Restoration of sustained attention following virtual nature exposure: Undeniable or unreliable?

by

Lydia Hicks

A thesis

presented to the University of Waterloo

in fulfilment of the

thesis requirement for the degree of

Master of Arts

in

Psychology

Waterloo, Ontario, Canada, 2020

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

Lydia Hicks was the primary author of the manuscript, conceived the experimental design, and analyzed the data.

Alyssa Smith assisted with editing of the manuscript and data analysis.

Brandon Ralph programmed the experiment and assisted with writing the manuscript.

Daniel Smilek guided the research process, including the experimental design, data analysis, and the writing of the manuscript.

Abstract

Building on previous research examining the influence of natural versus urban images on attention, the purpose of the present experiments was to examine attention restoration with (1) two large samples, (2) a broader image set that was more representative of typical natural and urban environments, and (3) an increased number of task trials to increase the likelihood of more thorough attentional depletion. In both experiments, participants completed the Sustained Attention to Response Task (SART; a measure of sustained attention) before and after they viewed either natural or urban images. When compared to the urban condition, participants in the nature condition did not demonstrate improved performance in the post-image exposure SART. Bayesian analyses also indicated support for the null hypothesis. These findings were replicated in the second experiment, which served to address some additional confounding issues within the stimulus set. These experiments provide evidence that is inconsistent with the foundational finding that images of natural settings are more restorative for attention than images of urban settings.

Acknowledgements

I would like to thank my supervisor, Dan Smilek, for his guidance, encouragement, and enthusiasm while navigating topics both in and outside of our research area. You have taught me how to (a) grapple with data that directly contradicts what I believe and (b) reorganize my perspective based on that data – two very necessary lessons for someone who wants to be as data-driven as possible in academia and beyond. I would also like to thank Alyssa Smith and Brandon Ralph for their encouragement and support, which was specifically helpful during this current project, but also while navigating the ups and downs of grad school. Thank you as well to Jeremy Marty-Dugas and Anna Hudson for their mentorship and friendship throughout this masters, and thank you to the rest of my lab mates for their support and camaraderie as well: Emilie, Allison, Tyler, Bruno, and Marty. Pertaining specifically to this master's project, the process was made much easier with help from fellow graduate students Austin Hurst and Sarah McCrackin, as well as one very dedicated research assistant, Brian Kim – thank you to each of you. Finally, thank you to the Natural Science and Engineering Research Council of Canada (NSERC) who supports our lab through a discovery grant to Daniel Smilek. This work is currently under review at the *Journal of Environmental Psychology*.

Dedication

This thesis is dedicated to my two sets of parents: Mom and Bob, and Dad and Lorrie. Thank you for your unconditional love, encouragement, and support, which helped me to realize my own potential and brought me to this point. As well, this is dedicated to my stand in third set of "parents" – Aunt Cathy and Uncle Tom – thank you for always treating me like one of your own; I know the doors are always open, the conversation is always stimulating, and the nature is always restorative at your welcoming home.

Table of Contents

List of Figures

List of Tables

Introduction

As attention is neccesary for human performance (James, 1890), understanding how we can restore our attention capacities after a period of fatigue has become an important question. Researchers have proposed that exposure to natural environments may more effectively restore depleted attentional capacities than exposure to urban environments, and that this nature-related restoration leads to improved post-exposure attentional performance (Berman, Jonides, & Kaplan, 2008; Berto, 2005; Kaplan & Kaplan, 1989; for a review, see Ohly et al., 2016, or Stevenson, Schilab, & Bentsen, 2018). To explain the potential restorative benefits of natural environments, Kaplan and colleagues (Kaplan, 1995; Kaplan & Kaplan, 1989) proposed the Attention Restoration Theory (ART). ART posits that there are two distinct types of attention: 'directed attention', which is controlled and effortful; and 'involuntary attention', which is uncontrolled and effortless. ART rests on the assumption that directed attention depends on a limited resource that can be depleted, and therefore, requires periods of rest in which it may be restored. According to ART, natural environments hold more restorative potential than urban environments because they are more likely to gently, and involuntarily, guide the viewer's attention in such a way that directed attention is not required and has a chance to rest (a situation referred to as "soft fascination", Kaplan, 1995). For example, one may think of how his or her attention is guided while viewing aspects of natural environments, such as a sunset or a calmly flowing river. In direct contrast, urban environments are proposed to attract attention in a more forceful manner (i.e., "hard fascination"; Kaplan, 1995), thus demanding the engagement of directed attention to buffer against the distracting stimuli that surround the individual. For example, one may imagine the multitude of distractions experienced while walking down a busy sidewalk in Time Squares in New York City. Urban environments are therefore less likely to provide opportunity for restoration of one's directed attention resources.

ART has been tested by examining how people's performance during attention tasks is impacted by exposure to various type of environments. Much of the past research has examined the influence of *real-world* natural and urban environment exposure on performance in attention-related tasks (Berman, Jonides, & Kaplan, 2008a; Berman, et al., 2012; Bodin & Hartig, 2003; Hartig, Mang, & Evans, 1991; Hartig et al., 2003; Johansson, Hartig, & Staats, 2011; Mayer, McPherson Frantz, Bruehlman-Senecal, & Dolliver, 2009; Perkins, Searight, & Ratwick, 2011; Shin, 2011; Taylor & Kuo, 2009). However, the influence of other types of exposures on attention has also been examined, including *window-view* exposures (Rich, 2008; Tennessen & Cimprich, 1995), and *virtual image/video* exposures (Berman, Jonides, & Kaplan, 2008b; Berto, 2005; Hartig et al., 1996; Laumann, Gärling, & Stormark, 2003; Van den Berg, Koole, & Van der Wulp, 2003). Examination of these alternative exposure types is important because such exposures may provide opportunities for attention restoration for individuals who face barriers to accessing real-world natural environments (e.g., accessibility limitations, time limitations, financial limitations, pandemic-related isolation, and other socioeconomic factors); these alternatives are especially relevant in our modern, technology-driven society, in which we are constantly surrounded by screens that hold the potential for attention restoration in a variety of contexts. Here we focus on how attention is influenced by exposure to natural and urban photographs.

One study that has provided foundational support for attention restoration via virtual exposures to photographs was reported by Berto (2005). In this study, 32 participants completed two rounds of an attention task, separated by a period in which participants either viewed a set of images rated as "restorative" (which also all happened to be all natural environments, $n = 16$), or a set of images rated as "non-restorative" (which also happened to be all urban environments, $n = 16$). These images for the two conditions (high restoration and low restoration) were selected based on ratings of predicted restoration (using the Perceived Restorativeness Scale (PRS); Hartig et al., 1997), collected in a pre-scaling study. The attention task used was the Sustained Attention to Response Task (SART), which is a measure commonly used to detect changes in one's ability to maintain focus on a task to detect infrequent target stimuli (Roberston, Manly, Andrade, Baddeley, & Yiend, 1997). The SART consists of a series of digits ranging from 1 to 9, which are presented one at a time at a constant rate on the computer screen. Participants are required to make key-press responses to non-target stimuli (e.g., 1, 2, 4-9) and withhold the key-press response to a target stimulus (e.g., 3). The dependent variables of interest in this task include response times to non-target stimuli and failures to withhold a response when the target appears (i.e., commission errors); however, Berto (2005) reported results in terms of correct responses for target trials (the opposite of commission errors), so that is discussed here instead. In contrast to participants who viewed the images with low restoration ratings, participants who viewed the images with higher restoration ratings made significantly more correct responses on target trials, had significantly faster response times, and demonstrated improved sensitivity for detecting a target (d-prime) during the post-image-exposure SART in comparison to the pre-image-exposure SART. Based on these results, Berto (2005) concluded that images with high subjective ratings of perceived restoration were indeed objectively restorative as indexed by performance on an attention task, while images with low subjective ratings of perceived restoration were not. *Importantly,* Berto then stated that the results of the experiment(s), "underscore ART's premise that natural environments are more restorative than urban environments," (Berto, 2005, p. 258).

