 8.45)	81 (SE 15.50)	93 (SE 17.96)	NR	$F(1, 10) = 17.089$, Prep = 0.99
		Attention Network Test: Orienting: Response time (msec)	47 (SE 6.46)	55 (SE 7.33)	46 (SE 10.01)	43 (SE 4.73)	NR	NR
		Attention Network Test: Alerting Response time (msec)	32 (SE 6.86)	31 (SE 5.23)	36 (SE 6.52)	46 (SE 5.63)	NR	NR
Nonrandomized crossover trial	van den Berg, 2011 n = 12	Test of Everyday Attention: Difference in time between two tests (sec)	NR	3.20 (1.39)	NR	3.82 (2.47)	eta squared = 0.21; $p = .07$	NR
Nonrandomized controlled trial	Wu, 2008 n = 23	Chu's Attention Test: number of correct answers	45.78	52.89	53.11	47.66	No significant difference	NR
		Chu's Attention Test: number of errors	7.18	6.38	6.29	2.83	Significant difference (no p value)	NR

Note. Experiment number in parentheses; each experiment has distinct sample. T1, T2, and T3 used where attention measures were taken at three time points (as described in nature group cells). Underscore: inverse outcome, and therefore lower score indicates better performance.

Table 5L. Continued—natural experiment.

Study design	Author, year, and sample (n)	Attention test and outcome	Subgroups within sample	Natural settings: Follow-up mean (SD) and n	Nonnatural settings: Follow-up mean (SD) and n	Difference between groups at follow-up
Natural experiment	Taylor, 2002 $n = 169$	Concentration: combined z-scores (SDMT, DSB, AB, NCPC)	Girls	Green view NR	Barren view NR	$F(1, 76) = 10.9$; $B = 0.23$; $p < .01$
			Boys	Green view NR	Barren view NR	No significant difference
		Impulse inhibition: combined z-scores (MFF, SCW, CM)	Girls	Green view NR	Barren view NR	$F(1, 76) = 3.8$; $B = 0.17$; $p = .05$
			Boys	Green view NR	Barren view NR	$F(?, ?) = 2.3$; $B = 0.12$; $p = .13$
		Delay of gratification score	Girls	Green view NR	Barren view NR	$F(1, 76) = 12.7$; $B = 0.42$; $p < .001$
			Boys	Green view NR	Barren view NR	No significant difference
		Self-discipline score: average of the above three scores	Girls	Green view NR	Barren view NR	$F(1, 76) = 19.4$; $B = 0.27$; $p < .001$
			Boys	Green view NR	Barren view NR	No significant difference

Note. There were no baseline values for this study, therefore no pre–post exposure (baseline to follow-up) change values,

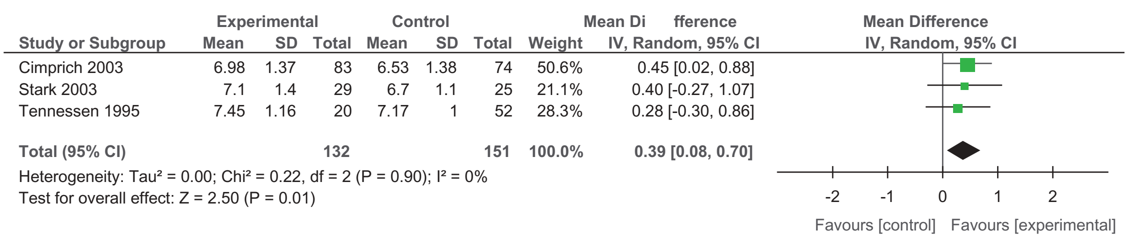

Figure 2. Forest plot showing meta-analysis for DSF.

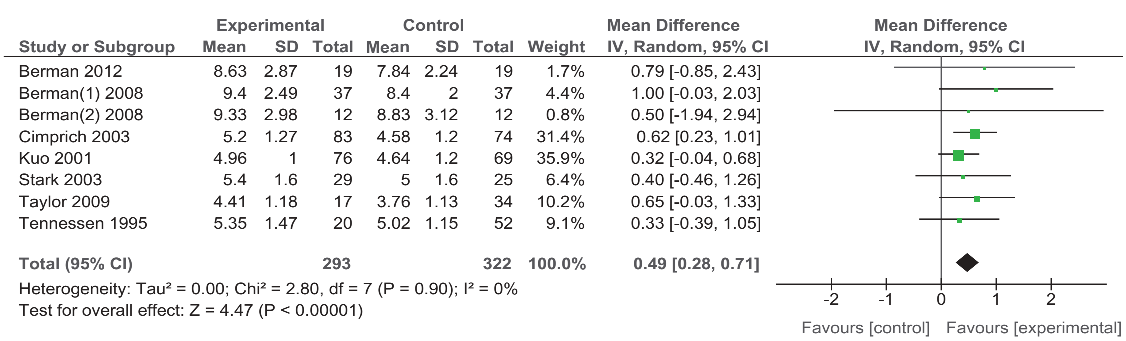

Figure 3. Forest plot showing meta-analysis for DSB.

(urban walk, and passive relaxation indoors). This was an RCT but it was not clear whether the groups were balanced at baseline. Lack of data precluded meta-analysis.

Necker Cube Pattern Control (NCPC)

An image of a three-dimensional cube is presented, which may be perceived from alternative perspectives resulting from reversal of the foreground and background. The participant needs to indicate the number of times the cube appears to "flip" or change perspectives in a short, timed period. The test is performed twice: first with the participant just observing the cube (baseline) and the second time attempting to hold one perspective (controlled). The score may be calculated in various ways, including percent reduction in reversals between the two tests (higher scores indicate better performance), the difference in reversals between the two tests (higher scores are better), or the number of reversals in the controlled test (lower scores are better).

