THE UNIVERSITY OF CHICAGO

THE ROLE OF PREFERENCE IN THE AFFECTIVE AND COGNITIVE BENEFITS OF NATURE

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE SOCIAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF PSYCHOLOGY

BY
KIMBERLY LEWIS MEIDENBAUER

CHICAGO, ILLINOIS
JUNE 2020

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

Listing out and thanking all the people who have helped me on my journey to this point would probably double the length of this already long dissertation. While I cannot name every person who has provided support for me intellectually, socially, physically (as I am quite prone to injury!), or emotionally along the way, know that I wholeheartedly appreciate and am grateful for all of you.

First and foremost, I would like to thank Marc, who took a chance on me, advised me in my research for the past several years, encouraged me to think critically, and provided me with the space and resources to grow as a scientist. I am a better researcher because of my time working under you in the Environmental Neuroscience Lab, and I'm immensely grateful for all that you have taught me. Whether it was a statistical technique I'd never heard of, tips on how to apply for academic jobs, or simply how to fit 12 pyrex containers of food into a messenger bag, you have always been generous in sharing your knowledge with me and the rest of the lab. I cannot thank you enough.

I would also like to thank Alex, Ed, and Chris for being on my committee and dedicating time and energy to my dissertation during a global pandemic. I'd like to give an extra shout out to Chris who advised me as an undergrad and made my first ever exposure to psychology and neuroscience research an overwhelmingly positive one. I might not have continued on to get my PhD if I hadn't loved being an RA in your lab as much as I did.

Every experience I had working in a research lab taught me something valuable and made me a better scientist through exposure to different methods, ideas, and people that I would not have had the opportunity to learn from otherwise. I am grateful to all the PIs, graduate students, post-docs, and RAs that I've encountered in my now over 10 years of working in psychological research. Special thanks to my former lab mates who now live hundreds of miles away (lookin' at you Lauren, Jali, and Cliff) but still keep me going with their friendship and support from a distance.

A huge thank you to the rest of the ENL family. Omid and Elliot, though you guys started out as just fellow cohort members, I'm glad I got the opportunity to work with and learn from you both. I'll never forget Elliot's "poetry" skills in year 1, and I don't think you will either, Omid Kard-an. Kyoung, I might not have guessed on day one of joining the lab that you'd end up being such a good friend and mentor, but I'm glad it worked out that way. Even though I now have Permanent head Damage, I don't think I'll ever forget the difference between good coffee and trash coffee. I sincerely thank you for shaming me into stepping up my coffee game. Carlos, Kate, Gaby, Andrew, and all the past and present grad students, lab managers, and RAs in the lab, you all have provided me with support in one way or another and made coming into lab fun on days where it might have been frustrating or boring. I'm excited to share the same physical space with you all again soon and stick my popsicle stick head above the whiteboard.

Nick, Julianne, Casey, and Becky, I'm grateful to have started this crazy journey with you guys. The 2014 cohort might be the smallest cohort in recent history, but also maybe the best cohort (in my unbiased opinion). Also, thank you to my Zumba family. Teaching classes and dancing with you all brought me so much joy and was the best source of endorphins in insane times. Val, thank you for being my friend for 27 years and offering to help me with design support for my defense powerpoint. I probably should have taken you up on that. Becca and Luiza, you are two of my favorite humans and role models, and I've benefitted from your friendship for literal decades. Thank you for all you've done for me.

Mom and dad, I learned how to be hardworking, compassionate, and curious because of the example you set for me. I'm proud to be your daughter and I cannot wait to celebrate being a doctor once we can see each other again. I love and appreciate you both infinitely. To my gran, who passed away a year and a half ago, I wish you could be here to celebrate with me too. You were one of the only non-academics I knew who actually enjoyed listening to me talk about my research. If there is a genetic component underlying the desire to learn about psychology and human behavior, I definitely got it from you. Russ and Debbie, I could not have asked for a better family to become a part of. I am so grateful for all you have done for me over the past 12 years.

And last but not least, thank you to my cats, Sydney and Uno… just kidding! (though they do deserve an honorable mention for being super cute work-from-home buddies while I was writing up this up). Mike, you are everything I could have asked for in a partner. Moving from Wisconsin to Chicago to pursue my PhD was one of the most difficult things I ever did. No words can express how grateful I am that you supported me and my ambitions for the past 12 years(!), from when we were undergrads at UWM, to the brutal 3 years we did long distance, to the best years of my life living with you in Hyde Park. I owe a major part of my finishing this dissertation while fighting off covid-19 to you, and there is nobody on the planet I would rather be quarantining with in a 1-bedroom apartment for months. You are the best person I know. I love you. And thank you for having a distinctive last name that I could take for SEO purposes.

ABSTRACT

Exposure to natural environments over urban ones has beneficial effects on human psychological functioning, particularly in improving affective state and providing cognitive restoration. Because nature is so highly preferred, it has been difficult to disentangle what effects result from nature itself and what are simply due to exposure to a highly preferred stimulus. In this dissertation, I examine whether the aesthetic preference for nature accounts for any, some, or all of its cognitive and affective benefits. In Chapter 1, I investigate whether the highly documented nature preferences observed in adults are also found in 4- to 11-year-old children. I found that, compared to adults, children show stronger preferences for urban environments, though this urban preference lessens with age. Further, though children do not like nature as much as adults, children with more nearby nature had lower parent-reported inattentiveness, suggesting this benefit is not dependent upon liking nature. In Chapter 2, I tested whether natural environments have some additional positive effect on mood, above and beyond what can be attributed to preference. The results of this set of studies supported an overwhelmingly preference-based account of nature's short-term affective benefits. In Chapters 3 and 4, I examine whether cognitive restoration is elicited by images or videos of nature when compared to equally preferred urban environments. Unfortunately, these studies did not provide evidence for either a strong role of preference or environment type in predicting cognitive restoration. To examine whether neural indices of cognitive restoration could be found in natural over urban environments, Chapter 4 also used functional near-infrared spectroscopy (fNIRS) to measure changes in prefrontal and parietal cortical activation during cognitive tasks and virtual environmental exposure. While no effects of environment type were found in fNIRS activity during the post-video cognitive task, fNIRS activity did, overall, reliably map on to cognitive load and performance. Taken together, these studies suggest that nature preferences are not universal across ages, and that preference matters more for the affective than for the cognitive benefits of nature exposure.

GENERAL INTRODUCTION

The documented effects of interactions with natural environments are robust and span many domains. Visible nearby nature is related to improved physical health (Kardan et al., 2015; Parsons, Tassinary, Ulrich, Hebl, & Grossman-Alexander, 1998; Ulrich et al., 1991) and emotional well-being (MacKerron & Mourato, 2013; Ulrich, 1983; Vujcic et al., 2017). Brief walks in natural environments or exposure to nature photos or sounds have improved performance on cognitive tasks (Berman, Jonides, & Kaplan, 2008; Berto, 2005; Van Hedger et al., 2018) and have been linked to prosocial behaviors (Guéguen & Stefan, 2014; Zhang, Piff, Iyer, Koleva, & Keltner, 2014).

Several theories attempt to explain why nature is beneficial. The most widely used frameworks are that of Attention Restoration Theory (Kaplan, 1995), Stress Reduction Theory (Ulrich et al., 1991), the Perceptual Fluency Account (Joye & van den Berg, 2011), and Biophilia (Kellert & Wilson, 1995). Biophilia proposes that humans have evolved in natural settings and thus have an innate affinity for biological things rather than those that are built. Biophilia cites this innate preference for nature as an explanation for why nature has positive health and affective benefits. Stress Reduction Theory (SRT) also uses an evolutionary account of nature preferences and puts forth that an improvement in mood and a reduction in stress are the cause for the observed cognitive benefits. Perceptual Fluency (PFA) states that certain features of natural environments, such as fractality or self-similarity, are more easily or fluently processed by our brains. This ease of processing is less cognitively demanding, which improves mood, which in turn, frees up attention resources. Thus, in PFA and SRT, nature primarily improves state affect and this has a downstream effect on cognitive performance, suggesting that the cognitive benefits arise from the affective ones. Attention Restoration Theory (ART) focuses on the cognitive benefits, suggesting that natural environments (compared to urban ones) place minimal demands on effortful, directed attention and have features that gently capture bottom-up, involuntary attention. Because endogenous directed attention is thought to be a finite resource that can be depleted, a natural environment that is engaging but does not require effortful attention is well suited to restore attention and improve cognition. It is noteworthy that these theories do not all converge on whether there is a significant role of preference in emotional or cognitive changes obtained through nature interactions, and there is disagreement as to whether the emotional and cognitive benefits are connected.

Because nature is so highly preferred (Kaplan & Kaplan, 1989), the effects of nature itself are typically confounded with the effects of exposure to something that an individual likes. Therefore, the goal of this dissertation was to identify the role of preference in the cognitive and affective benefits of nature. Four major studies were conducted to answer this question. Study 1 (Chapter 1: The Gradual Development of the Preference for Natural Environments) examines the question of whether the highly documented nature preferences observed in adults are also found in young children. This work tests the universality of nature preferences by studying a previously unexamined subset of the population, and examines how children's preferences relate to age, nature exposure, and parental preferences. This study also tests whether children's environmental preferences relate to any of the cognitive or affective benefits of nature exposure.

Study 2 (Chapter 2: The Affective Benefits of Nature Exposure: What's Nature Got to Do with It?) tests whether nature improves state affect when preference is accounted for. A set of studies examining the mood-boosting effects of nature images and equally preferred images of other categories were conducted to answer whether nature scenes contribute some unique benefit to improved affect above and beyond it being a highly preferred environment.

Study 3 (Chapter 3: Preference and the Cognitive Benefits of Nature Exposure) tests the effects of preference (high and low) and of environment type (nature and urban) in predicting change in performance in a cognitively demanding dual n-back task. The relationships between preference, affect change, and dual n-back accuracy change are examined to see whether preference and/or affect play a role in cognitive restoration.