In the present investigation, our goal was to more directly test the quoted conclusion drawn by Berto (2005): specifically, are natural environments *indeed* more restorative than urban environments. As Berto (2005) is continuously cited as an example of the general idea that nature environments confer greater attentional restoration than do urban environments (e.g., Abraham, Sommerhalder, & Abel, 2010; Atchley, Strayer, & Atchley, 2012; Berman, Jonides, & Kaplan, 2008; Bratman, Daily, Levy & Gross, 2015; Bratman, Hamilton, & Daily, 2012; Carrus et al., 2015; Kaplan & Berman, 2010; Sonnentag, Venz, & Casper, 2017; van den Berg, Maas, Verheig, & Groenewgen, 2010), rather than being cited as an example of images with high restoration ratings confer greater attentional restoration than images with low restoration ratings, we found this an important issue to bring to light within the field of attention restoration. Our extension of Berto's (2005) work was based on a reconsideration of the experimental design, the findings, and the conceptual claims made in the initial report, as well as on a consideration of the subsequent replications of the initial findings (Craig, 2016; Craig & Klein, 2012; Craig, Klein, Menon, & Rinaldo, 2015; Dumke, 2014; Lee, Williams, Sargent, Williams, & Johnson, 2015; Nguyen, Nelson, & Klein, 2018, Sanguinetti, 2009; Valtchanov, 2010).

As a starting point, it is important to consider the stimuli used in the original studies and how these support the commonly drawn conclusion from Berto's studies. Recall that the critical inference made by Berto (2005)—that natural environments are more restorative than urban environments—is what is often referenced when other researchers discuss Berto's work. However, this inference may be problematic because the images used in Berto's study (2005) were selected based on predicted *restorativeness* (high vs. low), and not exclusively on *environment type* (nature vs. urban). The images selected for the high-restoration group only happened to be all natural images, while the images selected for the low-restoration only happened to be all urban images. However, it could be possible that there are natural environments that are low in predicted restorativeness (e.g., a swamp bursting with mosquitos and black flies) and urban environments that are high in predicted restorativeness (e.g., a calm and quiet museum) that were not represented in the stimulus set used by Berto (2005). In other words, Berto's stimulus set contained an explicit confound between predicted restorativeness and environment type (nature vs. urban). Contrary to the common citation of Berto's study as reflecting restoration differences between exposure to urban and nature images, the findings should have been attributed to the criterion for image selection, which was 'restorativeness'.

Perpetuating the foregoing problem, several attempted replications of Berto's (2005) findings have continued to use confounded stimuli. In fact, some attempts at replication have used the exact same stimulus set that Berto (2005) used (Craig, 2016; Craig & Klein, 2012; Craig, Klein, Menon, & Rinaldo, 2015), with the authors sometimes explicitly noting that natural images were selected based on their higher perceived restorativeness ratings, while urban images were selected based on their lower perceived restorativeness ratings. Other attempts at replication used different stimulus sets, but continued the tradition of selecting stimuli more on the basis of restorativeness than the nature vs. urban distinction (Dumke, 2014, Nguyen, Nelson, & Klein, 2018), thus maintaining the possible environment type vs. restorativeness level confound.

Unfortunately, previous attempts at conceptual replications of Berto (2005) have had other stimulus limitations. Specifically, in several studies, researchers used only one stimulus in each of the urban and nature conditions. For instance, Valtchanov (2010) compared restoration with only one virtual reality (VR) nature environment and one VR urban environment in the investigation. Lee, Williams, Sargent, Williams, and Johnson (2015) included a single image of each condition – the same urban rooftop with and without greenery in their study. And, Sanguinetti's (2009) stimuli consisted of one view from an open window that included nature and another view of the closed window (without nature). Critically, the use of a single stimulus per condition is problematic because it substantively limits the generalizability of the findings and allows for the possibility that the stimuli selected differ on a variety of dimensions other than the nature vs. urban distinction, such as beauty and familiarity, to give a few examples.

Finally, we note that the extant replications of Berto's (2005) findings have yielded mixed results. Some of these replication attempts supported the idea that natural environments could lead to improved SART performance relative to urban environments (Craig, 2016; Craig et al., 2015; Dumke, 2014), while others were unable to find differences in SART performance after exposure to either natural or urban environment type (Craig & Klein, 2012; Nguyen, Nelson, & Klein, 2018; Sanguinetti, 2009; Valtchanov, 2010). The inconsistency in the results across these studies may be in part due to the underwhelming sample sizes used in the extant studies. For example, in Berto's original study, only 16 participants were included in each group. There were less than 25 participants per group in the studies by Craig (2016), Craig and colleagues (2015), and Valtchanov (2010), and less than 20 participants per group in the studies by Dumke (2014) and Nguyen, Nelson, and Klein (2018). Sanuinetti (2009) and Craig and Klein (2012) had the fewest participants in their studies, with a total of 11 and 10 participants, respectively. The only study that had a more acceptable sample size required for identifying a potentially small effect was one reported by Lee and colleagues (2015), with around 75 participants per group. While the results of Lee and colleagues (2015) supported the idea that nature environments can provide cognitive benefits, as discussed above, their findings were based on the comparison between groups who saw either one image of a rooftop with greenery or one image of a rooftop without greenery, which, as we noted above, limits the generalizability of the findings.

It is important to note that since all of the studies mentioned above were sizably underpowered (with the exception of Lee and colleagues, 2015), and the number of unpublished successful and failed replications is unknown, it would be impossible to generate a reliable meta-analysis on these data to predict the replicability of Berto's (2005) results. Critically, such a meta-analysis would not be useful because many of the studies also confounded the concepts of restorativeness and nature. Thus, it is clear that a well-defined, well-designed, and well-powered replication attempt of Berto (2005) is necessary to uncover the replicability of these results.

The Present Studies

Given the abovementioned concerns, the purpose of the current investigation is threefold. First, we aimed to more directly test the conclusion drawn by Berto (2005)—that exposure to natural environments is more attentionally restorative than exposure to urban environments—using a stimulus set that is representative of participants' conceptions of environments that are typical of natural settings and environments that are typical of urban settings. To gauge participants' conceptualizations of typical natural and urban environments, we collected participants' ratings of how typical various images were of natural and urban settings. Importantly, our stimuli were selected on the basis of *typicality* of each environment type, rather than on the basis of restorativeness, making this more a conceptual replication rather than a direct replication.

Second, we aimed to test Berto's (2005) conclusion with two much larger samples that would be sufficient to detect a fairly small effect size. As the previous studies have included small samples, it was decided that the effect sizes from those studies would not be used to calculate the sample size for this study. Instead, the required sample size was calculated assuming a small effect size. We determined that a minimum of 80 participants per group was required to detect a fairly small effect size (d = 0.30), with a power of .80 and a two-tailed alpha value of .05.

Finally, in our experiments, we increased the number of trials in an attempt to more thoroughly deplete attention during the pre-exposure SART and to gain a better understanding of the potential attention restoration that might occur during the post-exposure SART. As previous research examining sustained attention has demonstrated that performance declines as a function of time-on-task (a.k.a., the 'vigilance decrement'; Head, Russell, Dorahy, Neumann, & Helton, 2012; Helton & Russell, 2011; Helton & Warm, 2008; Mackworth, 1948), we predicted that attention depletion would be the greatest at the end of the pre-exposure SART trials, and that attention restoration would be the strongest at the beginning of the post-exposure SART trials. In an attempt to more similarly match the analysis of Berto (2005) that occurred with far fewer trials, we planned to contrast performance in the last portion of the pre-exposure SART and first portion of the post-exposure SART, allowing for the most optimal conditions to find evidence for attentional restoration from exposure to natural, rather than urban, images.

Experiment One

In Experiment 1[1], we sought to provide a conceptual replication of Berto (2005) with images selected based on ratings of environment typicality rather than ratings of perceived restorativeness. If images of natural environments are indeed more restorative for attentional performance than images of urban environments, then there should be a significant interaction between the exposure period (pre vs. post-image SART block) and the type of images presented (nature vs. urban), which should occur for at least one of the SART measures (i.e., attentional performance should be better for those who viewed images of nature, but only during the post-exposure SART).

Method

Participants. Two hundred and two undergraduate students (*mean age* = 19.77, *SD* = 1.89) from the University of Waterloo participated in this study. Thirteen people were excluded because they participated prior to the finalization of the experimental protocol, and nine participants were excluded due to deviations in the experimental protocol, resulting in a total of 180 participants to be included in the analysis. One-hundred and twenty-three participants were female, 54 were male, one participant identified as male-to-female, and two participants' genders were unknown. Participants were randomly assigned to either the nature or the urban condition. Participants received partial course credit in exchange for participating in the experiment. This study was reviewed and approved for ethic clearance through the University of Waterloo Research Ethics Committee (ORE# 40442).

Materials.