Four studies reported NCPC scores (Table 5E) (Cimprich and Ronis 2003; Hartig et al. 2003; Ottosson and Grahn 2005; Tennessen and Cimprich 1995). Meta-analysis was conducted separately for different calculations of the NCPC score.

For percentage reduction in reversals, two studies were included in the meta-analysis. Study populations were either not balanced at baseline (Cimprich and Ronis 2003) or balance was unknown (Tennessen and Cimprich 1995). There was little evidence of a difference between groups at follow-up. Wide confidence intervals indicate substantial uncertainty in the pooled effect estimate (Figure 5).

For number of reversals, data from two independent groups, reported in the same study, were included in the meta-analysis (Hartig et al. 2003). One group completed a mental loading task prior to the environmental exposure; the other group did not. The groups were not balanced at baseline. Pooled results indicated little evidence of a significant difference between groups at follow-up (Figure 6).

	Experimental			Control				Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI
Bodin(1) 2003	11.92	2.31	12	12.17	2.89	12	63.6%	-0.25 [-2.34, 1.84]
Bodin(2) 2003	10.58	3.68	12	11.25	3.22	12	36.4%	-0.67 [-3.44, 2.10]
Total (95% CI)			**24**			**24**	**100.0%**	**-0.40 [-2.07, 1.27]**

Heterogeneity: Tau² = 0.00; Chi² = 0.06, df = 1 (P = 0.81); I² = 0%
Test for overall effect: Z = 0.47 (P = 0.64)

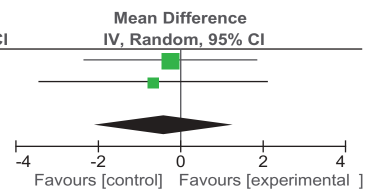

Figure 4. Forest plot showing meta-analysis for combined DSF/DSB.

	Experimental			Control				Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI
Cimprich 2003	19.48	37.8	83	13.87	62.88	74	46.9%	5.61 [-10.86, 22.08]
Tennessen 1995	62.35	18.1	20	35.86	37.6	52	53.1%	26.49 [13.55, 39.43]
Total (95% CI)			**103**			**126**	**100.0%**	**16.70 [-3.72, 37.12]**

Heterogeneity: Tau² = 160.88; Chi² = 3.82, df = 1 (P = 0.05); I² = 74%
Test for overall effect: Z = 1.60 (P = 0.11)

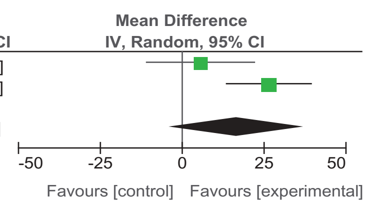

Figure 5. Forest plot showing meta-analysis for NCPC: percentage reduction in reversals.

	Experimental			Control				Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI
Hartig(1) 2003	3.98	2.44	26	4.67	2.63	27	38.0%	-0.69 [-2.06, 0.68]
Hartig(2) 2003	3.88	2.21	26	4.09	1.72	27	62.0%	-0.21 [-1.28, 0.86]
Total (95% CI)			**52**			**54**	**100.0%**	**-0.39 [-1.23, 0.45]**

Heterogeneity: Tau² = 0.00; Chi² = 0.29, df = 1 (P = 0.59); I² = 0%
Test for overall effect: Z = 0.91 (P = 0.36)

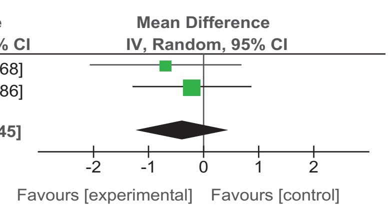

Figure 6. Forest plot showing meta-analysis for NCPC: number of reversals.

Search and Memory Task (SMT)

The participant memorizes 5 target letters, subsequently searches lines of 59 letters, and crosses off any target letters found. Subjects need to complete as many lines, and find as many target letters, as possible in 10 min. The number of target letters per line, and in total, varies between studies. The score may be calculated in various ways, including the percent missed targets (accuracy: lower scores indicate better performance), the number of errors per line (accuracy: lower scores are better), the number of letters searched in a given time (speed: higher scores are better), or accuracy multiplied by speed (higher scores are better). This task combines elements of vigilance and working memory capacity. As targets are essentially random, it might be argued that this is a more demanding task than proofreading.

Five studies, reported in three articles, reported SMT scores (Table 5F) (Hartig et al. 1996; 2003; Mayer et al. 2009). Meta-analysis was conducted separately for different methods of calculating the SMT score.

For percentage error (accuracy), data from three studies, reported in the same article, were pooled (Hartig et al. 1996). Baseline data were not reported, so balance between groups is unknown. The first study reported data for two independent groups, one of which completed a mental loading task prior to environmental exposure. These data appear as separate investigations in the forest plot. For both studies, data were reported in two blocks because data collection was conducted in two halves, completed back-to-back, in order to assess change in accuracy and speed over the course of the task (Table 5F) (Hartig et al. 1996). Data from both blocks were averaged for meta-analysis. There was no evidence of a significant difference between groups at follow-up. The confidence interval indicates substantial uncertainty in the pooled effect estimate (Figure 7).

For number of letters searched (speed), as already described, data from three studies, reported in the same article, were pooled (Hartig et al. 1996). The control groups (nonnatural) performed significantly better than the intervention groups. (Figure 8).