Study 4 (Chapter 4: Neural Correlates of Cognitive Effort and Restoration from Equally Preferred Nature and Urban Environments: an fNIRS Investigation) examines whether cognitive restoration, measured by behavior or neural activation, can be elicited by preference-equated nature and urban videos. This study had the additional aim of performing a robust validation of functional near-infrared spectroscopy (fNIRS) as a tool for future neuroimaging work in realistic natural and urban environments.

CHAPTER 1: The Gradual Development of the Preference for Natural Environments

A version of this chapter was published in the Journal of Environmental Psychology, *65, 101328 (2019)*.

Co-authors: Cecilia U. D. Stenfors, Jaime Young, Elliot A. Layden, Kathryn E. Schertz, Omid Kardan, Jean Decety, & Marc G. Berman

Abstract

Adults demonstrate aesthetic preferences for natural environments over urban ones. This preference has influenced theories like Biophilia to explain why nature is beneficial. While both adults and children show cognitive and affective benefits after nature exposure, it is unknown whether children demonstrate nature preferences. In the current study, 4-to-11-year-old children and their parents rated their preferences for images of nature and urban scenes. Parents' preferences matched those of a normative adult sample. However, children demonstrated robust preferences for urban over natural environments, and those urban preferences significantly decreased with age. Nature exposure around the home and nature-related activities, as reported by parents, did not predict children's preferences. Children with more nearby nature, however, had lower reported inattentiveness, but interestingly, this was unrelated to children's preferences for nature. These results provide an important step into future research on the role of preference in how children and adults benefit from nature.

1. Introduction

Humans have an incredibly strong preference for natural environments. People's preferences for natural environments over urban environments have been extensively documented (Ibarra et al., 2017; R. Kaplan et al., 1989; van den Berg et al., 2007). In fact, nature preferences are so strong that researchers have found that the distributions of adults' preference ratings for many kinds of nature and urban photos barely overlap (Berman, Hout, et al., 2014; S. Kaplan et al., 1972; Kardan, Demiralp, et al., 2015). These preferences are also observed with real world dependent measures such as property values, as home buyers pay a premium for homes near nature (Crompton, 2001) and vacationers pay extra for rooms with natural sea views (Fleischer, 2012). These nature preferences are fairly universal as adults' preferences for nature are found cross culturally (Kaplan & Yang, 1990; Ulrich, 1993). In fact, nature preferences are so strong that even within completely built structures, building interiors and exteriors that are perceived to be more natural are rated higher on aesthetic preference (Coburn et al., 2019).

Some theoretical accounts, such as Biophilia (Kellert & Wilson, 1995) and Stress Reduction Theory (SRT) (Ulrich et al., 1991), suggest that these preferences for nature arise from humans evolving in natural environments. Though there is disagreement on how this might occur—whether this innate affinity is genetically programmed or works through a form of biologically-prepared learning—a common explanation for why nature is preferred is that only a tiny fraction of evolutionary history has occurred within our current urban environments, and the remainder in more natural environments.

In recent years, theories that suggest nature preferences arise from humans' evolutionary history have been called into question. Some critics of Biophilia and SRT emphasize the lack of strong empirical data to support such evolutionary accounts (Joye & De Block, 2011; Y. Joye & van den Berg, 2018), and propose that positive feelings of restoration in nature are a byproduct of the ease with which humans are able to process features of the natural world (Joye et al., 2016). This idea, referred to as the Perceptual Fluency Account (PFA), suggests that it is this lessened cognitive processing of stimuli in the natural world which underlies many of the observed benefits of nature (Joye & van den Berg, 2011). While it can be argued why humans prefer nature, the persistent preferences for natural environments found in adults are well-established.

Though adults' preferences are well researched, very few studies have attempted to examine whether these preferences exist in children. One study conducted on this topic involved asking 9- to 12-year-old children to make a map or drawing of their favorite places. They found that the vast majority of the illustrations were outdoors, featuring lawns, playgrounds, and parks (Moore, 1986), suggesting that older children may display an affinity for nature that mirrors adult preferences. Those results, while highly informative, cannot speak to preferences for children under 9 years of age. In addition, from those results it is difficult to determine whether the children's preferences for natural spaces were indeed an indication of an affinity for nature, or a more general preference for areas associated with recreation. Other related research has addressed children's environmental reasoning, and found that elementary school children believe nature has intrinsic value and that preserving nature is important (Kahn, 1997). Children also seem to show visual preferences for certain types of natural environments (e.g., savannas and evergreen forests) over others (e.g., rainforests and deserts)(Balling & Falk, 1982). Overall, while there is some preliminary evidence that children may appreciate and enjoy nature, no empirical work has directly examined natural versus urban preferences in children across a broad age range and across a varied stimulus set.

Knowing whether children do indeed prefer natural environments can inform two outstanding questions related to environmental preferences. First, examining children's preferences and their consistency (or lack thereof) during development can inform current debates about whether there is an innateness to nature preferences. This study may speak to whether nature preferences are genetically programmed and result from our evolutionary history (i.e. Kellert, 1993) or are learned through experience. Evidence for the latter hypothesis comes from recent research demonstrating that certain sounds and sound features are not preferred when the source of the sound is completely artificial (e.g., computer generated sounds) or obscured through scrambling, but become preferred when the sound is thought to be generated from a natural source, e.g., bird song (Van Hedger et al., 2019). Thus, there may not be anything intrinsically preferred about the features (visual, acoustic, tactile, etc.) of natural environments, and humans' preferences for nature may be heavily influenced by learned expectations.

Second, the present study can inform current theories about the role of nature preferences in the observed benefits of interacting with natural environments. In adults, these benefits include improvements in mood(Bratman, Hamilton, et al., 2015; Hartig et al., 2003), positive physical health outcomes (Kardan, Gozdyra, et al., 2015; Nielsen & Hansen, 2007), and improved executive functioning (Berman et al., 2008, 2012; Berto, 2005; Van Hedger et al., 2018). Though generally not as well studied, many of these same positive effects of nature are also found in children. Nature interventions have been shown to 1) decrease children's levels of stress, 2) reduce attention deficit hyperactivity disorder (ADHD) symptoms and 3) decrease levels of inattentiveness (Amoly et al., 2014; Corraliza et al., 2012; Faber Taylor & Kuo, 2009; Mårtensson et al., 2009; Wells & Evans, 2003).

Importantly, theories for how and why nature provides psychological benefits differ in the extent to which they believe nature preferences are required or relevant. Proponents of SRT argue that this preference for nature can explain some of the reduced stress and improved mood after interactions with natural environments, and this, in turn, can improve cognitive performance. Conversely, Perceptual Fluency theorists suggests that a positive response to and preference for nature is a consequence of the fluent processing of natural features rather than a cause of restoration (Joye & van den Berg, 2011). Attention Restoration Theory (ART) (Kaplan, 1995; Kaplan & Berman, 2010) does not require that natural environments be preferred to obtain the cognitive benefits. ART proposes that the cognitive improvements seen after nature interactions are not driven by mood or preference, but rather result from resting top-down directed-attention resources. Given that nature has been found salubrious for psychological functioning for both adults and children, if children do not prefer natural environments to the same extent as adults, this can inform extant theoretical accounts about the role of preference in obtaining cognitive benefits. Previous research on adults has suggested that preference and subsequent mood improvements may not be necessary to obtain the cognitive benefits (Berman et al., 2008; Stenfors et al., 2019), but results from this study would test whether preference is an important factor to obtain cognitive benefits in children.

The primary goal of the current study was to examine whether this preference for natural environments exists in children, and whether children's preferences are consistent across development. To test whether children overall prefer natural or urban environments, we compared adults' and children's preferences for environments that varied on a spectrum of naturalness and preference. As adults' preference ratings for nature images are usually much higher than those for urban images, aesthetic preference and environment type are almost always confounded in studies using nature and urban stimuli (see Appendix A). Therefore, the current study used nature and urban images that were equated on aesthetic preference from a separate study, to unconfound aesthetics (high vs. low preference) and environment type (nature vs. urban). This is critical to ensure that children's environmental preferences are not simply a result of a difference in general aesthetic preferences that have nothing to do with natural or urban environments per se. In other words, if preference and environment type are confounded, as they are in most studies because adults prefer nature, it would be impossible to know if any child preference differences were due to liking nature less or due to a non-specific difference in preference that has nothing to do with nature or urban environments. In our experiment, we can separate out these two possibilities.

If children do not show the same preference for nature as adults, this would suggest that there is not an innate, present-from-birth affinity for natural environments. This result would also imply that children's nature preferences may not be necessary to reap the cognitive benefits of interacting with the natural world. Similarly, if the preference for nature is due to learning, children's preferences for natural environments may develop over time. If children's preferences look very similar to those of adults', this would be a compelling case for a truly innate nature preference, consistent with a strong interpretation of the Biophilia Hypothesis (Kellert, 1993).

We also examined the developmental trajectory of environmental preferences to test the question of whether an affinity for nature is learned. If preferences change over time to look more similar to those of adults or increasingly relate to nature exposure with age, this would be evidence for a learned account. To directly test this question, analyses were conducted investigating children's preferences as a function of the naturalness of their home, school, and play environments, as well as parental preferences, and the interaction of these variables with age.

A final aim was to directly examine whether some of the observed cognitive benefits of nature related to nature preferences in the current sample. To test this, measures of emotional, social, and cognitive functioning (using the Strengths & Difficulties Questionnaire) were collected and related to the amount of nature exposure and environmental preferences of the children. Overall, the current study sought to examine 1) whether children have environmental preferences that match those of adults, 2) whether these preferences are stable or change over time, 3) what factors may relate to individual differences and the development of environmental preferences, and 4) whether preferences relate to any observed cognitive/social/emotional benefits of nature in our sample. In addition to the fundamental importance of systematically testing nature preferences in a broad age-range of children and across a wide range of environments, this study will provide novel insights into current theories of nature preferences and the potential mechanisms of nature's psychological benefits.