Demographic Questionnaire. Demographic information (age and gender) was collected from each participant at the beginning of the experimental session.

Image Selection. One hundred images were selected from various free desktop background websites. Fifty images contained mostly natural settings and 50 images contained mostly urban settings. The goal was to sample a wide selection of images that varied on a spectrum from natural to urban environments. No settings contained people. Image selection was based on the types of environments that Berto (2005) included in the original picture set: lakes, rivers, seas, hills, woods, orchards, forests, city riversides, city streets, industrial zones, housing, porches, urban areas, and skyscrapers.

Prior to the commencement of the experiment proper, a pre-scaling study was conducted to determine which subset of images from the initial set of 100 images would be included in the experiment proper. In this pre-scaling study, 100 participants from the Amazon Mechanical Turk participant pool evaluated the initial set of images in a Human Intelligence Task (HIT) posted on Amazon Mechanical Turk (www.mturk.com). These participants were compensated $3.50 for a HIT that lasted about 30 minutes and consisted of a brief demographic questionnaire followed by the image-rating task. During the image rating task, participants viewed each image for 6 seconds. As the type of computer screen was likely to vary for each participant who completed the task, the images were first cropped into squares, which were then programmed to take up 50 percent of each participant's monitor size. Thus, the images were programmed to appear in the same dimensions, but pixel size was calculated relative to each participant's monitor size. After viewing each image, participants were asked to respond to two slider scales. The first slider scale read: “That is a place that matches my idea of a typical natural environment”, and the second slider read: “That is a place that matches my idea of a typical urban environment”; both sliders were anchored by, “(0) Strongly Disagree” and “(100) Strongly Agree”. All participants viewed and rated the same two practice images at the beginning of the experiment, which were used to familiarize participants with the task and were not included in the ratings. After reporting their responses for a given image, participants would then rate the next image. This continued until participants had viewed and rated all 100 images. The pre-scaling experiment was created using HTML, Javascript, CSS and PHP.

The ratings of the images from the pre-scaling study were then used to select which images would make up the stimulus sets for the urban and natural conditions in the replication of Berto (2005). To select the images, the ratings across participants were averaged for each image, and then difference scores were calculated for each image from those ratings by subtracting the average urban typicality values from the average nature typicality values. Then, the twenty-five images with the largest positive difference scores were selected as the images for the nature condition, and the twenty-five images with the largest negative difference scores were selected as the images for the urban condition. Once the images were selected, they were balanced on contrast and luminance using Matlab programming (McCrackin & Itier, 2018; Willenbockel et al., 2010).

Sustained Attention to Response Task (SART). The SART (Robertson et al., 1997) is a computer task in which participants are presented a series of digits (1-9) one at a time and must actively monitor for infrequent targets. Participants were instructed to press the space bar for each “go” non-target digit (*1, 2, 4-9*), and withhold responses to the “no-go” target digit, (*3*). For each trial, a single digit from 1 to 9 was randomly selected without replacement to appear for 250 ms in the center of the screen, followed by a "+" mask presented for 900 ms, resulting in a total trial duration of 1,150 ms. After each digit from the set of 1 to 9 was presented once (which defined a 'block'), the process began again. This meant that the critical no-go target digit was presented pseudo-randomly over the duration of the task. The SART was comprised of 70 blocks of nine trials each, for a total of 630 trials, resulting in a total task length of 12 minutes and 4.5 seconds. All digits were presented in white on a black background, using the Courier New typeface in one of five randomly selected sizes: 120, 100, 94, 72, and 48 points. The SART was programmed using Python 2 and was presented on Philips LCD (Model: 224E2SB/27) monitors presenting in 1920x1080p resolution.

The primary measures of interest obtained from the SART were commission errors, defined as instances when participants failed to withhold a response to a no-go trial and therefore an important measure for indexing failures in sustained attention (Robertson et al., 1997). Secondary measures of attention lapses included go-trial reaction times (RTs) – defined as the mean response latencies for all of the go-trials in which a response was made – and omission errors – defined as instances when participants failed to respond to a go-trial (Cheyne, Solman, Carriere, & Smilek, 2009; Robertson et al., 1997).

Image Viewing Task. Using Python 2, the fifty images (25 per condition) selected from the pre-scaling study were presented after the first SART and before the second SART in the experiment proper. The images were presented in colour with dimensions of 800x800px in the center of the screen. The rest of the screen surrounding the images was black. Each image was presented for 15 seconds, resulting in a total image-viewing period of 6 minutes and 15 seconds.

Procedure. Participants were tested either individually, or in groups ranging from two to four, in a laboratory room at the University of Waterloo. Prior to the experimental session, participants were assigned to either the nature or the urban conditions, which dictated the kind of images that were presented to them during the experiment. After providing informed consent, participants responded to demographic questions presented on their computer workstation. Participants were then provided with verbal instructions via the experimenter and written instructions via the computer program, regarding the tasks they were to complete during the experimental session. Specifically, they were told that they would first complete 18 practice trials of the SART, then complete the first set of experimental SART trials, then view a series of images, and finally complete a second set of experimental SART trials. Participants were given thorough instructions regarding how to complete the SART and were told that they should try to respond as quickly and as accurately as possible. The experimenters never referred to the SART as a "sustained attention task" when speaking to the participants; instead they referred to is as a "computer-target-detection task."

After the instructions were given, participants were asked to place *Peltor Sport Shotgunner* headphones over their ears. This was done to prevent participants from being distracted by the sounds (e.g., key presses) generated by other participants in the room. Participants then completed the practice trials, and if they misunderstood the goal of the task, the experimenter corrected them. When all participants understood the task, they were prompted to begin the experiment proper, and the experimenter left the room. Participants completed 630 trials of the first SART, which was followed by the image viewing task consisting of 25 images in their respective condition (i.e., either nature or urban). After the image viewing task was completed, participants completed another 630 trials of the SART. In total, the experiment took a total of about 50 minutes to complete.

Results

Data Exclusion and Analysis Overview. Data was first analyzed for violations of normality. Initial descriptive statistics revealed that, while skew and kurtosis values were normal for response times and commission errors, the outputs for omission errors were not within the acceptable skew and kurtosis range ($|skew| < 3.0$ and $|kurtosis| < 10$, as per the recommendations of Kline, 1998). This was due to a few outliers in the high end of the omission distribution. To correct for this, participants with an average omission error rate of 3 standard deviations or greater from the mean was excluded listwise from each condition pre- and post-image exposure (as per of the protocol of Seli, Jonker, Solman, Cheyne, & Smilek, 2013). This was done as higher omission rates may be indicative of participants who failed to understand the instructions or willfully decided not to follow instructions. After removing those participants, the skew and kurtosis values for the omissions of each group fell within the normal range (Kline, 1998). The final sample consisted of 171 participants, with 85 participants in the nature condition, and 86 participants in the urban condition.

The data were analyzed in two ways. The first analysis examined each of the SART dependent variables (commission errors, RTs, and omission errors) based on data from all trials of the pre- and post-exposure SARTs. The second analysis included only the last 243 trials of the pre-exposure SART and the first 243 trials of the post-exposure SART, in an attempt to more directly mimic Berto’s (2005) analysis. In addition, analyses are reported with omission outliers both removed and included.

To determine whether the data provided evidence for the null hypothesis (i.e., that image type does not influence SART performance), we conducted a Bayes analysis. The Bayes factor allowed us to assess the evidence in favour of the null hypothesis (H_0) and the alternative hypothesis (H_1) given a distribution for the prior probability of these hypotheses. A BF_{01} is the ratio of the likelihood of the data under the null hypothesis compared to the alternative hypothesis. We interpreted the Bayes factors based on the recommendations of Jeffreys (1961).

Analyses of All SART Trials. The data for each performance metric of the full-length pre- and post-image exposure SARTs were submitted to separate 2 x 2 mixed ANOVAs with Image Type (Nature vs. Urban) as a between participant variable and SART Block (Pre-exposure vs. Post-exposure) as a within-participant variable. The means (and confidence intervals) of each of the SART dependent variables for each of the pre-and post-exposure SARTs, averaged across participants in each of the Image Type groups, are shown in Table 1.