Sustained Attention to Response Test (SART)

One digit (1–9) is assigned as the target. Digits are presented to the participant on a computer screen in quick succession. Individuals need to press the space bar every time a nontarget digit is seen, and avoid pressing the space bar when viewing the target. The one paper that used this test focused on four separate scores: reaction times (lower scores indicate better performance), number of incorrect responses (lower scores are better), quantity of correct responses (higher scores are better), amount of incorrect responses (lower scores are better), and sensitivity (or d-prime), which takes into account correct and incorrect responses simultaneously (higher scores are

	Experimental			Control				Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI
Hartig (1_1) 1996	38.7	17.15	17	38.45	14.9	34	42.1%	0.25 [-9.32, 9.82]
Hartig (1_2) 1996	39.05	17.9	17	38.47	16.7	34	37.1%	0.58 [-9.61, 10.77]
Hartig (2) 1996	36.69	13.88	9	37.63	15.55	9	20.8%	-0.94 [-14.56, 12.68]
Total (95% CI)			**43**			**77**	**100.0%**	**0.13 [-6.08, 6.33]**

Heterogeneity: Tau² = 0.00; Chi² = 0.03, df = 2 (P = 0.98); I² = 0%
Test for overall effect: Z = 0.04 (P = 0.97)

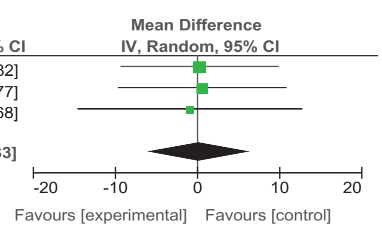

Figure 7. Forest plot showing meta-analysis for SMT: percentage error (accuracy).

	Experimental			Control				Mean Difference
Study or Subgroup	Mean	SD	Total	Mean	SD	Total	Weight	IV, Random, 95% CI
Hartig (1_1) 1996	645.45	176.7	17	733.25	205.98	34	45.4%	-87.80 [-196.65, 21.05]
Hartig (1_2) 1996	560.45	154.35	17	670.87	223.28	34	48.9%	-110.42 [-215.38, -5.46]
Hartig (2) 1996	638.94	142.67	9	688.3	447.99	9	5.7%	-49.36 [-356.53, 257.81]
Total (95% CI)			**43**			**77**	**100.0%**	**-96.66 [-170.03, -23.29]**

Heterogeneity: Tau² = 0.00; Chi² = 0.18, df = 2 (P = 0.91); I² = 0%
Test for overall effect: Z = 2.58 (P = 0.010)

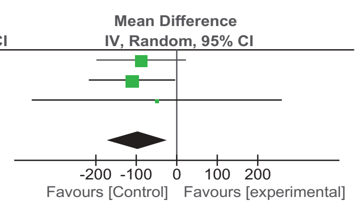

Figure 8. Forest plot showing meta-analysis for SMT: number of letters searched (speed).

better). The SART was designed to measure ability to withhold responses to infrequent and unpredictable stimuli during a period of rapid and rhythmic response to frequent stimuli and so may be viewed as a vigilance task (Robertson et al. 1997).

Three studies within one article reported SART scores (Table 5G) (Berto 2005). Meta-analysis was conducted separately for the four approaches to measuring SART outcomes. However, there are some concerns that outcomes reported in these experiments are not the same as those intended by the developers of SART (Robertson et al. 1997).

For reaction time (milliseconds), the three studies were balanced at baseline (Berto 2005). There was no evidence of a significant difference between groups at follow-up (Figure 9).

For number of correct responses, the three studies were not balanced at baseline (Berto 2005). There was no evidence of a significant difference between groups at follow-up (Figure 10).

For number of incorrect responses, of the three meta-analyzed studies, only one (study 3) was balanced at baseline (Berto 2005). There was little evidence of a significant difference between groups at follow-up (Figure 11). This was also true for the balanced study alone (Berto 2005) (Figure 11).

For sensitivity (or d-prime), the three investigations were not balanced at baseline (Berto 2005).

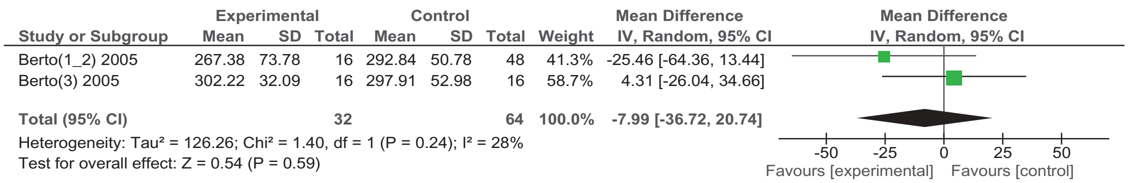

Figure 9. Forest plot showing meta-analysis for SART: reaction time (msec).

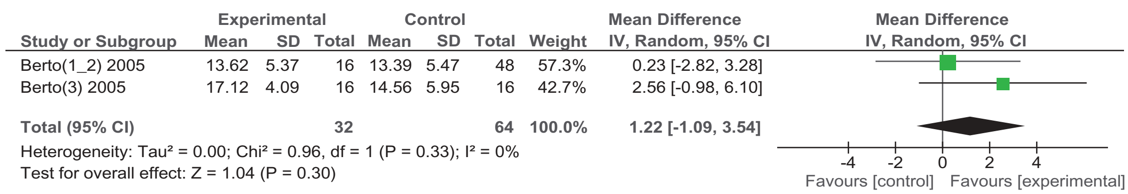

Figure 10. Forest plot showing meta-analysis for SART: number of correct responses.

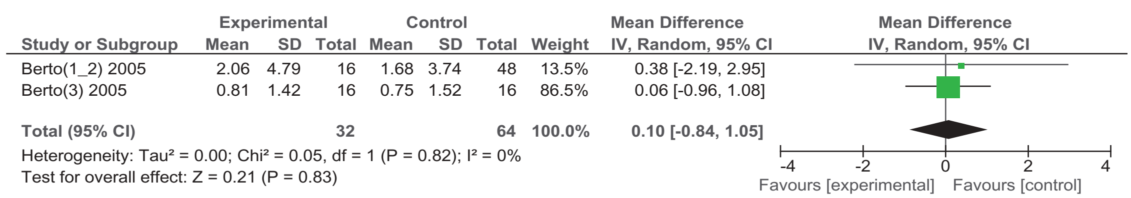

Figure 11. Forest plot showing meta-analysis for SART: number of incorrect responses.