2. Method

2.1 Participants

2.1.1. Sample Information

Data from 251 children and 187 parents or guardians were collected. Twelve children were excluded from analysis due experimenter note that the child had clear difficulty understanding the task or failed to complete all trials of the task. The final sample included 239 children between the ages of 4 and 11 years and 182 adults (162 parents and 16 grandparents/other non-parent guardians). Gender split of the children in the final sample was relatively even (112 male, 127 female). Of the 239 usable children, 61 were siblings of another child participant. All study procedures were approved by the University of Chicago Institutional Review Board.

2.1.2. Sample Size and Age Range

The lower limit of the age range was selected based on a short pilot study previously conducted with 3- to 6-year-olds, which showed that children under 4 years of age had a very difficult time understanding the task. Our goal was to collect usable data from at least twenty children per one-year age bin, and we stopped data collection when we reached this goal. This number was determined from a quasi-power analysis using results from previous stimulus validation studies in our lab.

These studies have found that obtaining preference ratings from about twenty adult participants is sufficient to gain reliable estimates of image preference (Kotabe et al., 2017). Though the extent to which children would be similarly consistent in evaluating the images in this experiment were unknown, all planned analyses were to be conducted using the full sample of children from 4 to 11 years old, rather than analyzing within a certain year of age. As such, the problem of insufficient statistical power did not seem likely with a minimum of 20 children per year of age. Additional sample information by age bin can be found in **Table 1.1**.

Table 1.1. Child participant breakdown by age and gender

Age Bin	**N (% of total)**	**# F (# M)**
4 years	21 (8.8%)	11 (10)
5 years	29 (12.1%)	16 (13)
6 years	42 (17.6%)	16 (26)
7 years	34 (14.2%)	15 (19)
8 years	30 (12.6%)	18 (12)
9 years	27 (11.3%)	15 (12)
10 years	30 (12.6%)	19 (11)
11 years	26 (10.9%)	17 (9)
Total	**239**	**127 (112)**

2.1.3. Participant Geography and Nature Exposure

As much of our recruitment occurred in a museum in Chicago which draws many tourists, the children in our study came from incredibly varied geographic locations. Of those who reported location either via zip code or city name (N = 213), 42 children came from Chicago, 52 from other parts of Illinois, 108 from other places in the US (across 23 other states), and 11 from other English-speaking countries (UK, Canada, Switzerland, Australia, and New Zealand).

Additionally, child participants came from locations that varied greatly on the extent to which they were urban or rural. Using National Land Cover Database data taken from US zip code, percentage of developed land was estimated for each zip code and used as a proxy for more versus less urbanized zip codes. A histogram of this data for all children who provided a zip code is found in **Figure 1.1**. The values for percentage developed range from 3% to 80%, where zip codes with 3% developed land are primarily located in small, mostly rural towns (populations of < 2,000 people) and those with over 60% are primarily zip codes in major US cities.

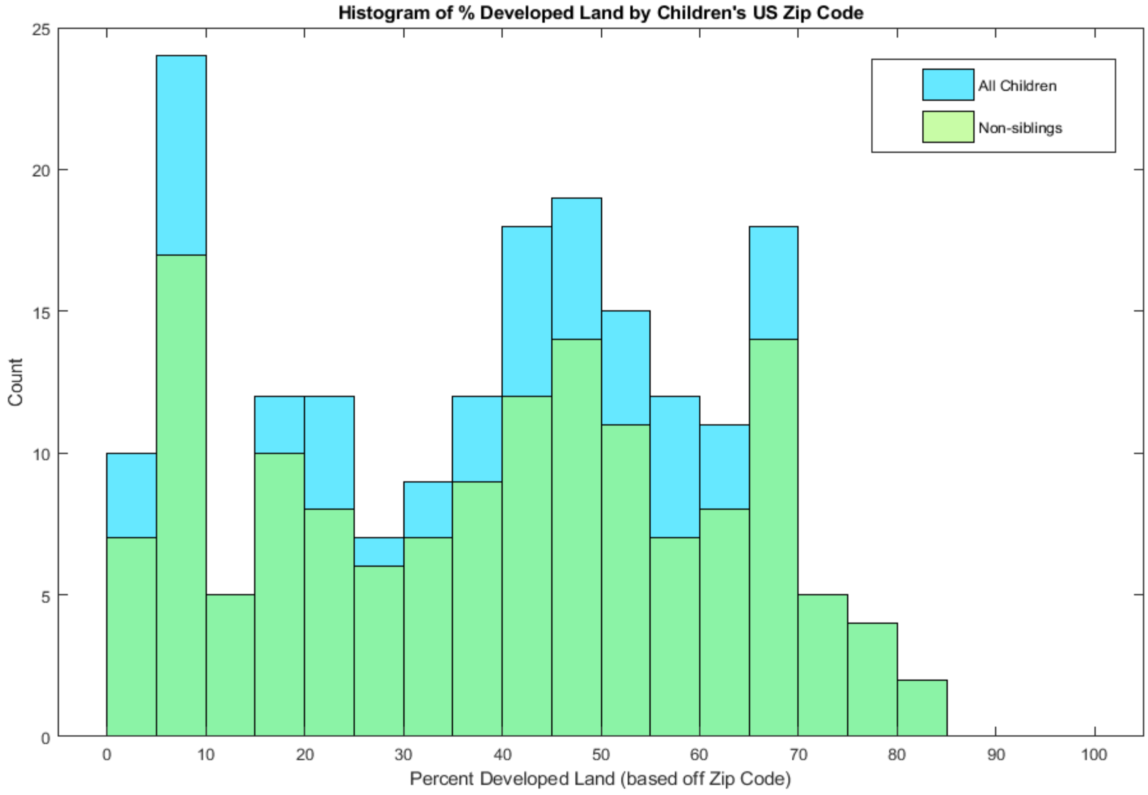


Figure 1.1. Level of urbanization for children's homes based on US zip code Percentage of developed land is used as a proxy of urbanization, where lower values indicate less urbanized/developed zip codes and higher percentages indicate more urbanized/developed land. In blue are all children in the sample with US zip code information. In green, family-level data are presented which exclude redundancy of information due to possible siblings.

2.1.4. Siblings in the Sample

A number of children in our study were the sibling of another child in the study, leading to non-independence of these data points. Consequently, we also conducted analyses on a subset of our sample which did not include siblings. To create this subset, we pseudo-randomly selected one child out of all instances of siblings. We attempted to keep at least 22 children in each age bin by choosing more siblings in the outer age bins (which were typically sparser than the middle ages) and filling in other gaps as needed. **Table 1.2** displays the breakdown of children by age included those analyses: i) without any siblings (all included), ii) in the sibling subset (1 child chosen for age balance), iii) the new non-correlated sample (non-siblings + sibling subset), and iv) the original total sample (including siblings).

Table 1.2. Breakdown of all children and independent (non-correlated) subset

Age	i) No siblings	ii) Sibling subset	iii) Non-correlated sample total	iv) Original sample
4 y	13	7	20	21
5 y	20	2	22	29
6 y	24	0	24	42
7 y	15	7	22	34
8 y	11	11	22	30
9 y	11	11	22	27
10 y	13	9	22	30
11 y	12	10	22	26
All	119	57	176	239

2.2. Procedure

2.2.1. Instructions

Data collection occurred in lab and at a nearby museum. In the museum, experimenters directly approached families that appeared to have children in the correct age range to invite them to participate in a short research study. In both cases, parents (or guardians) provided informed consent for their child's participation before any additional study procedures occurred. Once parental consent and child assent were obtained, the child went through the picture sorting task procedure with an experimenter while another experimenter ran through the same task with the parent. The instructions for the task were as follows: "You are about to see sets of four pictures, and you will be asked to put them in order based on how much you like them. On the one end you will see a frowny face, and on the other you will see a smiley face. I want you to move these pictures around so that the pictures are in order of the one you like the least by the frowny face to the one you like the most by the smiley face. When you've put the photos in order of your least to your most favorite, you can press the green button to go onto the next set of pictures." All child participants had these basic instructions explained to them (and additional information and clarification added with 4- to 6-year old children), then completed four practice trials with the experimenter where they were asked to sort images of children's bedrooms before continuing onto the real task. To ensure that children understood the task, the researchers took the children through these practice trials very deliberately and carefully, asking children to verbally indicate their preference for the images as they moved them along the frowny-to-happy-face scale. Child participants who struggled with comprehending the task were still run through full procedures, but their lack of understanding was noted, and they were subsequently excluded from analysis. Adults were provided with the same general instructions but did not complete the practice trials.

2.2.2. Stimuli Selection: Ratings from Validation Study

The specific stimuli used were taken from an image set which was rated on several attributes (including aesthetic preference and naturalness) in a previous validation study. In this validation study, adult participants rated a set of over 300 nature and urban images on a 1-7 Likert scale (1 = strongly dislike to 7 = strongly like). The preference ratings from this validation study were used to select the particular stimuli for the current study. Our goal was to find sets of nature and urban images which were rated very similarly on aesthetic preference to ensure that we'd be able to examine environmental preferences in children and any observed effects would not be attributable simply to differences in aesthetics. Full sized versions of all stimuli can be found here: https://osf.io/axn9q/?view_only=fa88e665f5a74885bd857d79f5a7ce4a, and a detailed explanation of the stimuli ratings, rationale for specific stimuli selection, and smaller versions of all images used can be found in Appendix A.