Table 1
Descriptive Statistics for All SART Trials

Image Type	SART Block	N	Commission Errors	Reaction Time	Omission Errors
Urban	Pre-Exposure	86	.47 [.43, .52]	381 [361, 401]	.02 [.02, .03]
	Post-Exposure	86	.49 [.44, .54]	382 [362, 403]	.04 [.03, .06]
Nature	Pre-Exposure	85	.44 [.40, .48]	390 [371, 409]	.02 [.01, .02]
	Post-Exposure	85	.45 [.40, .50]	386 [366, 405]	.03 [.02, .03]

Note. Square brackets denote upper and lower limits of 95% confidence intervals.

Commission errors. When examining the *commission errors*, there were no significant main effects of either SART Block ($F(1, 169) = 1.96$, $p = .163$ $\eta_2 = 0.01$) or

Image Type (F(1, 169) = 1.62, p = .205, η^2 < .01). There was also no significant interaction between SART Block and Image Type (F(1, 169) = 0.59, p = .445, η^2 < .01). The Bayes factor also provided substantial evidence that a model excluding the interaction better approximates the data than a model with the interaction, BF_{01} = 5.00. Without excluding outliers, all tests remain non-significant (all F(1, 178) < 1.49, all p > .224, and evidence against favoring a model with the interaction was again substantial BF_{01} = 6.27).

Reaction times. When examining the *reaction times*, there was no main effect of SART Block (F(1, 169) = 0.21, p = .646, η^2 < 0.01), no main effect of Image Type (F(1, 169) = 0.21, p = .646, η^2 < 0.01), nor a significant interaction between SART Block and Image Type (F(1, 169) = 0.53, p = .467, η^2 < 0.01). The Bayes factor provided substantial evidence against the model with the interaction, BF_{01} = 5.41. With no exclusions, all tests remain non-significant (all F(1, 178) < 0.95, all p > .330, and evidence against the inclusion of the interaction remained consistent, BF_{01} = 3.62).

Omission errors. When examining the *omission errors*, there was a significant main effect of SART Block (F(1, 169) = 25.53, p < .001, η^2 = 0.13), such that participants made more omission errors in the post-exposure SART than pre-exposure SART. In addition, there was a significant main effect of Image Type (F(1, 169) = 5.99, p = .015, η^2 = 0.03), whereby participants in the Urban condition made more omission errors than those in the Nature condition. However, the interaction between SART Block and Image Type did not reach significance (F(1, 169) = 2.88, p = .092, η^2 = 0.02). There was anecdotal evidence against the inclusion of the interaction, with a Bayes factor of BF_{01} = 1.81.With no exclusions, the main effects of SART Block and Image Type remained significant (F(1, 178) = 17.62, p < .001; F(1, 178) = 4.05, p = .046, respectively), while the interaction remained non-significant (F(1, 178) = 1.41, p = .236). However, with no exclusions, the Bayes factor now provided substantial evidence against the inclusion of the interaction, BF_{01} = 3.45.

Analyses of Reduced SART Trials. To examine SART performance in the number of trials that better matched the number of trials used in Berto (2005)—which only included 240 trials per SART—each performance metric was submitted to separate 2 x 2 mixed ANOVAs conducted with only the last 243 trials of the pre-exposure SART and the first 243 trials of the post-exposure SART. As with the previous analysis, Image Type (Nature vs. Urban) was included as a between participant variable and SART Block (pre-exposure vs. post-exposure) was included as a within-participant variable. Descriptive statistics for SART performance metrics corresponding to the reduced SART trials are included in Table 2.

Table 2
Descriptive Statistics for Reduced SART Trials

Image Type	SART Block	N	Commission Errors	Reaction Time	Omission Errors
Urban	Pre-Exposure	86	.52 [.46, .57]	384 [361, 406]	.03 [.02, .04]
	Post-Exposure	86	.47 [.41, .52]	381 [359, 403]	.03 [.02, .05]
Nature	Pre-Exposure	85	.49 [.44, .54]	387 [365, 409]	.02 [.02, .03]
	Post-Exposure	85	.42 [.37, .48]	383 [362, 404]	.02 [.01, .03]

Note. Square brackets denote upper and lower limits of 95% confidence intervals.

Commission errors. When examining the *commission errors* of the last 243 trials of the pre-exposure SART and the first 243 trials of the post-exposure SART, there was a main effect of SART Block (F(1, 167) = 22.11, p < .001, η^2 = 0.12), such that less commission errors were made (regardless of Image Type) during the post-exposure SART compared to the pre-exposure SART. However, there was no significant main effect of Image Type ($F(1, 167) = 1.15$, $p = .286$, $\eta_2 < 0.01$), and no significant interaction between SART Block and Image Type ($F(1, 167) = 0.48$, $p = .489$, $\eta_2 < 0.01$). The Bayes factor was also calculated and it provided substantial evidence that the model without the interaction was better than the model with an interaction, $BF_{01} = 4.99$. With no exclusions, the main effect of SART Block remained significant ($F(1, 178) = 21.32$, $p < .001$), while the main effect of Image type and the interaction remained non-significant (both $F(1, 178) < 0.93$, both $p > .335$). The Bayes factor continued to provide substantial evidence against the inclusion of the interaction, $BF_{01} = 5.67$.

Reaction times. When comparing the *reaction times* of the last 243 trials of the pre-exposure SART and the first 243 trials of the post-exposure SART, there was no main effect of SART Block ($F(1, 167) = 0.94$, $p = .334$, $\eta_2 < 0.01$), no main effect of Image Type ($F(1, 167) = 0.04$, $p = .842$, $\eta_2 < 0.01$), nor an interaction between SART Block and Image Type ($F(1, 167) = 0.05$, $p = .818$, $\eta_2 < 0.01$). The Bayes factor indicated substantial evidence against the inclusion of the interaction in a model of the data, $BF_{01} = 5.83$. With no exclusions, all tests remain non-significant all ($F(1, 178) < 1.93$, all $p > .167$), and the Bayes factor continued to provide substantial evidence against the inclusion of the interaction, $BF_{01} = 5.70$.

Omission errors. When comparing the *omission errors* of the last 240 trials of the pre-exposure SART and the first 240 trials of the post-exposure SART, there was a significant main effect of Image Type ($F(1, 167) = 5.35$, $p = .022$, $\eta_2 = 0.03$), such that participants in the nature condition made fewer omission errors (regardless of exposure period) than those in the urban condition. However, there was not a significant main effect of SART Block ($F(1, 167) = 0.14$, $p = .708$, $\eta_2 < .01$), and there was not a significant interaction between SART Block and Image Type ($F(1, 167) = 0.57$, $p = .451$, $\eta_2 = 0.01$). Again, the Bayes factor provided evidence favoring a model of the data without an interaction, $BF_{01} = 4.53$. With no exclusions, the main effect of Image Type remained significant ($F(1, 178) = 4.21$, $p = .042$), while both the main effect of SART Block and the interaction remained non-significant (both $F(1, 178) < 1.96$, both $p > .163$). The Bayes factor continued to provide substantial evidence against the inclusion of the interaction, $BF_{01} = 6.16$.