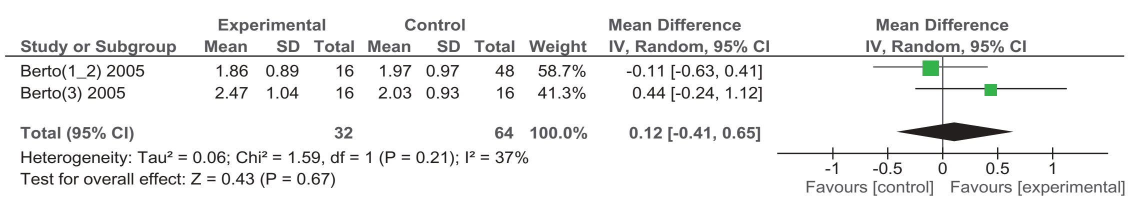

Figure 12. Forest plot showing meta-analysis for SART: sensitivity (or d-prime).

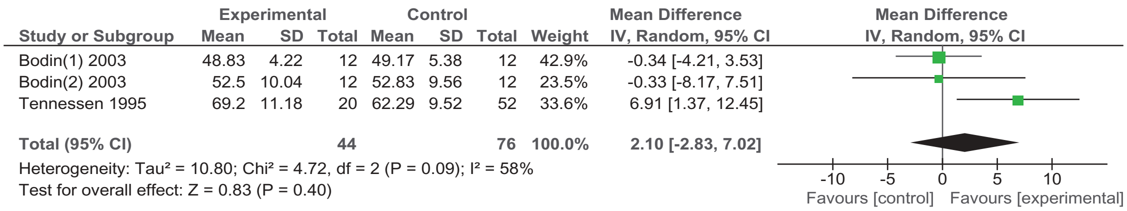

Figure 13. Forest plot showing meta-analysis for SDMT.

There was little evidence of a significant difference between groups at follow-up (Figure 12).

Symbol Digit Modalities Test (SDMT)

The participant is given nine pairs of symbols and digits (e.g., 1#, 2X, 3$, . . . 9%). After practicing writing the correct number under the corresponding symbol on a test sheet, the participant is given a blank copy of the test and asked to write the correct number for each symbol in 90 sec. This is repeated orally. The numbers of correct symbol/digit pairs for the written and oral tests are combined, with higher scores indicating better performance.

Given the complexity of the task, it probably reflects several perceptual, attentional, and executive function processes. Pfeffer et al. (1981) suggested that it is one of the best tools to distinguish between early signs of dementia and depression, indicating that while many of the attentional tests may be affected by mood, the SDMT is tapping into cognitive function over and above any mood effects (p. 524).

Three studies reported SDMT scores as presented in Table 5H (Bodin and Hartig 2003; Ottosson and Grahn 2005; Tennessen and Cimprich 1995). The meta-analysis included data from two studies (Bodin and Hartig 2003; Tennessen and Cimprich 1995). Only the latter was balanced at baseline (Bodin and Hartig 2003). This study reported data for men and women as subgroups and these appear as separate rows in the forest plot. There was little evidence of a significant difference between groups at follow-up (Figure 13). This was also the case for the balanced study alone (Figure 13).

Symbol Substitution Test (SST)

As for the SDMT, participants are given pairs of nine symbols and digits. For the SST, they are asked to assign the correct digits to a series of blanks, each paired with a symbol. After a practice trial, the participant is given 60 sec to fill in as many of the 110 available blanks as possible. The score is the number of correct assignments completed, and higher scores indicate better performance. Comments already given about the processes for SDMT measures also apply to SST, where speed is an even more important consideration. It is unclear why this test was renamed and administered for a shorter test period compared to the SDMT.

Only one investigation reported SST scores (Table 5I) (Johansson, Hartig, and Staats 2011). This study was a crossover design in which participants walked four times: alone and with a friend, in natural and nonnatural settings. SST scores declined for all four conditions, with the greatest

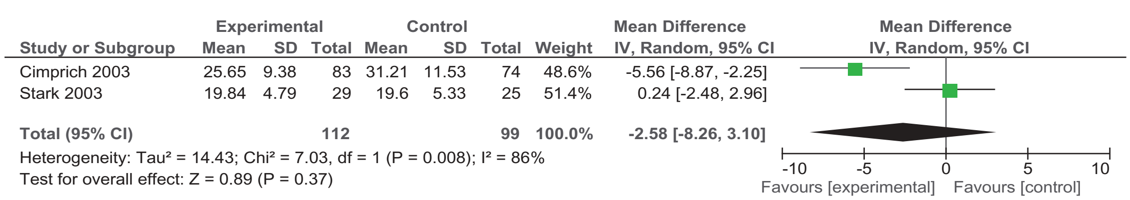

Figure 14. Forest plot showing meta-analysis for TMTA.

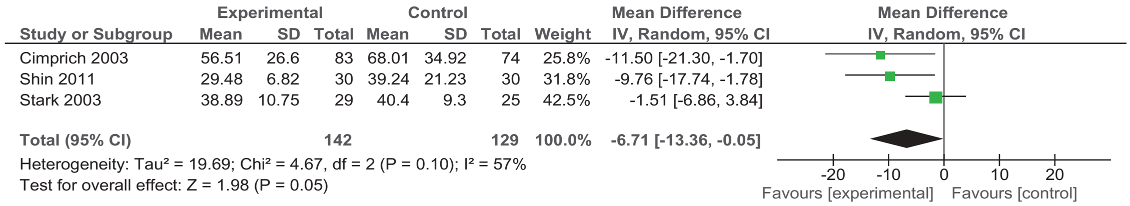

Figure 15. Forest plot showing meta-analysis for TMTB.

decline observed in the natural setting walked with a friend. Differences in change between the natural and nonnatural groups, regardless of social context, were significant. However, SST performance differences between groups were greater at baseline and gaps closed during the intervention, which may reflect regression to the mean. Meta-analysis was not appropriate for this outcome measure because the two environment groups were not independent (Johansson, Hartig, and Staats 2011). Although this test was virtually the same as the SDMT, the difference in test periods (60 vs. 90 sec) indicated that it was not possible to meta-analyze SST/SDMT scores together.