2.2.3. Stimulus Conditions

There were two different image sets used in this experiment (parents and children always completed the task with the same image set). In each set, there were 6 categories of images based on previous normative adult evaluations of their aesthetic value from the validation study described above. These categories were high aesthetic value nature (two images), low aesthetic value nature (two images), high aesthetic value urban (two images), low aesthetic value urban (two images), very high aesthetic value nature (one image), and very low aesthetic value urban (one image). Images in the high aesthetic value nature and high aesthetic value urban categories were matched on preference, as were those in low aesthetic value nature and urban (see **Table 1.3** for ratings of the current study stimuli from the validation study). The unmatched images (very high aesthetic value nature and very low aesthetic value urban) were included based on research in our lab that finds these images reliably elicit more extreme ratings in an adult sample. We have been unable to find sufficiently highly preferred urban images to match the very high aesthetic value nature (and nature to match the very low aesthetic value urban) which inevitably leads to a design in which conditions are not completely crossed. However, we chose to include these images as an additional, separate test of whether children's preferences map onto the preferences we see in adults.

Table 1.3. Average image aesthetic value pre-ratings
Ratings on a 1-7 scale (1 = strongly dislike, 7 = strongly like) for images in each picture set. These ratings were gathered from a separate validation study with a normative adult sample. The images in each picture set were chosen with the goal of ensuring that the nature and urban images in the same aesthetic value category (i.e. High Aesthetic Value) were very closely matched on preference.

	Very High Aesthetic Value Image	High Aesthetic Value Image #1	High Aesthetic Value Image #2	Low Aesthetic Value Image #1	Low Aesthetic Value Image #2	Very Low Aesthetic Value Image
Picture Set 1						
Nature	6.31	5.30	5.12	3.28	3.12	
Urban		5.29	5.11	3.28	3.06	2.09
Picture Set 2						
Nature	6.19	5.02	4.86	3.30	3.22	
Urban		5.04	4.88	3.30	3.22	1.77

2.2.4. Task

The task was completed on a touch-screen tablet, which allowed participants to drag the images left and right to put them in the preferred order. Using this comparison approach rather than a Likert scale rating on single images was decided based on data from a pilot study. This pilot study showed that the use of Likert-type response scales on singular images for children in the lower end of our age range tended to result in a more binary decision (choosing the anchors of happy vs. frowny face) rather than along a continuum. The task included 10 trials where four images were shown at a time. The presentation of images was randomized across trials and across starting positions within a trial. Because of this randomization, any set of four images from the 10 image set could appear in a trial, but the task used an algorithm to ensure that each of the 10 images were compared to every other image in the set at least once (see **Figure 1.2** for a display of the task).

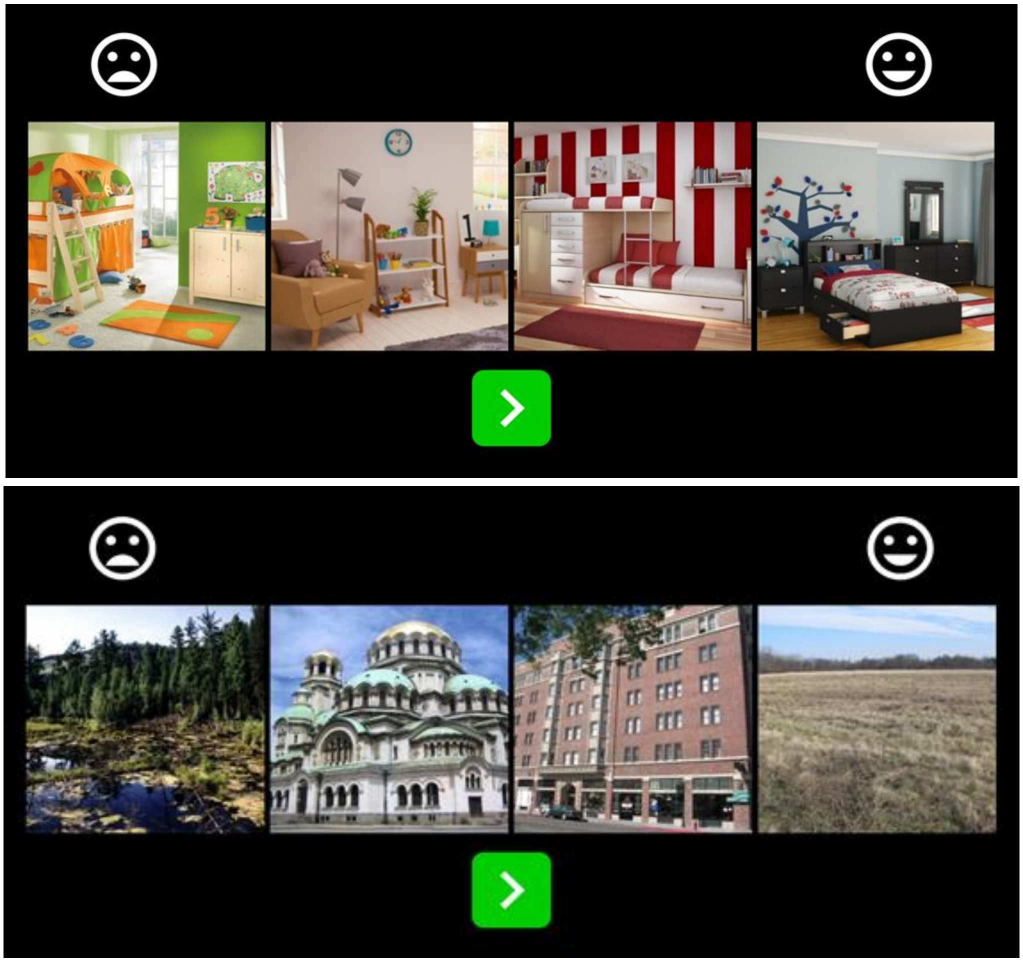


Figure 1.2. Task Design
Upper panel depicts a trial from the practice rounds. Lower panel depicts a sample trial from the actual experiment, which contains images in the following conditions (from left to right): high aesthetic value nature, high aesthetic value urban, low aesthetic value urban, low aesthetic value nature.

2.2.5. Additional Measures

Parents were also asked to complete a number of optional questionnaires about their child. Standard demographic measures were collected (birthdate, gender, ethnicity, household income, parental education) as well as zip code, which was used to calculate objective greenspace/land cover types from the 2011 National Land Cover Database (NLCD). From the NLCD data, we calculated amount of natural features nearby (summed coverage of values for water, deciduous forest, evergreen forest, mixed forest, shrub, grassland, pasture, cultivated land, woody wetlands, and herbaceous wetlands) as well as a ratio of low to high developed land, calculated by taking the amount of open-to-low developed land and dividing by the amount of medium-to-high developed land.

We also collected parent-reported natural features near the child's home (and school or daycare if applicable). This questionnaire (adapted from Tilt et al., 2007), asks about the presence of nine types of natural features within an approximate half mile distance from home or school. The total number of features near home (or the average of home and school/daycare if both included) was used to calculate parent-reported nearby natural features.

Additionally, the types of children's play environments outside of school and daycare hours and during school/daycare hours (if applicable) were assessed. The play environments questionnaire (adapted from Amoly et al., 2014; Faber Taylor & Kuo, 2011) asked parents to indicate the environments that their child typically plays in most of the time during a warm week in autumn or spring. Play in more natural versus more built spaces was calculated by taking the number of natural play environments (big trees and grass, open grass, "wild" places, waterfronts, deserts, and farms) and subtracting the number of built or indoor environments (deep indoors, indoors with windows, paved or built places, public indoors

Lastly, we asked parents to fill out the Strengths & Difficulties Questionnaire (Goodman et al., 2010), which assesses peer problems, conduct problems, emotional problems, hyperactivity/ inattentiveness, and prosocial behaviors.

Not all parents provided full questionnaire data during the study. Of the usable sample of 239 children, age and gender were collected for all 239 participants, 235 have basic demographic information (ethnicity, income, parental education), 195 provided their home zip code, 200 completed the parent-reported nearby natural features questionnaire, 171 completed the typical play environments questionnaire, and 151 completed the Strengths and Difficulties Questionnaire.

2.3. Data Analysis

2.3.1. Data Structure and Cleaning

The output from the picture sorting task was 10 trials of pictures ranked from 1 (least preferred in trial) to 4 (most preferred in trial). Occasionally, participants accidentally hit the advance button twice in a row, skipping the trial. To account for this, trials that had a duration less than 1 second were removed from the data.

The average rating used for demonstration purposes and data visualization (in **Figure 1.3**) was calculated by taking the average position of each photo across the 4 trials in which it occurred, resulting in a value between 1 (always chosen as least preferred) and 4 (always chosen as most preferred) for each picture. For subsequent statistical analyses, these average ratings were sorted from highest to lowest to create a full 1 to 10 ranking of all images in the set. In the case where multiple images had the same average rating, the higher ranking was given to the image that was more preferred in the trial(s) that included both images. Because the statistical analysis procedures used (described in the Statistical Analysis section) are not easily graphed, these average ratings were used to visualize the pattern of results.

A subset of adult participants (15 out of 182) were excluded from analysis after being identified as likely completing the task in reverse. This assessment was based on having both exceptionally low ratings (1 to 1.75 out of 4) of the very high aesthetic value nature images and high ratings (3.25 to 4) of the very low aesthetic value urban images. The very high aesthetic value nature images and very low aesthetic value urban images received very reliable ratings in previous stimulus validation studies and across the rest of the adult sample in the current study. Given that the adults were given the instructions but did not do any practice trials with an experimenter, and on a few occasions parents realized they were doing the task in reverse and told the experimenter this was the case, we felt confident that these adults were likely not paying close attention to the frown/smile anchors and simply made their rankings backwards.