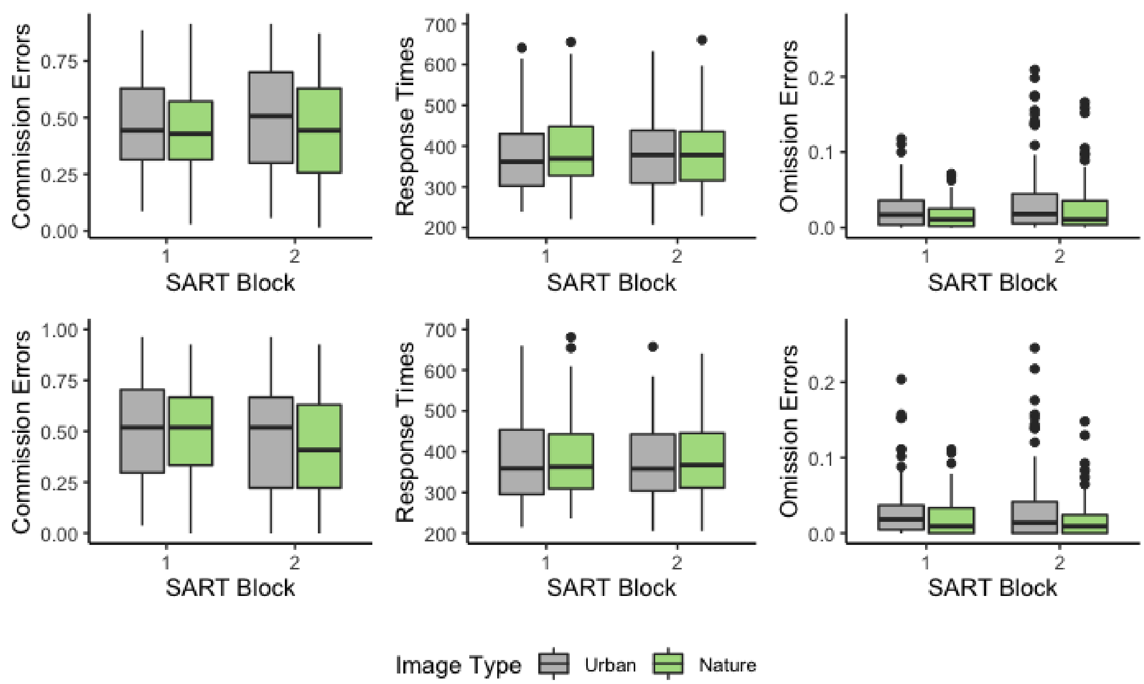

Figure 1. Box and whisker plots (boxplots) representing data from each measure in both analyses. Data from all SART trials are portrayed in the top row, while data from the reduced SART trials are portrayed in the bottom row. From left to right, the graphs show data from commission errors, response times, and omission errors. Data from the urban condition are portrayed in grey, while data from the nature condition are portrayed in green. No significant interactions were found for any of the measures.

Discussion

Experiment 1 was an attempt to directly test a commonly cited conclusion drawn by Berto (2005), namely that exposure to images of natural environments restores attention-related performance more effectively than does exposure to images of urban environments. In this experiment, it was reasoned that if the type of environment to which participants were exposed influences attention-related performance, then there should be a significant interaction between the relative time the attention task was completed (pre vs. post-image SART block) and the type of images presented (nature vs. urban), and that this interaction should occur for at least one of the SART performance measures. While time-on-task effects were evident via the main effects of SART block found for commission errors and omission errors, we found no evidence for interactions between SART Block and Image Type for *any* of our measures. Indeed, Bayes analyses suggested strong evidence for the null hypotheses. Therefore, our results are not consistent with the conclusion forwarded by Berto that images of natural environments are more restorative for attention than images of urban environments.

It is important to note here a few potential limitations of the stimulus set used in Experiment 1. First, small areas of pixilation were introduced to some images by the algorithm used to match the images on contrast and luminance. Second, some could argue that perhaps there were confounding differences between the groups of images in characteristics such as scope, visual angle, and artistic quality – important characteristics to consider as they may impact perceived beauty, which has been considered as a dimension of restorative quality (Herzog, Black, Fountaine, & Knotts, 1997; Van den Berg, Joy, & Koole, 2016). The purpose of Experiment 2 was to provide another replication of this experiment while also addressing these two abovementioned concerns.

Experiment Two

In Experiment 2, we sought to replicate the results from Experiment 1 with a large sample while addressing two main concerns: pixilation of images and other potential aesthetic confounds. We aimed to address these concerns by (a) using the images as they came instead of balancing them on contrast and luminance (making for a more realistic viewing experience), and (b) balancing images across conditions on ratings of scope, visual angle, and artistic quality. As before, image selection within each condition was also based on typicality.

Method

Participants. One hundred and seventy undergraduate students (*mean age* = 19.92, *SD* = 2.51) from the University of Waterloo participated in this study. Data from 10 participants was deemed preliminary as they were collected while training research assistants on how to run the experimental protocol. Two additional participants were excluded because of deviations to the protocol during the experiment, resulting in 158 participants to be included in the analysis. Again, participants were randomly assigned to either the nature or the urban condition. Participants received partial course credit in exchange for participating in the experiment. This study was reviewed and approved for ethic clearance through the University of Waterloo Research Ethics Committee (ORE #40442).

Materials.

The materials were identical to those used in Experiment 1, except for the images.

Image Selection. To reiterate, in Experiment 1, we selected the 25 images with the largest positive difference scores for the nature condition, and the 25 images with the largest negative difference scores or the urban condition; once the images were selected, they were balanced on contrast and luminance. In contrast, in this experiment we first ran another round of the pre-scaling study on Amazon's Mechanical Turk to collect typicality ratings for an additional 100 images (50 nature and 50 urban). Participants who completed this study were also compensated $3.50 per HIT which lasted about 30 minutes and consisted of demographic questions followed by the same image-rating task as the previous experiment. Programming for this pre-scaling study was the same used in the previous experiment. Following the pre-scaling study, ratings for the 100 images collected here and for the 100 images collected in the first pre-scaling study from experiment one were combined such that typicality ratings for 200 images were available. We then selected the 40 images with the largest positive difference scores to begin selecting for the nature condition, and the 40 images with the largest negative difference scores to begin selecting for the urban condition. Following this, the experimenter gave each of the 80 images subjective ratings on scope (1 being "close up", 2 being "middle", and 3 being "panoramic"), visual angle (1 being "ground", and 2 being "sky"), and artistic quality (1 being "low", 2 being "moderate", and 3 being "high"). After that, those scores were averaged across the top 25 images for each group and compared. Neither groups differed significantly in their scores of scope, visual angle, and artistic quality. In addition, these images were not balanced on contrast and luminance in this version of the experiment to remove the potential confound of pixilation that had occurred in some of the images in the previous experiment.

Procedure. The procedure was identical to that of Experiment 1.

Results

Data Exclusion and Analysis Overview. The data from the 158 participants (81 in the urban condition, 77 in the nature condition) were analyzed for violations of normality. Again, initial descriptive statistics revealed that, while skew and kurtosis values were normal for response times and commission errors, the outputs for omission errors were not within the acceptable skew and kurtosis range (|skew| < 3.0 and |kurtosis| < 10, as per the recommendations of Kline, 1998). To correct for this, participants with an average omission error rate of 3 standard deviations or greater from the mean was dropped listwise from each condition pre- and post-image exposure (as per of the protocol of Seli, Jonker, Solman, Cheyne, & Smilek, 2013). After removing those participants (four from the urban condition and four from the nature condition), the skew and kurtosis values for the omissions of each group fell within the normal range (Kline, 1998). The final sample consisted of 150 participants, with 77 participants in the nature condition, and 73 participants in the urban condition.

The data were again analyzed in two ways: first with an analysis of all trials, and then with an analysis of reduced trials (only the last 243 trials of the pre-exposure SART and the first 243 trials of the post-exposure SART). Analyses are also reported first with the omission outliers removed and second with the outliers included. Bayes factors in favour of the null hypothesis are reported as well, and again interpreted based on the recommendations of Jeffreys (1961).

Analyses of All SART Trials. The data for each performance metric of the full-length pre- and post-image exposure SARTs were submitted to separate 2 x 2 mixed ANOVAs with Image Type (Nature vs. Urban) as a between participant variable and SART

Block (Pre-exposure vs. Post-exposure) as a within-participant variable. The means (and confidence intervals) of each of the SART dependent variables for each of the pre-and post-exposure SARTs, averaged across participants in each of the Image Type groups, are shown in Table 1.

Table 1
Descriptive Statistics for All SART Trials

Image Typey	SART Block	N	Commission Errors	Reaction Time	Omission Errors
Urban	Pre-Exposure	73	.46 [.42, .50]	382 [364, 400]	.02 [.02, .03]
	Post-Exposure	73	.49 [.44, .54]	389 [367, 410]	.04 [.02, .05]
Nature	Pre-Exposure	77	.47 [.43, .52]	375 [356, 393]	.02 [.02, .03]
	Post-Exposure	77	.53 [.47, .58]	370 [351, 389]	.04 [.03, .05]

Note. Square brackets denote upper and lower limits of 95% confidence intervals.

Commission errors. When examining the *commission errors*, there were no significant main effects of Image Type ($F(1, 148) = 0.55$, $p = .460$, $\eta_2 < 0.01$). There was a significant main effect of SART Block ($F(1, 148) = 17.38$, $p < .001$, $\eta_2 = 0.11$), such that participants made more commission errors in the post-exposure SART than the pre-exposure SART. However, the interaction between SART Block and Image Type was not significant ($F(1, 148) = 0.83$, $p = .364$, $\eta_2 = 0.01$). The Bayes factor also provided substantial evidence that a model excluding the interaction better approximates the data than a model with the interaction, $BF_{01} = 3.70$. Without excluding omission error outliers, the main effect of SART Block remained significant ($F(1, 156) = 14.56$, $p < .001$, $\eta_2 = 0.09$), while the main effect of Image Type and the interaction remained non-significant ($F(1, 156) = 0.49$, $p = .485$, $\eta_2 < 0.01$; $F(1, 156) = 1.08$, $p = 0.299$, $\eta_2 = 0.01$, respectively). Evidence against favoring a model with the interaction was again substantial, $BF_{01} = 3.50$.