Trail Making Test A (TMTA)

On paper or computer, participants must connect 25 numeric targets (1, 2, 3, 4, 5, etc.) in the correct ascending order as quickly as possible. The score is completion time, so lower scores indicate better performance. Unlike most, the TMTA has a strong motor component, since the respondent has to coordinate actions as well as attention.

Two studies reported TMTA scores and both were included in the meta-analysis (Table 5J) (Cimprich and Ronis 2003; Stark 2003). Only one was balanced at baseline (Stark 2003). There was no evidence of a marked difference between groups at follow-up (Figure 14). The one balanced experiment on its own also provided no evidence of a significant difference between groups at follow-up (Stark 2003) (Figure 14).

Trail Making Test B (TMTB)

This test follows the same procedures as Trail Making Test A, but targets alternate numbers and letters (1, A, 2, B, 3, C, etc.). It places more demands on working memory as participants need to know whether they are shifting from numbers to letters, or vice versa.

All three studies reporting TMTB scores were included in the meta-analysis (Table 5K) (Cimprich and Ronis 2003; Shin et al. 2011; Stark 2003). Only one study was balanced at baseline (Shin et al. 2011). The intervention groups (natural) performed significantly better than controls (Figure 15). This finding was supported by the balanced study on its own (Shin et al. 2011) (Figure 15).

Other Measures of Attention

Ten studies uniquely reported other measures of attention (Table 5L). Three investigations showed that attention performance improved in both (or all) groups, with a significantly greater improvement in the natural group (Chen, Lai, and Wu 2011; Cimprich and Ronis 2003; Stark 2003). The other studies reported incomplete data (Laumann, Gärling, and Stormark 2003; Perkins, Searight, and Ratwik 2011; Taylor, Kuo, and Sullivan 2002; van den Berg and van den Berg 2011) and/or found no significant differences in change between the natural and nonnatural groups (Berman et al. 2008; van den Berg and van den Berg 2011; Wu et al. 2008).

Discussion

Key Findings

This systematic review is based on 31 studies with a variety of study designs, reported in 24 articles, and found some empirical evidence to support ART for three measures such as DSF, DSB, and TMTB; the meta-analyses demonstrated significant evidence that participants exposed to natural settings displayed better postexposure attention scores than those exposed to nonnatural settings. However, meta-analyses for 10 other attention outcomes (using 7 different attention measures) did not show any marked differences between settings. Further, meta-analysis for one attention outcome, SMT, indicated that participants exposed to nonnatural settings displayed significantly better postexposure attention scores than participants exposed to natural settings.

Several measures demonstrating significant effects, such as ANT, have thus far only been employed in a single published study, precluding synthesis. ANT is the only measure that attempts to delineate which of the attention processes (alerting, orientating, or executive processes; Jonides et al. 2008)) may be restored through exposure to natural environments. As noted earlier, more agreement about the most appropriate measures of attention restoration is needed in the field. Further trials using these agreed-on measures would help future appraisals of the theory because more studies may then be included in fewer meta-analyses, resulting in greater power.

It was not always clear how meaningful improvements were. The pooled data from the DSF test represents a mean increase of 0.39 digits recalled by those exposed to natural compared to nonnatural settings. Despite being significant, practical significance in real-world settings is unclear.

A full critique of the outcomes used for ART is beyond the scope of this review, but one can make some observations. Some tasks, including proofreading and SART, appear to measure vigilance processes, with relatively limited demands on executive functions such as working memory. These require that attention is inhibited from shifting to more interesting stimuli than the task, but require few things to be remembered or cognitively manipulated. Other tasks, such as the DSF, DSB, SMT, SDMT, and SST, involve more obvious demands on working memory, in terms of either the amount of information remembered or the need to manipulate it. Further, these measures encompass graded demands on cognitive processes; for instance, DSB involves more working memory and executive function than DSF. Berman et al. (2008) suggested that those tasks concerned with working memory may be most likely to be affected by natural exposure and so most relevant for measuring its impact on attention. However, not all the review analyses support this. Some tasks that imposed higher levels of demand on working memory failed to show significant effects for exposure to nature (SDMT). Conversely, DSF and DSB displayed similar significant effects for exposure to nature compared to controls, although DSB places greater demands on working memory than DSF. It not known whether these anomalies are related to lack of study power and limited numbers of studies contributing to each meta-analysis, low-moderate quality investigations, or inappropriate outcome measures. Again, better understandings of the mechanisms for attention restoration, and the best ways to measure them, are needed.

Only two studies measured attention during the exposure to nature, as well as before and after the exposure (Hartig et al. 2003; Wu et al. 2008). This is potentially important because any positive effects on attention during exposure may also be considered beneficial, even if they do not persist beyond the exposure. Future experiments might consider including “during exposure” measures to determine whether effects of nature are short-lived or longer lasting.

Review Strengths

This is the first systematic review regarding attention restoration potential of natural compared to other settings to focus on objective measures and use systematic methods to identify, select, appraise, and synthesize relevant experimental studies. By focusing on attention, it was possible to include many more relevant studies than Bowler et al. (2010), and also to conduct meta-analysis on several attention measures that had been utilized across several studies, allowing greater confidence in results (e.g., in the effect of nature to impact DSB scores). Nonetheless, a key outcome from the review process as a whole demonstrates that the field has yet to arrive at a clear consensus regarding exactly how to best operationally define “directed attention” as conceptualized by ART. By clearly identifying those attention-related tasks that are most affected by nature, future research may therefore help refine ART by increasing our understanding of which precise attentional processes may be the most relevant.