2.3.2. Statistical Analysis of Task Data

As the task data were ordinal and included repeated measures, we conducted regression analyses using a proportional odds mixed model (McCullagh, 1980), fit using the "ordinal" package (Christensen, 2018) in R 3.5.1 (R Foundation for Statistical Computing, www.rproject.org). This analysis models the effect of predictors across the "cut-points" between categories of the ordinal criterion variable. In the case of our data, there were 10 categories (ranks 1 through 10) resulting in nine cut-points. The proportional odds model predicts the log odds of a given response being below each cut-point, under the assumption that a predictor's effects do not significantly differ across cut-points (i.e., the proportional odds assumption). If the proportional odds assumption is met, the model yields cumulative odds ratios that do not depend on the specific cut-points used. To test the proportional odds assumption, we modelled predictor separately as nominal effect and as an ordinal effect, and the model fits were compared via likelihood-ratio test. In all cases, the model fits did not significantly differ, indicating that the proportional odds assumption was met (i.e., the effect of predictors did not differ across cut-points). Maximum likelihood parameter estimates were obtained using an adaptive Gauss-Hermite quadrature approximation using 11 quadrature points (Lesaffre & Spiessens, 2001), implemented through the nlminb function in R.

Participant was included as a random intercept in our analyses to account for repeated measures. To specifically analyze age-related changes in the very high aesthetic value nature and very low aesthetic value urban categories, proportional odds modelling was conducted using the lrm function in the R package 'rms' (Harrell, 2018).

2.3.3. Analysis of Environmental Exposure and SDQ variables

To examine the relation between individual differences in children's nature exposure and cognitive functioning as measured by the Strengths & Difficulties Questionnaire, a multiple imputation procedure (Rubin, 1987) was first employed to handle participants with varying amounts of missing questionnaire data. Multiple imputation was performed using the "mice" package in R (van Buuren & Groothuis-Oudshoorn, 2010). Results with and without multiple imputation were very similar, and as such, the results reported in the manuscript are those from the raw (not imputed) data. Results including imputation to account for data missingness can be found in Appendix B.

2.3.4. Analysis of Parent-Child Similarity of Preference.

To analyze whether children's preferences reflect those of their parents, the Euclidean distance between children's rankings of images and their parents' rankings were calculated and compared to non-parents. For this analysis, the subset of data was used which only included independent parent-child pairs, as the inclusion of multiple siblings with a single parent would directly violate the assumption of independent samples in subsequent analyses. A standard distance calculation was performed, using the image ranks for all 10 images for kids and parents:

$$\sqrt{(\text{img1rank}_{\text{child}}-\text{img1rank}_{\text{parent}})^2+(\text{img2rank}_{\text{child}}-\text{img2rank}_{\text{parent}})^2+\dots+(\text{img10rank}_{\text{child}}-\text{img10rank}_{\text{parent}})^2}$$

This calculation was also performed on each child and every adult other than the child's parent in the same picture set. These calculations were then averaged, to create a value of the average distance of the child and the **n** other adults. That is, (

$$\sqrt{(\text{img1rank}_{\text{child}}-\text{img1rank}_{\text{adult 1}})^2+(\text{img2rank}_{\text{child}}-\text{img2rank}_{\text{adult 1}})^2+\dots+(\text{img10rank}_{\text{child}}-\text{img10rank}_{\text{adult 1}})^2}$$

$$+ \quad \dots \quad +$$

$$\sqrt{(\text{img1rank}_{\text{child}}-\text{img1rank}_{\text{adult n}})^2+(\text{img2rank}_{\text{child}}-\text{img2rank}_{\text{adult n}})^2+\dots+(\text{img10rank}_{\text{child}}-\text{img10rank}_{\text{adult n}})^2}$$

) divided by **n**.

2.3.5. Supplementary Analyses

A series of additional analyses were conducted to promote transparency and to ensure the reliability of our results. These results are reported in Appendix B, and any discrepancies between original results and these analyses are described in the results section. For analyses that involved the adult participants (section 3.1), we also conducted the same statistical tests including those 15 adult participants that were excluded from our sample. All analyses involving child participants were conducted using only the non-correlated subset (described in 2.2 Participants) which excludes siblings. Any analyses that involved our multiple imputation procedure (section 3.3) are also reported in this supplement.

3. Results

3.1. Environmental Preferences differ between Children and Adults

To test for preference differences between adults (i.e., parents and guardians in the current study) and children, a proportional odds mixed model was conducted using rank (1-10) as an outcome variable, environment type (nature/urban) and aesthetic value (high/low) as within-subject predictors, categorical age (adult/child) as a between-subjects predictor, and participant as a random variable. The model using the full interaction of predictor variables did not yield a significant 3-way interaction, and did not differ in goodness of fit from the model including only 2-way interactions, so the results of the more parsimonious model are reported.

A significant interaction of categorical age and environment type was found (B = 1.44, Z = 11.45, OR = 0.27, 95% CI [0.18, 0.30], p < 0.001), where children showed a greater preference for urban images compared to adults. Adults and children also differed in their preference for images based on aesthetic value (B = 0.16, Z = 2.90, OR = 0.70, 95% CI [0.54, 0.89], p = 0.003). Specifically, adults preferred the high aesthetic value images (compared to low aesthetic value ones) to a greater extent than did children (see **Figure 1.3**). When the 15 adults previously excluded (due to probable reversed responding) were included in this analysis, the interaction between categorical age and aesthetic value was no longer significant (p.s. = 0.3) but the interaction with environment was unaffected.

3.2. Children's Environmental Preferences

To test for children's preferences and how they may be influenced by age, a proportional odds mixed model was again employed using rank (1-10) as an outcome variable, environment type (nature/urban) and aesthetic value (high/low) as within-subject factors, child age as a continuous predictor, and subject as a random variable.

3.2.1. Group Effects

Results of this analysis showed a main effect of environment, where children generally preferred the urban environments compared to natural ones (B = -0.83, Z = -7.10, OR = 0.44, 95% CI [0.35, 0.55], p < 0.001). There was also a main effect of aesthetic value where children exhibited greater preferences for high aesthetic value images (B = 1.73, Z = 14.5, OR = 5.65, 95% CI [4.47, 7.14], p < 0.001). A significant interaction of environment and aesthetic value was also found (B = 0.70, Z = 4.31, OR = 0.50, 95% CI [0.36, 0.68], p < 0.001) where children exhibited a larger ranking difference between high and low aesthetic value nature images compared to urban images (see **Figure 1.3**). (Group effect analyses for adults can be found in Appendix B).

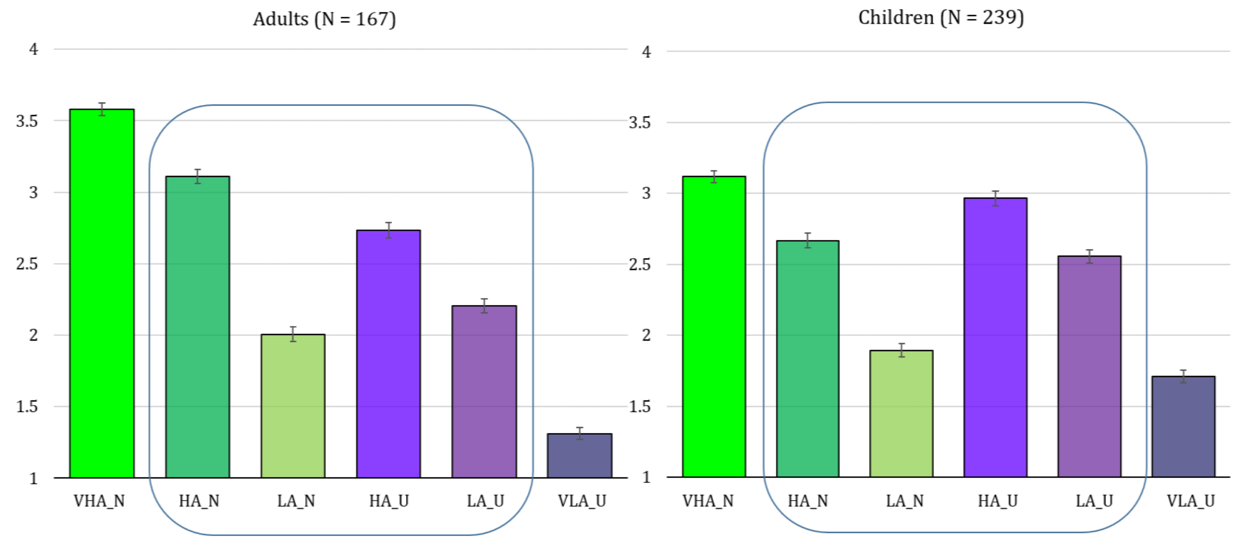


Figure 1.3. Preference for Environment Types in Adults and Children
Plotted are the average ratings for each image type (VHA_N = Very High Aesthetic value Nature, HA_N = High Aesthetic value Nature, LA_N = Low Aesthetic value Nature, HA_U = High Aesthetic value Urban, LA_U = Low Aesthetic value Urban, VLA_U = Very Low Aesthetic value Urban) in the adult and child samples. The boxes represent the four conditions that are completely crossed for statistical analysis. In this chart, higher ratings (closer to 4) represent more favored image types and lower ratings (closer to 1) represent less liked image types. Error bars indicate +/- SEM.

3.2.2. Age-Dependent Preferences

Importantly, both children's preference for environment type and aesthetic value showed significant interactions with age. With increasing age, children showed a lessened preference for urban environments over natural ones (B = 0.32, Z = 4.30, OR = 1.39, 95% CI [1.18, 1.63], $p < 0.001$). Additionally, children preferred images of high over low aesthetic value to a greater degree with increasing age (B = 0.27, Z = 3.31, OR = 1.31, 95% CI [1.11, 1.53], $p < 0.001$; **Figure 1.4**).

Age related changes in the very high aesthetic value nature images and very low aesthetic value urban images were examined in separate proportional odds regressions, predicting the rank of the image (1 to 10) by child age. A significant effect of age was found on rankings of very high aesthetic value nature images (B = -0.43, Z = -3.62, OR = 0.65, 95% CI [0.51, 0.82], p < 0.001), where higher preference rankings were associated with increasing child age. There was also a significant effect of age on ranking of very low aesthetic value urban images (B = 0.68, Z = 5.59, OR = 1.98, 95% CI [1.56, 2.52], p < 0.001). For this category, as age increased, children's preferences for these images decreased (**Figure 1.4**). In summary, as children aged, their preferences began to look more similar to those of adults.