Reaction times. When examining the *reaction times*, there was no main effect of Image Type ($F(1, 148) = 1.00$, $p = .319$, $\eta^2 = 0.01$), no main effect of SART Block ($F(1, 148) = 0.05$, $p = .821$, $\eta^2 < 0.01$), nor a significant interaction between SART Block and Image Type ($F(1, 148) = 1.50$, $p = .223$, $\eta^2 = 0.01$). The Bayes factor provided anecdotal evidence against the model with the interaction, $BF_{01} = 2.87$. With no exclusions, all tests remain non-significant (all $F(1, 156) < 1.29$, all $p > .258$). However, with no exclusions, the Bayes factor now provided substantial evidence against the inclusion of the interaction, $BF_{01} = 3.22$.

Omission errors. When examining the *omission errors*, there was a significant main effect of SART Block ($F(1, 148) = 21.06$, $p < .001$, $\eta^2 = 0.13$), such that participants made more omission errors in the post-exposure SART than pre-exposure SART. However, there was no significant main effect of Image Type ($F(1, 148) < 0.01$, $p = .966$, $\eta^2 < 0.01$), nor a significant interaction between SART Block and Image Type ($F(1, 148) = 0.03$, $p = .866$, $\eta^2 < 0.01$). There was substantial evidence against the inclusion of the interaction, with a Bayes factor of $BF_{01} = 5.75$. With no exclusions, the main effect of SART Block remained significant ($F(1, 156) = 23.22$, $p < .001$, $\eta^2 = 0.13$), while the main effect of Image Type and the interaction remained non-significant ($F(1, 156) = 0.06$, $p = .804$, $\eta^2 < 0.01$; $F(1, 156) < 0.01$, $p = 0.956$, $\eta^2 < 0.01$, respectively). Evidence against favoring a model with the interaction was again substantial, $BF_{01} = 6.06$.

Analyses of Reduced SART Trials. To examine SART performance in the number of trials that better matched the number of trials used in Berto (2005)—which only included 240 trials per SART—each performance metric was submitted to separate 2 x 2 mixed ANOVAs conducted with only the last 243 trials of the pre-exposure SART and the first 243 trials of the post-exposure SART. As with the previous analysis, Image Type (Nature vs. Urban) was included as a between participant variable and SART Block (pre-exposure vs. post-exposure) was included as a within-participant variable. Again, statistics are reported first with omission outliers removed and second with outliers reported. In this reduced analysis, only three participants from the urban condition and three participants from the nature condition were removed, resulting in the inclusion of 152 participants in the outliers removed analysis and 158 in the outliers included analysis. Descriptive statistics for SART performance metrics corresponding to the reduced SART trials are included in Table 2.

Table 2
Descriptive Statistics for Reduced SART Trials

Image Type	SART Block	N	Commission Errors	Reaction Time	Omission Errors
Urban	Pre-Exposure	74	.49 [.44, .55]	394 [371, 417]	.03 [.02, .04]
	Post-Exposure	74	.46 [.40, .52]	388 [365, 411]	.03 [.02, .04]
Nature	Pre-Exposure	78	.50 [.45, .56]	384 [363, 404]	.03 [.02, .04]
	Post-Exposure	78	.49 [.43, .55]	377 [356, 397]	.04 [.02, .05]

Note. Square brackets denote upper and lower limits of 95% confidence intervals.

Commission errors. When examining the *commission errors* of the last 243 trials of the pre-exposure SART and the first 243 trials of the post-exposure SART, there was no main effect of Image Type ($F(1, 150) = 0.19$, $p = .660$, $\eta_2 < 0.01$), no main effect of SART Block ($F(1, 150) = 3.47$, $p = .064$, $\eta_2 = 0.02$), nor a significant interaction between Image Type and SART Block ($F(1, 150) = 0.33$, $p = 0.567$, $\eta_2 < 0.01$). The Bayes factor was also calculated, and it provided substantial evidence that the model without the interaction was better than the model with an interaction, $BF_{01} = 4.64$. With no exclusions, the main effect of SART Block became significant ($F(1, 156) = 5.83$, $p = .017$, $\eta_2 = 0.04$), while the main effect of Image Type and the interaction remained non-significant ($F(1, 156) = 0.33$, $p =$ 566; $F(1, 156) = 0.37$, $p = .544$, respectively). The Bayes factor continued to provide substantial evidence against the inclusion of the interaction, $BF_{01} = 4.72$.

Reaction times. When comparing the *reaction times* of the last 243 trials of the pre-exposure SART and the first 243 trials of the post-exposure SART, there was no main effect of SART Block ($F(1, 150) = 1.73$, $p = .190$, $\eta_2 = 0.01$), no main effect of Image Type ($F(1, 150) = 0.54$, $p = .462$, $\eta_2 < 0.01$), nor a significant interaction between SART Block and Image Type ($F(1, 150) = 0.03$, $p = .864$, $\eta_2 < 0.01$). The Bayes factor indicated substantial evidence against the inclusion of the interaction in a model of the data, $BF_{01} =$ 6.16. With no exclusions, all tests remain non-significant (all $F(1, 156) < 2.11$, all $p > .148$), and the Bayes factor continued to provide substantial evidence against the inclusion of the interaction, $BF_{01} = 5.93$.

Omission errors. When comparing the *omission errors* of the last 240 trials of the pre-exposure SART and the first 240 trials of the post-exposure SART, there no main effect of Image Type ($F(1, 150) = 0.04$, $p = .835$, $\eta_2 < 0.01$), no main effect of SART Block ($F(1, 150) = 1.59$, $p = .209$, $\eta_2 = 0.01$), nor a significant interaction between SART Block and Image Type ($F(1, 150) = 0.40$, $p = .530$, $\eta_2 < 0.01$). Again, the Bayes factor provided evidence favoring a model of the data without an interaction, $BF_{01} = 4.70$. With no exclusions, all tests remain non-significant (all $F(1, 156) < 2.70$, all $p > .104$), and the Bayes factor continued to provide substantial evidence against the inclusion of the interaction, $BF_{01} = 4.63$.

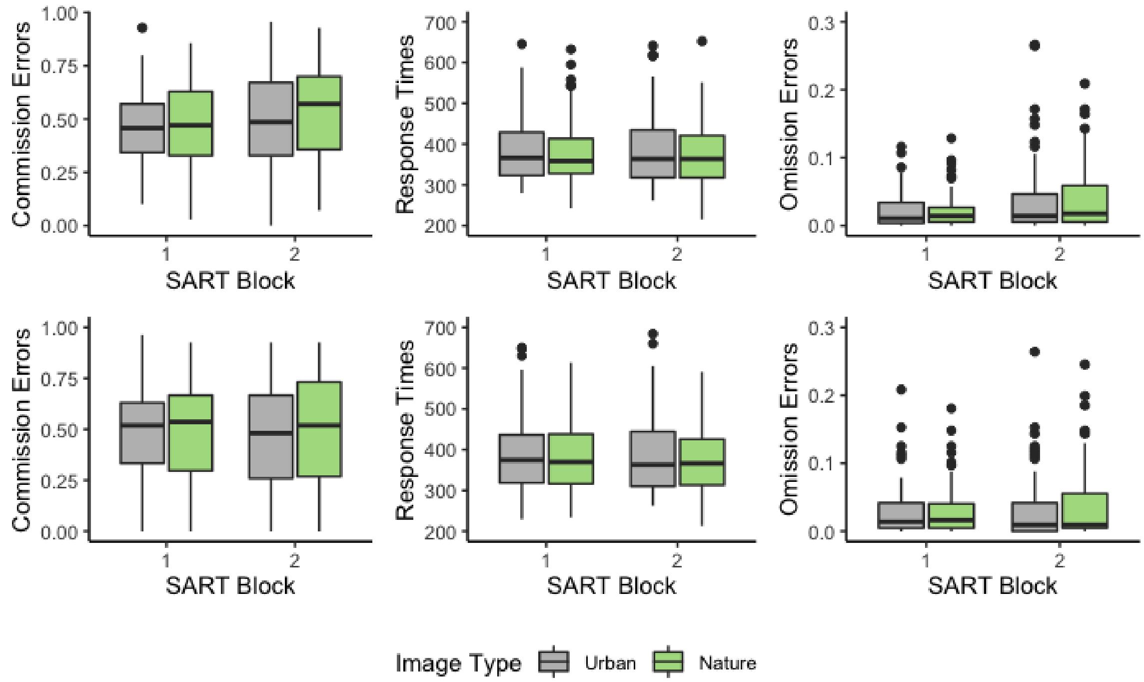

Figure 2. Box and whisker plots (boxplots) representing data from each measure in both analyses. Data from all SART trials are portrayed in the top row, while data from the reduced SART trials are portrayed in the bottom row. From left to right, the graphs show data from commission errors, response times, and omission errors. Data from the urban condition are portrayed in grey, while data from the nature condition are portrayed in green. No significant interactions were found for any of the measures.