Review Limitations

The studies included in this systematic review were heterogeneous in terms of study design, experimental population, sample size, attention capacity at baseline, type of natural setting, type of exposure to and engagement with nature, duration of exposure, measures of attention, outcomes, statistical methods used, and data reported. This limited the scope for meta-analysis and it was not possible to determine which groups of individuals, settings, and exposures might result in the greatest attention restoration. Many meta-analyses contained few investigations, reducing their power and preventing statistical determination of publication bias. It is recommended that future studies consider using standardized approaches, to enable subsequent systematic reviews to explore potential differential effects more completely.

The quality of most (22 of 31) of the included studies was rated “moderate,” with two “low” and only seven rated “good.” This was partly because some aspects of experimental design known to be particularly important (such as blinding and randomization procedures) were rarely reported in the included investigations (Schulz et al. 1995). It is also important to acknowledge that some of these studies were published more than 20 years ago, and since then, reporting standards have progressed considerably. As research in this discipline has not previously been judged against systematic review quality appraisal systems, some of the apparent limitations in studies where the authors did not reply to our request for further information may be explained by lack of reporting, rather than by deficits in conduct. Only 9 of 24 investigators replied to our request for further information, and it is possible that this may have introduced an element of bias into our rating process. In addition, it may be hard for subsequent reviewers to fully replicate the current evidence synthesis. In medical and health services research there is greater consensus around best practice reporting standards than in this field (Schulz, Altman, and Moher 2010). There is thus a need for researchers, journal editors, and reviewers to come to an agreement regarding key elements of study conduct required to be reported in experiments in this field, in order for quality to be judged fairly. All of the quality indicators were given equal weighting in this system of appraisal. Therefore, although our quality appraisal system necessarily simplifies a complex issue, it is postulated that this provides much-needed impetus and guidance for better research practice and reporting in future studies.

Ordinarily, quality appraisal in a systematic review would include an appraisal of the validity and reliability of outcome measures reported in the included papers. However feedback from peer reviewers led us to reconsider this. It was suggested that “directed attention,” “voluntary attention,” and “top-down attention” are synonymous, and thus from this perspective any measure considered to be an appropriate way of operationalizing either of these concepts is, in theory, appropriate for either concept. However, it is recognized that any given task may be associated with demands on other resources over and above directed attention. Moreover, since nearly all measures used in the ART literature are examples of widely used attention measures with reliable psychometric properties, virtually all measures might be considered valid and reliable. However, it was also suggested that “directed attention,” as defined by ART, is not clearly elucidated, making it unclear how validity should be determined in the context of measures employed to appraise the attention restoration value of different environments. In this case, it is difficult for researchers to know whether they have adopted an appropriately valid measure. The debate examined was broad and beyond the scope of our review, and subsequently removed this criterion form the quality appraisal tool. It is postulated that the ART community needs to attempt to address this shortcoming in the future such that there is clearer agreement regarding which tools are deemed valid and reliable measures of “directed attention” as defined by ART in future studies.

In order to reduce bias in the meta-analyses, only the follow-up data were used from studies that had not employed appropriate analysis of covariance methods to adjust for baseline imbalance. A potential limitation of discarding baseline data is that information is lost, making it less likely that analyses might detect differences between the trial arms or estimate differences precisely. Where studies use multiple measures, it is possible that performance may be influenced by the order in which they were administered, with previous tests impacting on performance in subsequent ones. It was not possible to account for this possibility in the meta-analyses. There are complexities around establishing the need for restoration in a studied population, which may be influenced by the characteristics of the participants, the nature and length of any loading task and the nature of the measurement task itself, and their interactions. Clear reasons were provided for our assessment of the need for restoration in studies included in this review, but the true picture may be more complicated and requires further investigation.

Finally, the current review was not designed to examine attention outcomes alongside the range of other outcomes claimed to be associated with exposure to natural environments, such as recovery from physiological stress, improvements in mood, encouragement to exercise, facilitating social contact, encouraging optimal development in children, providing opportunities for personal development, and a sense of purpose, despite many of our reviewed studies also including one or more of these outcomes (Mayer et al. 2009). Nonetheless, as the causal mechanisms for attention restoration may be co-related with other restorative effects, further synthesis, building on theoretical developments on how nature may restore individuals through multiple and interacting pathways, is needed.

Research Recommendations

This systematic review highlighted a number of issues for the future research agenda: (1) It is unclear how validity and reliability of measures of directed attention should be assessed, and which are likely to be most sensitive to nature exposure. More needs to be done to articulate which specific characteristics of a task might be important and thus gain a better understanding of exactly which underlying attentional processes nature may influence the most. (2) Meta-analysis would be facilitated if the ART community could articulate more clearly which measures of attention are likely to measure the impact of restoration most appropriately, and then use these measures in a consistent way across multiple studies. (3) Researchers and journal editors should encourage complete reporting of experimental outcomes, including publishing negative findings, so that accurate assessments can be made of the attention restoration potential of natural settings. (4) Investigators and journal editors should work together to agree the key elements of research reporting and experimental conduct to allow an accurate and fair appraisal of study quality to be made by readers and reviewers. (5) Future studies could usefully assess the impact of employing multiple measures, and the order in which they are administered, on the outcomes of attention themselves.

Funding

We thank the authors of previous studies who responded to our queries relating to this review; Rachel Wigglesworth for her help with double data extraction; and Shanker Venkatasubramanian for his advice about cognitive tests. We thank the peer reviewers for their insightful comments on a previous version of this article. The European Centre for Environment and Human The authors of this review are supported by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care (CLAHRC) for the South West Peninsula (AB, OU, VN, RG) and by the European Regional Development Fund and the European Social Fund Convergence Programme for Cornwall and the Isles of Scilly (HO, MW, BW, RG). The views expressed in this publication are those of the authors and not necessarily those of the NHS, the NIHR, the Department of Health in England, or the European Union.