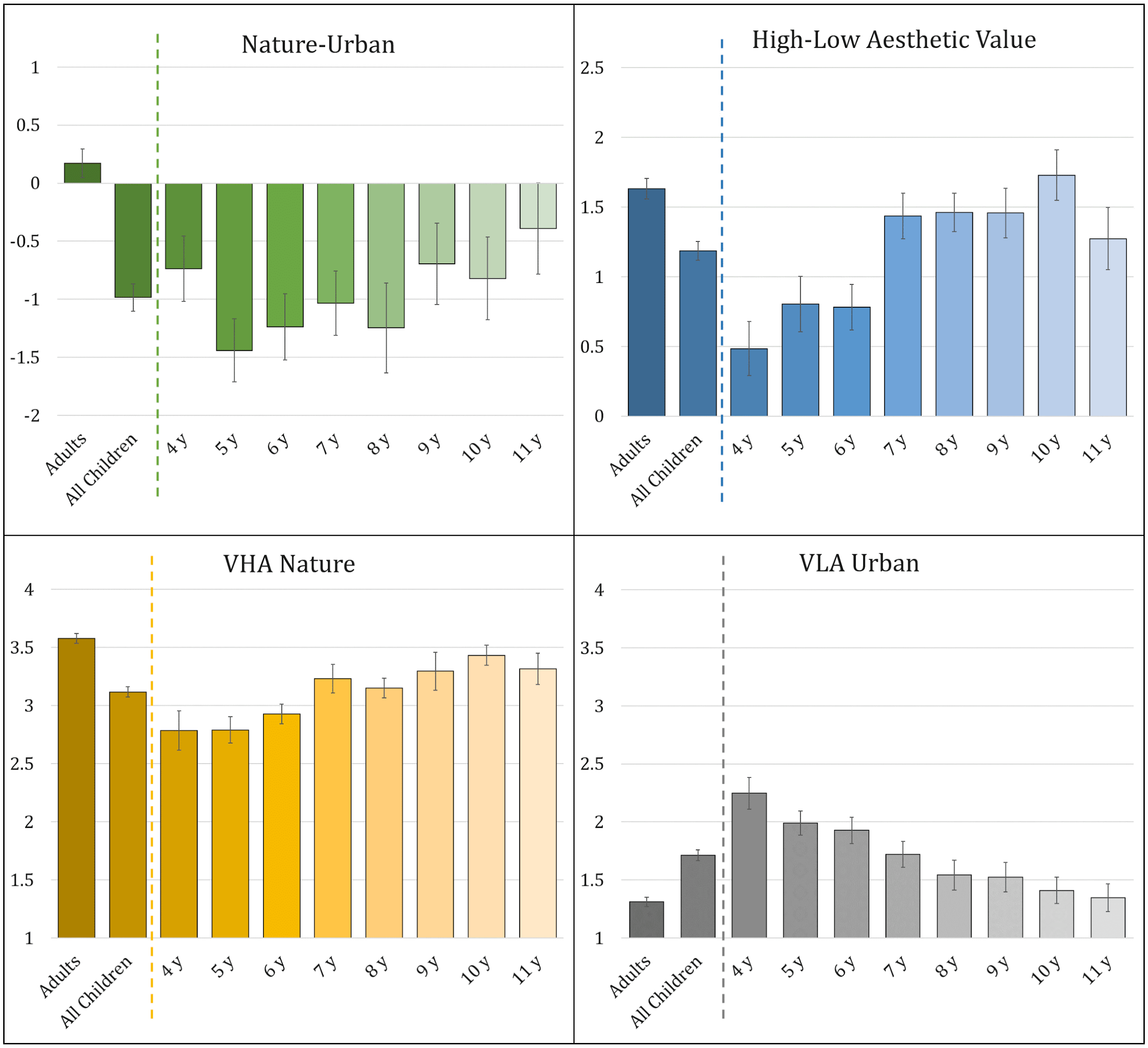


Figure 1.4. Developmental Changes in Environmental and Aesthetic Value Preference

The top panels represent differences calculated from average ratings, plotted across age groups. The environment preference measure ("Nature - Urban") displayed in the upper left panel was calculated by subtracting the average of the high and low aesthetic value Urban categories from the average of the high and low aesthetic value Nature categories. Similarly, the aesthetic value measure ("High - Low Aesthetic Value") displayed in the upper right panel was calculated by subtracting the average of ratings for Low Aesthetic value Nature and Urban images from the High Aesthetic value Nature and Urban images. The bottom panels represent average ratings for the Very High Aesthetic value Nature category (lower left) and Very Low Aesthetic value Urban category (lower right), divided across age bins. Error bars represent +/- SEM.

3.3. Relations between nature exposure, nature preference, and children's behavior

A goal of this study was to examine whether any of the observed benefits of nature exposure on children's cognitive, social, and emotional functioning could be explained by individual differences in children's nature preferences. Another aim was to see if nature exposure was linked to preference.

3.3.1. Nearby Nature Predicts Children's Attention

Based on extensive prior research that looked at the effects of nature exposure on children's attentional functioning, a confirmatory analysis was conducted examining the effects of nearby nature and play in nature on the parent-reported hyperactivity/ inattentiveness subscale from the SDQ. As predicted, greater parent-reported nearby natural features were correlated with lower scores on the SDQ hyperactivity/ inattentiveness (r = -0.21, 95% CI [-0.36, -0.05], p = 0.009). However, income was also negatively correlated with hyperactivity/ inattentiveness (r = -0.21, 95% CI [-0.36, -0.04], p = 0.015). Thus, separate analyses were run to see whether natural features were still predictive when income was included in the model. Results from a linear regression indicated that both greater parent-reported natural features and higher income were significant independent predictors of children's inattentiveness/hyperactivity (Natural features: B = -0.31, p = 0.015; η_p^2 = 0.04; Income: B = -0.31, p = 0.023, η_p^2 = 0.04). When this analysis was conducted on the non-correlated subset of the data which reduced our sample size for this analysis, this effect became non-significant (p = 0.1). Surprisingly, play in more natural environments was not significantly related to this SDQ measure (*ps* = 0.2).

3.3.2. Other Nature-Behavior Relations

An exploratory analysis was conducted to examine whether parent-reported interactions with nearby nature (natural features near home/school), parent-reported play in natural over built environments, or objective measures of nearby nature (zip code based NLCD measures of natural features and high versus low developed space) were correlated with any of the other SDQ subscales: conduct problems, emotional problems, peer problems, and prosociality.

Results of the full correlation matrix can be found in **Table 1.4.** As expected, there were strong correlations between NLCD natural features and parent-reported nearby natural features (r = 0.44, 95% CI [0.32, 0.55], p < 0.001), as well as NLCD natural features and play in natural environments (r = 0.32, 95% CI [0.18, 0.44], p < 0.001).

More parent-reported nearby natural features were modestly correlated with lower scores on the SDQ conduct problems subscale (r = -0.15, 95% CI [-0.3, 0.01], p = 0.077), but was not related to any other SDQ subscales. Play in natural environments and nearby natural features calculated from zip-code based national land cover data (NLCD) were not related to any SDQ measures (all $p > 0.13$).

Table 1.4. Correlation matrix of nearby nature, SDQ, demographics, and child environmental preference

Correlation coefficients are listed for all correlations between parent-reported and objective nature, SDQ subscales, SES measures, and children's environmental preferences. Coefficients with ** indicate a *p*-value of less than 0.01, and those with * indicate $p < 0.05$.

	PR Nat Play	NLCD Nat Feat	NLCD Dev	SDQ Emot	SDQ Cond	SDQ Inatt/ Hyper	SDQ Peer	SDQ Prosoc	Income	Mom Ed	Child Env Pref
PR Nat Feat	.37**	.44**	.32**	.00	-.15	-.21**	-.04	.15	.14	.01	-.02
PR Nat Play		.32**	.19**	-.07	-.09	-.11	-.08	.13	.02	.14	.00
NLCD Nat Feat			.56**	-.13	-.09	-.12	-.1	.14	-.06	.01	-.07
NLCD Dev				-.08	-.08	-.1	.02	.14	.05	.06	-.01
SDQ Emot					.31**	.36**	.34**	.26**	-.24**	-.04	.00
SDQ Cond						.49**	.08	-.43**	-.14	-.09	-.13
SDQ Inatt							.16	-.34**	-.21*	-.19*	-.05
SDQ Peer								-.12	-.15	-.02	.05
SDQ Pros									.05	.05	-.03
Income										.45**	-.13
Mom Ed											-.08

[Legend: PR Nat Feat = Parent-reported nearby natural features, PR Nat Play = Parent-reported measure of child's play in nature over built space, NLCD Nat Feat = natural features calculated from zip-code level NCLD data, NLCD Dev = ratio of open/low to medium/high developed space from zip-code level NLCD data, SDQ Emot = Emotional problems subscale, SDQ Conduct = Conduct problems subscale, SDQ Inatt = Inattentive/Hyperactive subscale, SDQ Peer = Peer problems subscale, SDQ Pros = Prosocial behavior subscale, Income = family household income, Mom Ed = Maternal Education, Child Env Pref = Child's preference for natural over urban environments]

3.4. Children's Nature Preferences and Environmental Exposure

To examine whether individual differences in nature preferences related to their nearby nature exposure, a metric of nature versus urban preference was also examined in the correlations. This metric was calculated by taking the average rank for the four nature images and subtracting the average rank of the four urban images.

3.4.1. Preference and Exposure Across All Children

The results of these correlations indicated no significant correlations between children's environmental preferences and any of the nearby nature exposure or behavioral measures (see **Table 1.4**).