Discussion

Experiment 2 replicated the results found in Experiment 1, even after addressing concerns related to image pixilation and aesthetic confounds. While time-on-task effects were evident via the main effects of SART block found for commission errors and omission errors, we again found no evidence for interactions between SART Block and Image Type for *any* of our measures. In addition, Bayes analyses suggested strong evidence for the null hypotheses. Therefore, this experiment provided a second demonstration that images of natural environments *are not* more restorative for SART performance than images of urban environments.

General Discussion

The present experiments were conducted in an effort to test the replicability of a commonly cited conclusion by Berto (2005): exposure to images of natural environments restores attention-related performance more effectively than does exposure to images of urban environments. In two experiments, participants completed the SART, then were exposed to images of either nature or urban settings, and then completed the SART once more. Had a significant interaction occurred in at least one of the measures, it could be reasoned that the environment that participants were exposed to influenced their performance on the SART. However, in *both* experiments, we found no evidence for such this interaction. Bayes analyses also suggested strong evidence for the null hypothesis instead. Thus, our results from two well-powered replication attempts are not consistent with the conclusion that images of natural environments are more restorative for attention than images of urban environments.

There were two important differences between the present studies and the study reported by Berto (2005) that might account for the differences in outcomes across studies. The first major difference across studies concerns the way the stimuli were selected. Our approach differed from Berto's such that our stimuli were selected based on ratings of typicality rather than restorative potential, as we wanted to avoid confounding environment type with predicted restorativeness level. We also deviated from the design of Berto (2005) in the number of trials included in the SART task. Each iteration of our SART consisted of 630 trials, which is substantially more the 240 trials included in Berto's (2005) original study. We increased the number of trials in an attempt to more thoroughly deplete attention during the pre-exposure SART, and to obtain more observations during the post-exposure

SART. It is possible that any restoration that might have occurred was only short-lived and could not be detected once performance was averaged across all of the trials of the post-exposure SART. This concern is alleviated by the fact that we did not find any difference in performance across nature and urban conditions when only the first 243 trials of the post-exposure SART were compared to the last 243 trials of the pre-exposure SART (matching Berto's analysis).

One possible reason why we failed to replicate the results reported by Berto (2005) is that Berto's initial findings might be unreliable owing to the relatively small sample used in the study. The issue of underpowered studies within the attention restoration literature is not limited to studies using the SART, nor is it limited to studies using virtual exposure. Reviews of the attention restoration literature reveal that studies using various measures of attention and cognition (other than the SART) to index attention restoration also often rely on inadequate sample sizes; this is true not only for initial reports of a particular effect, but also for subsequent replications (for meta-analyses, see Ohly et al., 2016 or Stevenson, Schilhab, & Bentsen, 2018). Specifically, Ohly et al. (2016) note that many of the studies included in their meta-analyses were low to moderate quality. That is, limited studies had been conducted, and many of those that are available lack the power needed for obtaining reliable results. As an example, we can consider a rather impactful study reported by Berman, Jonides, and Kaplan (2008), which reported two studies of nature (compared to urban) exposure. In the first study, participants viewed real-world environments, while in the second, they viewed virtual environments. The real-world study, conducted via a within-participants design with 38 participants, showed improved performance on a backwards-digit span task following walks in a natural setting, but not an urban setting.

The second study found similar improvements performance on the attention network task (ANT) following exposure to natural (but not urban images); importantly, however, this within-participant study was conducted with only 12 participants. Studies attempting to replicate Berman and colleagues' (2008) virtual exposure study (Emfield & Neider, 2014; Gamble, Howard, & Howard, 2014; Kuziek & Mathewson, 2017) and real-world study (Bratman, Daily, Levy, & Gross, 2015) have all been conducted as between-participant designs with 30 participants or less per condition. Each of these attempts failed to replicate the findings from Berman et al. (2008); however, these nulls are not particularly convincing due to the small sample sizes and consequent lack of power.

The most convincing null result finally came from Boggs (2018), who included 90 participants in each condition of a between-participant virtual-exposure design. The results are more generalizable than prior replication attempts because each condition in the study contained *multiple* images, rather than just *one* – a problem present in a number of prior studies (for virtual studies: Lee, Williams, Sargent, Williams, & Johnson, 2015; Sanguinetti, 2009; Valtchanov, 2010; van den Berg, Koole, van der Wulp, 2003; for real world studies: Berman, Kaplan, & Jonides, 2008; Bodin & Hartig, 2003; Bratman, Daily, Levy, & Gross, 2015; Hartig, Evans, Jamner, Davis, & Gärling, 2003; Johansson, Hartig, & Staats, 2011; Mayer, Franz, Bruehlman-Senecal, & Dolliver, 2009; Perkins, Seawright, & Ratwick, 2011; Shin, Shin, Yeoun, & Kim, 2011; Faber Taylor & Kuo, 2009). Use of only one stimulus for each environmental condition is problematic because researchers cannot determine whether their results were due to the selected environments' categorical membership (e.g., natural vs. urban) or some other features unique to those specific environments. To improve generalizability, future studies aiming to understand attention restoration from both images and real-world settings should ensure that participants are exposed to multiple settings for each environment type that they view.

Finally, it is important to note that while we failed to find an influence of environment (nature vs. urban) on performance in the SART following virtual exposures, this does not mean that null findings will also occur in well-powered examinations using other types of attention tasks or other methods of nature exposure (i.e., real-world exposure). In a meta-analysis, Stevenson and colleagues (2018) examined several types of tasks that may hold greater potential for capturing subtle improvements in performance following natural vs. urban exposure, such as the Stroop Task and the Trail Making Task – version B. However, we hasten to add again that many of the studies considered in that review were stated by the authors to be underpowered, had only small to moderate effect sizes, and demonstrated no attentional improvements when the groups within the studies started from equivalent baseline measures (Stevenson et al., 2018). Additionally, we think it is important to note that the "File Drawer Problem" (Rosenthal, 1979, p. 638) – a longstanding issue in psychological research – also makes it difficult to be certain of effect sizes that come from these meta-analyses, as there is no telling how many other studies resulted in null results but were never published. With this in mind, we hope that researchers continue to critically examine the claims within the attention restoration literature, as it is only through several iterations of well-powered and well-designed studies that we can really be confident that an effect exists.

References

Abraham, A., Sommerhalder, K., & Abel, T. (2010). Landscape and well-being: a scoping study on the health-promoting impact of outdoor environments. *International Journal of Public Health*, *55*(1), 59–69.

Atchley, R. A., Strayer, D. L., & Atchley, P. (2012). Creativity in the wild: Improving creative reasoning through immersion in natural settings. *PLOS ONE*, *7*(12), e51474.

Bakker, M., van Dijk, A., & Wicherts, J. M. (2012). The rules of the game called psychological science. *Perspectives on Psychological Science*, *7*(6), 543–554.

Berman, M. G., Jonides, J., & Kaplan, S. (2008). The cognitive benefits of interacting with nature. *Psychological Science*, *19*(12), 1207–1212.

Berto, R. (2005). Exposure to restorative environments helps restore attentional capacity. *Journal of Environmental Psychology*, *25*(3), 249–259.

Bodin, M., & Hartig, T. (2003). Does the outdoor environment matter for psychological restoration gained through running?. *Psychology of Sport and Exercise*, *4*(2), 141–153.

Bratman, G. N., Daily, G. C., Levy, B. J., & Gross, J. J. (2015). The benefits of nature experience: Improved affect and cognition. *Landscape and Urban Planning*, *138*, 41–50.