Notes on contributor

All authors contributed to the design of this review, critically revised the article, and approved the final versions. HO contributed to all stages of the systematic review (searching, screening, data extraction, quality appraisal and synthesis) and drafted the article. MW and BW contributed to double data extraction and preparation of the article. AB devised the search strategy, ran the literature searches, carried out citation searching, and contributed to double screening. OU and VN provided statistical advice and designed and conducted the meta-analyses. RG conceived the idea for the review, contributed to double screening, double data extraction, quality appraisal, and preparation of the article, and is the guarantor.

ORCID

Ruth Garside http://orcid.org/0000-0003-1649-4773

References

Baumeister, R. F., K. D. Vohs, and D. M. Tice. 2007. The strength model of self-control. *Current Directions in Psychological Science* 16:351–55. doi:10.1111/cdir.2007.16.issue-6.

Berman, M. G., J. Jonides, and S. Kaplan. 2008. The cognitive benefits of interacting with nature. *Psychological Science* 19:1207–12. doi:10.1111/psci.2008.19.issue-12.

Berman, M. G., E. Kross, K. M. Krpan, M. K. Askren, A. Burson, P. J. Deldin, S. Kaplan, L. Sherdell, I. H. Gotlib, and J. Jonides. 2012. Interacting with na ture improves cognition and affect for individuals with depression. *Journal of Affective Disorders* 140:300–05. doi:10.1016/j.jad.2012.03.012.

Berto, R. 2005. Exposure to restorative environments helps restore attentional capacity. *Journal of Environmental Psychology* 25:249–59. doi:10.1016/j.jenvp.2005.07.001.

Bodin, M., and T. Hartig. 2003. Does the outdoor environment matter for psychological restoration gained through running? *Psychology of Sport and Exercise* 4:141–53. doi:10.1016/S1469-0292(01)00038-3.

Bowler, D. E., L. M. Buyung-Ali, T. M. Knight, and A. S. Pullin. 2010. A systematic review of evidence for the added benefits to health of exposure to natural environments. *BMC Public Health* 10. doi:10.1186/1471-2458-10-456.

Centre for Reviews and Dissemination. 2009. *Systematic reviews: CRD's guidance for undertaking reviews in health care*. York, UK.

Chen, C., Y.-H. Lai, and J.-P. Wu. 2011. Restorative affections about directed attention recovery and reflection in different environments. *Chinese Mental Health Journal* 25:681–85.

Cimprich, B., and D. L. Ronis. 2003. An environmental intervention to restore attention in women with newly diagnosed breast cancer. *Cancer Nursing* 26:284–92. doi:10.1097/00002820-200308000-00005.

Cohen, J.. 2011. A power primer. *Psychological Bulletin* 1992. 112:155–159.

Critical Appraisal Skills Programme. 2013. CASP UK critical appraisal checklists. *Critical Appraisal Skills Programme 2013*. http://www.casp-uk.net (accessed September 2013).

Effective Public Health Practice Project. 2013. *Quality assessment tool for quantitative studies*. http://www.ephpp.ca/tools.html (accessed September, 2013).

Fan, J., B. D. McCandliss, J. Fossella, J. I. Flombaum, and M. I. Posner. 2005. The activation of attentional networks. *Neuroimage* 26:471–79. doi:10.1016/j.neuroimage.2005.02.004.

Fan, J., B. D. McCandliss, T. Sommer, A. Raz, and M. I. Posner. 2002. Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience* 14:340–47. doi:10.1162/089892902317361886.

Fan, M., and Y. Jin. 2013. Obesity and self-control: Food consumption, physical activity, and weight-loss intention. *Applied Economic Perspectives and Policy* 36:125–45. doi:10.1093/aepp/ppt034.

Gifford, R.. 2014. Environmental psychology matters. *Annual Review of Psychology* **65**:541–579.

Hare, T. A., C. F. Camerer, and A. Rangel. 2009. Self-control in decision-making involves modulation of the vmPFC valuation system. *Science* 324:646–48. doi:10.1126/science.1168450.

Hartig, T., A. Book, J. Garvill, T. Olsson, and T. Garling. 1996. Environmental influences on psychological restoration. *Scandinavian Journal of Psychology* 37:378–93. doi:10.1111/j.1467-9450.1996.tb00670.x.

Hartig, T., G. W. Evans, L. D. Jamner, D. S. Davis, and T. Gärling. 2003. Tracking restoration in natural and urban field settings. *Journal of Environmental Psychology* 23:109–23. doi:10.1016/S0272-4944(02)00109-3.

Hartig, T., M. Mang, and G. W. Evans. 1991. Restorative effects of natural environment experiences. *Environment and Behavior* 23:3–26. doi:10.1177/0013916591231001.

Herzog, T. R., A. M. Black, K. A. Fountaine, and D. J. Knotts. 1997. Reflection and attentional recovery as distinctive benefits of restorative environments. *Journal of Environmental Psychology* 17:165–70. doi:10.1006/jevp.1997.0051.

Herzog, T. R., P. Ouellette, J. R. Rolens, and A. M. Koenigs. 2010. Houses of worship as restorative environments. *Environment and Behavior* 42:395–419. doi:10.1177/0013916508328610.

Johansson, M., T. Hartig, and H. Staats. 2011. Psychological benefits of walking: Moderation by company and outdoor environment. *Applied Psychology: Health and Well-Being* 3:261–80. doi:10.1111/aphw.2011.3.issue-3.

Jonides, J., R. L. Lewis, D. E. Nee, C. Lustig, M. G. Berman, and K. S. Moore. 2008. The mind and brain of short-term memory. *Annual Review of Psychology* 59:193–224. doi:10.1146/annurev.psych.59.103006.093615.