3.4.2. Interaction of Nature Preference, Environmental Exposure, and Age

A proportional odds mixed model was conducted to examine the relationship between nearby natural features, preference for natural environments, and age. Results of this analysis showed a significant 3-way interaction of child age, environmental preference, and parent-reported nearby natural features (B = 0.39, Z = 4.6, OR = 1.49, 95% CI [1.26, 1.76], $p < 0.001$). To unpack this interaction, follow up proportional odds mixed models were performed separately on each of the datasets include 4-5 year-olds, 6-7 year-olds, 8-9 year-olds, and 10-11 year-olds.

None of the interactions between nearby natural features and preferences reached significance in the 4-to-5, 6-to-7, or 8-to-9-year old children. However, in the 10-to-11-year old group, this interaction was significant ($B = 0.68$, $Z = 3.45$, $OR = 1.96$, 95% CI = [1.33, 2.89], $p < 0.001$). The results of this analysis suggested that with the older children, nearby natural features were positively related to a greater preference for natural over urban images, but this was not the case for children under 10.

3.5. Parental Influences on Children's Preferences

Results of a repeated measures ANOVA comparing child-parent distance and child-nonparent average distance indicated that the distance between a child and their parent ($M = 9.96$, $SD = 3.17$) was smaller than that between the child and the average non-parent ($M = 13.8$, $SD = 1.0$; $F(1,143) = 152.9$, $p < 0.001$, $\eta^2 = 0.4$). This shorter distance indicates that, overall, children demonstrated more similar preferences to those of their own parent relative to parents of other children viewing the same images.

To see whether parent-child similarity was affected by child age, a mixed model ANOVA was run predicting distance from child-parent pairs versus child-non-parent pairs and child year of age, with a random effect for subject. This analysis yielded a significant interaction of age and parent vs. non-parent similarity ($F(7,136) = 2.83$, $p = 0.009$, $\eta^2 = 0.08$; **Figure 1.5**), where children's preferences were more similar to those of their parent over other adults with increasing age.

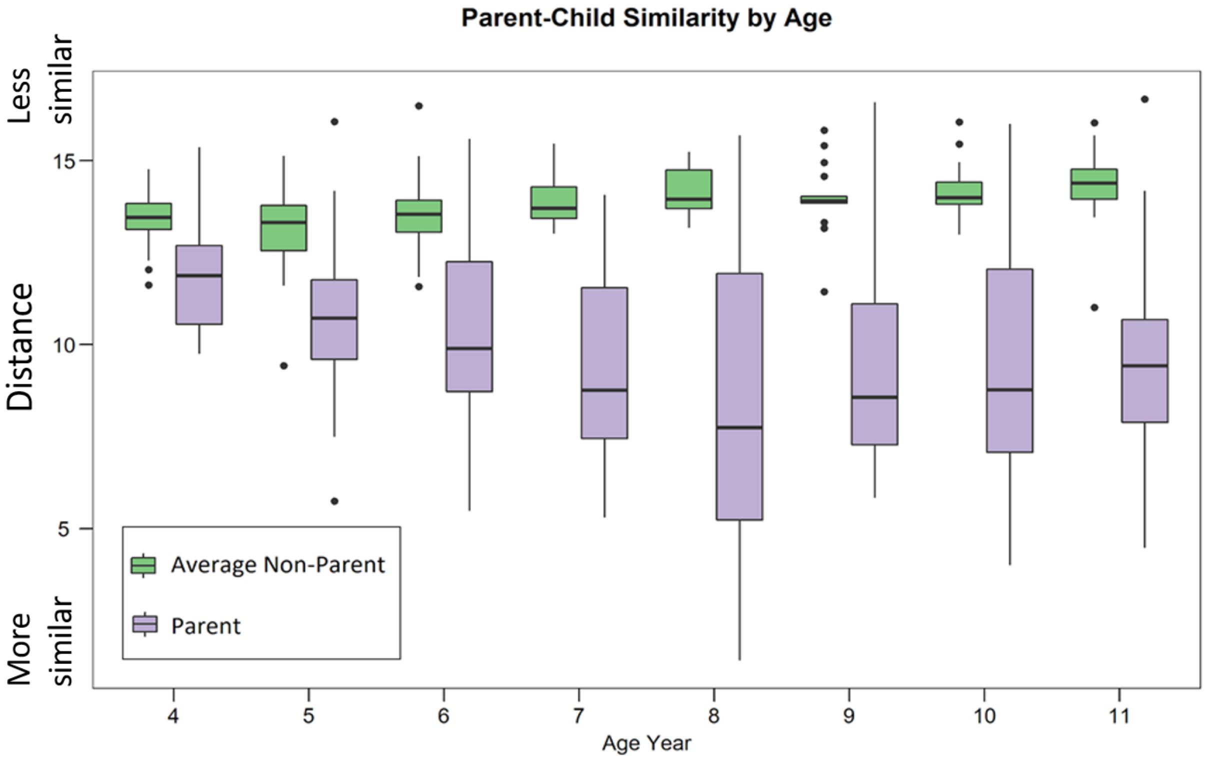


Figure 1.5. Age effects for Parent-Child Preference Similarity
Boxplot depicting the distance between preferences of children and their parents (purple) and children with the average parent rating the same picture set (green) plotted separately by year of age.

4. Discussion

Whether the pervasive preferences that adults show for nature (R. Kaplan & Kaplan, 1989) also exist in young children is an important, unanswered question. This study attempted to address this outstanding issue, and to understand how and when such preferences develop. Unexpectedly, this study found that children (4 to 11 years) have robust urban preferences, counter to adult preferences. However, this preference for urban environments decreased, almost linearly, with children's age. This key finding (among others) from the current study provide greater support for a learned affinity for nature, rather than an affinity that has been genetically programmed and present from birth. If an innate, biophilic response is indeed responsible for the preference for nature observed in adults, this developmental trajectory fits most readily with a biologically prepared learning account. This account has also been previously proposed in cross-cultural studies of rural vs. urban environmental effects on attention over development (Kardan et al., 2017).

Young children in our study exhibited a considerable urban preference, but this counterintuitive finding does not appear to have an obvious cause. One feature that strongly relates to children's preferences is stimulus novelty(Cantor & Cantor, 1964). A novelty account would make the argument that children find the urban images more interesting because the scenes are less familiar. This is an unlikely explanation for our results as overall, parent-reported and objective measures of nature near children's homes or schools were not related to their preferences (i.e. a more urban home environment did not predict preference for nature and vice versa).

The current study also found that children's preferences looked more similar to their parent's with increasing age, which suggests that with more time with one's parents/caregivers, child and parent preferences converge as compared to more divergence at younger child ages. Note, this is not just due to older kids exhibiting preferences more like adults in general, as these effects were specific to similarities in preference with one's own parents/caregivers vs. the other adults in the study (i.e., a difference score). Thus, it appears that parental preferences may be a key influence on what children prefer, but this influence seems to manifest later in development. This study has identified that environmental factors such as nearby natural features and parental preferences affect older child environmental preferences more than younger child environmental preferences, but our data cannot speak to any additional mechanisms or identify what specific factors might impact the preferences of young children. The observed effects do, however, provide an exciting avenue for additional research to identify what underlying mechanisms drive these preferences.

This study also provides support for the idea that children need not prefer natural environments to reap the cognitive benefits from interacting with natural environments. We identified a significant correlation between having greater nearby nature in children's home and school environments and reduced child inattentiveness and hyperactivity, but these effects were not related to children's preferences. This pattern is consistent with Attention Restoration Theory(S. Kaplan & Berman, 2010), which suggests that nature exposure improves cognitive functioning through replenished attentional capacity, and not through preference-driven mood changes. Additionally, if young children do not generally prefer nature, then it is possible that the cognitive benefits observed for children after nature exposure from other studies may also not be due to preference(Dadvand et al., 2015; Faber Taylor & Kuo, 2009, 2011; Wells, 2000). However, as this study does not involve directly manipulating nature exposure for these children, these data do not lend themselves to strong conclusions on the role of preference on cognitive effects. To directly test this, future empirical studies that utilize nature interventions on children's cognitive functioning should consider including a measure of children's environmental preferences, but that is careful to use stimuli that decouple preference from environment type.

Like all studies, our study is not without limitations. First, though we are using the same anchors and verbal descriptions of preference in both adults and children (i.e., "like", "favorite"), it is impossible to tell with the current design whether we are tapping into the same psychological construct. For example, it could be that children and adults differentially weight the desire to be in a given environment when making a preference evaluation. Again, future research would be required to rule out this explanation. Another limitation relates to the age range chosen. Though a pilot study indicated that we would have difficulty collecting data from children under 4 years of age with this paradigm, there are potentially fascinating and important developmental effects in infants and younger children that require investigation with age-appropriate experimental procedures. Our 11-year old participants did not display preference patterns identical to those of adults, and as such, extending this work into older ages would be required to fully examine the complete developmental trajectory. This is particularly important as previous research has shown that environmental preferences during the adolescent years may be quite different from those displayed in adulthood (Balling & Falk, 1982; R. Kaplan & Kaplan, 2002). Lastly, though the effect we observed relating attentional problems to less nearby nature has been found in other studies (Amoly et al., 2014; Dadvand et al., 2015; Faber Taylor & Kuo, 2009, 2011; M. Kuo et al., 2019), it is still correlational and uses subjective, parent reported inattentiveness, which is a limitation of this survey-based data.

Importantly, our study decoupled image preference from environment type and used multiple picture sets to allow for generalizing above a particular set of images. It would be beneficial to implement this paradigm with a wider array of images to ensure that these results can be generalized further, though it may be difficult to increase the image corpus much, as young children only have the attention span to perform the task for a few minutes. It was therefore outside the scope of the current study to include additional image sets that may not be preference-equated but chosen to represent more variety in nature (i.e. lakes, forests, deserts) or urban (i.e. houses, cityscapes) scenes. However, future research on this topic which includes a wider array of image types would be a valuable addition to the current results.