Bratman, G. N., Hamilton, J. P., & Daily, G. C. (2012). The impacts of nature experience on human cognitive function and mental health. *Annals of the New York Academy of Sciences*, *1249*(1), 118–136.

Carrus, G., Passiatore, Y., Pirchio, S., & Scopelliti, M. (2015). Contact with nature in educational settings might help cognitive functioning and promote positive social behaviour/El contacto con la naturaleza en los contextos educativos podría mejorar el funcionamiento cognitivo y fomentar el comportamiento social positivo. *Psyecology*, *6*(2), 191–212.

Cheyne, J. A., Solman, G. J., Carriere, J. S., & Smilek, D. (2009). Anatomy of an error: A bidirectional state model of task engagement/disengagement and attention-related errors. *Cognition*, *111*(1), 98–113.

Craig, C. M. (2016). *The green mind: Examining the intersection of executive and sustained attention with restorative environments* (Doctoral dissertation).

Craig, C., & Klein, M. I. (2012, September). Mental Demand, Asymmetries, and Tympanic Membrane Temperature (TMT). In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 56, No. 1, pp. 1529–1533). Sage CA: Los Angeles, CA: SAGE Publications.

Craig, C., Klein, M. I., Menon, C. V., & Rinaldo, S. B. (2015). Digital nature benefits typical individuals but not individuals with depressive symptoms. *Ecopsychology*, *7*(2), 53–58.

Dumke, R. (2014). *The effects of reflective prompts on attention in restorative and non-restorative environments*(Doctoral dissertation, Indiana State University).

Emfield, A. G., & Neider, M. B. (2014). Evaluating visual and auditory contributions to the cognitive restoration effect. *Frontiers in Psychology*, *5*(548), 1–11.

Faber Taylor, A., & Kuo, F. E. (2009). Children with attention deficits concentrate better after walk in the park. *Journal of Attention Disorders*, *12*(5), 402–409.

Gamble, K. R., Howard Jr, J. H., & Howard, D. V. (2014). Not just scenery: viewing nature pictures improves executive attention in older adults. *Experimental Aging Research*, *40*(5), 513–530.

Hartig, T., Evans, G. W., Jamner, L. D., Davis, D. S., & Gärling, T. (2003). Tracking restoration in natural and urban field settings. *Journal of Environmental Psychology*, *23*(2), 109–123.

Hartig, T., Korpela, K., Evans, G. W., & Gärling, T. (1997). A measure of restorative quality in environments. *Scandinavian Housing and Planning Research*, *14*(4), 175–194.

Head, J., & Helton, W. S. (2012). Natural scene stimuli and lapses of sustained attention. *Consciousness and Cognition*, *21*(4), 1617–1625.

Head, J., Russell, P. N., Dorahy, M. J., Neumann, E., & Helton, W. S. (2012). Text-speak processing and the sustained attention to response task. *Experimental Brain Research*, *216*(1), 103–111.

Helton, W. S., & Russell, P. N. (2011). Working memory load and the vigilance decrement. *Experimental Brain Research*, *212*(3), 429–437.

Helton, W. S., & Warm, J. S. (2008). Signal salience and the mindlessness theory of vigilance. *Acta Psychologica*, *129*(1), 18–25.

Herzog, T. R., Black, A. M., Fountaine, K. A., & Knotts, D. J. (1997). Reflection and attentional recovery as distinctive benefits of restorative environments. *Journal of Environmental Psychology*, *17*(2), 165–170.

James, W. (1890). *Principles of Psychology* (Vol. I). New York, NY: Holt and Company.

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). New York, NY: Oxford University Press.

Johansson, M., Hartig, T., & Staats, H. (2011). Psychological benefits of walking: Moderation by company and outdoor environment. *Applied Psychology: Health and Well-Being*, *3*(3), 261–280.

Kaplan, S., & Berman, M. G. (2010). Directed attention as a common resource for executive functioning and self-regulation. *Perspectives on Psychological Science*, *5*(1), 43–57.

Kaplan, S., & Kaplan, R. (1989). The experience of nature: A psychological perspective. New York: Cambridge University Press.

Kaplan, S. (1995). The restorative benefits of nature: Toward an integrative framework. *Journal of Environmental Psychology*, *15*(3), 169–182.

Kline, R. B. (1998). *Principles and practice of structural equation modeling*. New York: Guilford.

Kuziek, J., & Mathewson, K. E. (2017). Does viewing nature and urban environments change neuro-cognitive markers of attention?.

Lee, K. E., Williams, K. J., Sargent, L. D., Williams, N. S., & Johnson, K. A. (2015). 40-second green roof views sustain attention: The role of micro-breaks in attention restoration. *Journal of Environmental Psychology*, *42*, 182–189.

Mackworth, N. H. (1948). The breakdown of vigilance during prolonged visual search. *Quarterly Journal of Experimental Psychology*, *1*(1), 6–21.

Mayer, F. S., Frantz, C. M., Bruehlman-Senecal, E., & Dolliver, K. (2009). Why is nature beneficial? The role of connectedness to nature. *Environment and Behavior*, *41*(5), 607–643.

McCrackin, S. D., & Itier, R. J. (2018). Both fearful and happy expressions interact with gaze direction by 200 ms SOA to speed attention orienting. *Visual Cognition*, *26*(4), 231–252.

Nguyen, T., Neilson, B., & Klein, M. I. (2018, September). Greenery versus Aquatic: Cognitive Restoration of Different Components of Nature for Improvement in Environmental Design. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 62, No. 1, pp. 373–377). Sage CA: Los Angeles, CA: SAGE Publications.

Ohly, H., White, M. P., Wheeler, B. W., Bethel, A., Ukoumunne, O. C., Nikolaou, V., & Garside, R. (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.

Perkins, S., Searight, H. R., & Ratwik, S. (2011). Walking in a natural winter setting to relieve attention fatigue: A pilot study. *Psychology*, *2*(08), 777.

Robertson, I. H., Manly, T., Andrade, J., Baddeley, B. T., & Yiend, J. (1997). Oops!': performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia*, *35*(6), 747–758.

Rosenthal, R. (1979). The “File Drawer Problem” and tolerance for null results. *Psychological Bulletin, 86*(3), 638–641.

Sanguinetti, A. (2009). *Contact with Nature as a Motivating Operation Affecting Attention* (Doctoral dissertation, California State University, Stanislaus).

Seli, P., Jonker, T. R., Solman, G. J., Cheyne, J. A., & Smilek, D. (2013). A methodological note on evaluating performance in a sustained-attention-to-response task. *Behavior Research Methods*, *45*(2), 355–363.

Shin, W. S., Shin, C. S., Yeoun, P. S., & Kim, J. J. (2011). The influence of interaction with forest on cognitive function. *Scandinavian Journal of Forest Research*, *26*(6), 595–598.

Sonnentag, S., Venz, L., & Casper, A. (2017). Advances in recovery research: What have we learned? What should be done next?. *Journal of Occupational Health Psychology*, *22*(3), 365–380.

Stevenson, M. P., Schilhab, T., & Bentsen, P. (2018). Attention restoration theory II: A systematic review to clarify attention processes affected by exposure to natural environments. *Journal of Toxicology and Environmental Health, Part B*, *21*(4), 227–268.

Van den Berg, A. E., Maas, J., Verheij, R. A., & Groenewegen, P. P. (2010). Green space as a buffer between stressful life events and health. *Social Science & Medicine*, *70*(8), 1203–1210.

Valtchanov, D. (2010). *Physiological and affective responses to immersion in virtual reality: Effects of nature and urban settings* (Master's thesis, University of Waterloo).

Van den Berg, A. E., Joy, Y., & Koole, S. (2016). Why viewing nature is more fascinating and restorative than viewing buildings: A closer look at perceived complexity. *Urban Forestry & Urban Greening*, *20*, 397–401.

Van den Berg, A. E., Koole, S. L., & Van der Wulp, N. Y. (2003). Environmental preference and restoration:(How) are they related?. *Journal of Environmental Psychology*, *23*(2), 135–146.

Wicherts, J. M., Veldkamp, C. L. S., Augusteijn, H. E. M., Bakker, M., Van Aert, R. C.

M., & Van Assen, M. A. L. M. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: A checklist to avoid p-hacking. *Frontiers in Psychology*, *7*, 1832. *doi: 10.3389/fpsyg.2016.01832*

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: the SHINE toolbox. *Behavior Research Methods*, *42*(3), 671–684.