Jüni, P., D.G. Altman, and M. Egger. 2001. Restorative affections about directed attention recovery and reflection in different environments. *BMJ* 323.

Kaplan, R., and S. Kaplan. 1989. *The experience of nature: A psychological perspective*. New York, NY: Cambridge University Press.

Kaplan, S. 1995. The restorative benefits of nature: Toward an integrative framework. *Journal of Environmental Psychology* 15:169–82. doi:10.1016/0272-4944(95)90001-2.

Kaplan, S., and M. G. Berman. 2010. Directed attention as a common resource for executive functioning and self-regulation. *Perspectives on Psychological Science* 5:43–57. doi:10.1177/1745691609356784.

Kaplan, S., and J. Frey Talbot. 1983. Psychological benefits of a wilderness experience. In *Behavior and the natural environment*, eds. I. Altman and J. Wohlwill, 163–203. Boston: q6: Springer.

Kuo, F. E. 2001. Coping with poverty: Impacts of environment and attention in the inner city. *Environment and Behavior* 33:5–34. doi:10.1177/00139160121972846.

Laumann, K., T. Gärling, and K. M. Stormark. 2003. Selective attention and heart rate responses to natural and urban environments. *Journal of Environmental Psychology* 23:125–34. doi:10.1016/S0272-4944(02)00110-X.

Mayer, F., S. C. McPherson Frantz, E. Bruehlman-Senecal, and K. Dolliver. 2009. Why is nature beneficial?: The role of connectedness to nature. *Environment and Behavior* 41:607–43. doi:10.1177/0013916508319745.

Ottosson, J., and P. Grahn. 2005. A comparison of leisure time spent in a garden with leisure time spent indoors: On measures of restoration in residents in geriatric care. *Landscape Research* 30:23–55. doi:10.1080/0142639042000324758.

Perkins, S., H. Searight, and S. Ratwik. 2011. Walking in a natural winter setting to relieve attention fatigue: A pilot study. *Psychology* 2:777–80. doi:10.4236/psych.2011.28119.

Pfeffer, R. I., T. T. Kurosaki, C. H. Harrah, J. M. Chance, D. Bates, R. Detels, S. Filos, and C. Butzke. 1981. A survey diagnostic-tool for senile dementia. *American Journal of Epidemiology* 114:515–27.

Rich, D. 2008. Effects of exposure to plants and nature on cognition and mood: A cognitive psychology perspective. *Dissertation Abstracts International: Section B: Sciences Engineering* 68:4911.

Robertson, I. H., T. Manly, J. Andrade, B. T. Baddeley, and J. Yiend. 1997. `Oops!': Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia* 35:747–58. doi:10.1016/S0028-3932(97)00015-8.

Schulz, K. F., D. G. Altman, and D. Moher. 2010. CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. *British Medical Journal* 340:c332. doi:10.1136/bmj.c332.

Schulz, K. F., I. Chalmers, R. J. Hayes, and D. G. Altman. 1995. Empirical evidence of bias: Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *Journal of the American Medical Association* 273:408–12. doi:10.1001/jama.1995.03520290060030.

Shin, W. S., C. S. Shin, P. S. Yeoun, and J. J. Kim. 2011. The influence of interaction with forest on cognitive function. *Scandinavian Journal of Forest Research* 26:595–98. doi:10.1080/02827581.2011.585996.

Staats, H. 2012. Restorative environments. In *The Oxford handbook of environmental and conservation psychology*, ed. S. Clayton, 445–58. New York, NY: Oxford University Press.

Stark, M. A. 2003. Restoring attention in pregnancy: The natural environment. *Clinical Nursing Research* 12:246–65. doi:10.1177/1054773803252995.

Sutton, A. J., K. R. Abrams, D. R. Jones, T. A. Sheldon, and F. Song. 2000. *Methods for meta-analysis in medical research*. Wiley-Blackwell.

Taylor, A. F., and F. E. Kuo. 2009. Children with attention deficits concentrate better after walk in the park. *Journal of Attention Disorders* 12:402–09. doi:10.1177/1087054708323000.

Taylor, A. F., F. E. Kuo, and W. C. Sullivan. 2002. Views of nature and self-discipline: Evidence from inner city children. *Journal of Environmental Psychology* 22:49–63. doi:10.1006/jevp.2001.0241.

Tennessen, C. M., and B. Cimprich. 1995. Views to nature: Effects on attention. *Journal of Environmental Psychology* 15:77–85. doi:10.1016/0272-4944(95)90016-0.

van den Berg, A. E., and C. G. van den Berg. 2011. A comparison of children with ADHD in a natural and built setting. *Child: Care, Health Developments* 37:430–39.

van den Berg, A. E., S. L. Koole, and N. Y. Van Der Wulp. 2003. Environment preference and restoration: (How) are they related? *Journal of Environmental Psychology* 23:135–46. doi:10.1016/S0272-4944(02)00111-1.

Vickers, A. J., and D. G. Altman. 2001. Statistics notes: Analysing controlled trials with baseline and follow up measurements. *British Medical Journal* 323:1123–24. doi:10.1136/bmj.323.7321.1123.

Vohs, K. D., R. F. Baumeister, B. J. Schmeichel, J. M. Twenge, N. M. Nelson, and D. M. Tice. 2008. Making choices impairs subsequent self-control: A limited-resource account of decision making, self-regulation, and active initiative. *Journal of Personality and Social Psychology* 94:883–98. doi:10.1037/0022-3514.94.5.883.

Wu, S. H., C. L. Chang, J. H. Hsu, Y. J. Lin, and S. J. Tsao. 2008. The beneficial effects of horticultural activities on patients' community skill and motivation in a public psychiatric center. *Proceedings of the International Symposium on Horticultur Practices and Therapy for Human Well-Being* 775:55–70.