In summary, these data provide the first step in a broader series of studies that can attempt to understand why children and adults differ in their environmental preferences, what drives age-dependent changes in preference, and whether nature preferences are necessary for all, some, or none of the observed cognitive, emotional and health benefits of nature exposure in children.

CHAPTER 2: The Affective Benefits of Nature Exposure – What's Nature Got to Do with It?

Abstract

Nature interactions have been demonstrated to produce reliable affective benefits. While adults demonstrate strong preferences for natural environments over urban ones, it is not clear whether these affective benefits result from exposure to nature stimuli per se, or result from viewing a highly preferred stimulus. In one set of studies (Study 1 and 2), state affect before and after image viewing was examined as a function of both preference level (high, low, very high, or very low aesthetic value) and environment type (nature or urban). When aesthetic value was matched, no differences in affect change were found between environments. However, affect change was predicted by individual participants' ratings for the images. The largest affective benefits occurred after viewing very high aesthetic nature images, but Study 2 lacked an equivalently preferred urban image set. In a second set of studies (Study 3 and 4), new sets of very highly preferred images in categories other than nature scenes (urban scenes and animals) were employed. As before, individual differences in preference for the images (but not image category) was predictive of changes in affect. In a final study (Study 5), the nature and urban images from Study 1 and 2 were rated on beauty to assess whether the stimuli's preference ratings were capturing anything other than simple aesthetics. Results from study 5 showed that beauty/aesthetics and preference ('liking') were nearly

identical. Together, these results suggest that nature improves affective state because it is such a highly preferred environment.

Introduction

In the field of environmental psychology, a substantial body of research documents the relationships between the physical environment and human psychological functioning. One particular focus has been the link between nature exposure and emotional well-being. This link has been demonstrated using a variety of study designs, exposure types, and outcome measures (Bowler et al., 2010; McMahan & Estes, 2015). Frequently, such studies have employed controlled experimental designs in which the effects of nature interventions are compared with that of control interventions (often urban environment exposures).

Short-term effects of nature exposure have been examined using nature interventions that have varied in type and duration, but the observed benefits are consistent. Brief walks in natural settings have been shown to increase positive affect and decrease negative affect compared to urban walks, and the effects have been found in both healthy and clinical populations (Berman et al., 2012; Bratman, Daily, et al., 2015; Fuegen & Breitenbecher, 2018; Hartig et al., 2003; Johansson et al., 2011; Mayer et al., 2009). Passively viewing a natural environment (e.g. sitting in a forested area) improves self-reported and physiological measures of affect compared to spending the same time viewing a built environment (e.g. sitting in a parking lot) (Lee et al., 2009; Tsunetsugu et al., 2013). Simulated nature in the form of videos, image slideshows, and virtual reality (VR) elicit improvements in emotional state as well (Beute & de Kort, 2014; T. Hartig et al., 1996; Mayer et al., 2009; Valtchanov et al., 2010; van den Berg et al., 2003), though the effects are somewhat smaller than those of actual nature exposure (McMahan & Estes, 2015).

Benefits of nature exposure have also been examined longitudinally in epidemiological (e.g. White et al. 2013) and experience sampling studies. The latter combine regular assessments of emotional state (and other measures) with GPS location data from a mobile device. These studies find that being in an outdoor natural environment is related to more positive affect than being in other types of environments (e.g., indoors at home, outdoor urban environments, in transit, etc.) even when accounting for other relevant variables such as weather, daylight, or physical activity (Beute & de Kort, 2018; Glasgow et al., 2019; MacKerron & Mourato, 2013; Ryan et al., 2010).

While studies documenting benefits of nature exposure on emotional well-being are abundant, it is unclear whether these effects are the result of exposure to nature per se, or whether these effects are the result of viewing preferred stimuli. Research on adult environmental preferences has demonstrated that overwhelmingly, nature scenes are preferred over their urban counterparts (Kaplan & Herbert, 1987; Kaplan & Kaplan, 1989; Kaplan et al., 1972; Ulrich, 1979, 1983). Despite work identifying preference as an influence in the emotional benefits of nature exposure (Beute & de Kort, 2014; Browning et al., 2019; Mayer et al., 2009; White et al., 2010), many studies linking nature contact to psychological well-being do not assess preference for these environments, or examine how affect change relates to preference (Lee et al., 2009; Tsunetsugu et al., 2013; Valtchanov et al., 2010).

Those studies which have directly linked preference for natural environments to affective benefits have approached the question in different ways. Beute and de Kort (2014) showed participants an image slideshow and measured state affect and preference for the slideshows. The nature slideshow was preferred over the urban one, and a subsequent mediation analysis showed that positive affect change (operationalized as "hedonic tone") was mediated by preference for the slideshow. A different approach was taken by White et al. (2010), who did not employ the standard pre/post study design. Instead, they had participants rate a series of scenes on aesthetics (i.e "how attractive is the scene?"), behavioral preference (i.e. "how willing would you be to visit this scene?"), and their affective response to each image (i.e. "how does this photo make you feel?"), and then examined how these attributes related to each other and to the image types (i.e. fully built environments, green nature, nature with water, built environments with water, etc.). They found that images which were rated most highly on preference were also highly rated on positive affective responses. Additionally, they found that by incorporating bodies of water in images of built environments, these environments were also rated more favorably on both preference and affect measures, compared to urban images without water. Another approach measured affective restoration using videos of a built environment with water elements (dockland) and contrasting it with a video of a nature reserve (Karmanov & Hamel, 2008). Here, the nature video was rated as more attractive than the urban one, and the nature video caused relatively larger affective restoration, again suggesting that there is a potentially important role of aesthetic preferences.

These findings suggest an important question, namely whether there is something unique or special about visual exposure to nature itself (other than that it is preferred over other environments) which leads to affect change. For example, the theory of Biophilia (Kellert & Wilson, 1995) emphasizes nature's evolutionary significance and suggests that our emotional responses to nature result from our affinity towards natural things, though not necessarily aesthetically beautiful things. Other researchers have suggested that nature may be endogenously visually rewarding due to the distribution of visual spatial frequency information in nature scenes (Valtchanov & Ellard, 2015). Thus, two explanatory possibilities for the affective benefits of nature are that 1) exposure to natural scenery induces positive emotions, regardless of how they are aesthetically preferred, and that 2) nature scenes induce positive emotions because we have an aesthetic preference for such environments. However, unless participants are exposed to different types of environments that are equally preferred, it is hard to evaluate whether there is something unique about nature independent of preference. The current set of studies were designed to answer this outstanding question.

The aim of this research was to test the effects of environment type (nature vs. urban) and aesthetic value (preference level) on changes in affect from pre to post viewing of image sets. It should also be noted here that in each of these studies, nature is operationalized as scenes or environments with primarily natural (non-built) elements, rather than anything that is non-built or organic in nature. Study 1 was conducted to generate the preference-equated sets of natural and urban scenes. We reasoned that if nature has a positive effect on affective state that is not simply due to its status as a preferred environment, then exposure to natural environments should elicit larger positive affective changes than preference-equated urban environments. Conversely, if environment-type is less important than aesthetic preferences, then differences in affect should be observed primarily between image sets that vary on aesthetic value, but not on environment-type. Alternatively, both preference and environment type could have interactive effects on affect. All of these possibilities were tested in Study 2.

Study 1: Original Image Validation

1.1. Study Intro

Study 1 was conducted to find preference-equated sets of natural and urban scene images to be used in Study 2, which then examined how exposure to images of different environments across multiple levels of aesthetic value (preference) altered state affect.

1.2. Method

1.2.1. Participants

401 US-based adults (195 male, 201 female, 4 other, 1 no response) were recruited from Amazon Mechanical Turk. The age of participants ranged from 19 to 73 years (*M* = 36.0, *SD* = 11.0). The full study procedures were expected to take approximately 15 minutes and participants were compensated $1.50 for participating. Informed consent was administered by the University of Chicago Institutional Review Board (IRB).

1.2.2. Stimuli & Procedure

Participants were randomly assigned to rate 100 images out of a total of 375 potential images on a 1-7 Likert scale evaluating either their preference for the images (anchors of 1 = “strongly dislike” and 7 = “strongly like”; n = 200) or the naturalness of the image (anchors of 1 = “very man-made” and 7 = “very natural”; n = 201). The 375 images were taken from the SUN database (Xiao et al., 2010) as well as an online image search for non-copyrighted scene images. Criteria for image inclusion were that the images: 1) Do not contain people or animals, 2) Have sufficiently high resolution for clear viewing, 3) Have minimal trees/natural elements in the urban images and minimal man-made elements in the nature images, and 4) Have minimal text, signs, or graffiti. Additionally, images were selected containing a variety of viewpoints (horizons, slanted towards the ground, slanting upwards, etc.) and varied urban or natural forms (different types of nature, varied buildings or cities). The 100 images that participants saw were selected pseudorandomly, to show a relatively equal number of natural and urban images as well as images that varied on aesthetic value. Based on this pseudorandomization each image received a minimum of 33 preference ratings, with an average of roughly 51 ratings per image, and a minimum of 37 naturalness ratings, with an average of 53 ratings per image.

1.3. Results

From the 375 images examined, six sets of 45 images each were able to be created. Aesthetic preference ratings for the 45 images in each condition are presented in **Table 2.1**. Two pairs of preference-matched nature and urban image sets were created - High Aesthetic Value Nature (HA-Nat) and Urban (HA-Urb), and Low Aesthetic Value Nature (LA-Nat) and Urban (LA-Urb) conditions. The High and Low Aesthetic Value image sets had an average preference rating of 4.6 and 3.8, respectively. Unfortunately, there were very few strongly liked urban images or strongly disliked nature images. As such, there were not enough urban images to match the Very High Aesthetic Value Nature (VHA-Nat) images, nor were there enough disliked nature images to match those in the Very Low Aesthetic Value

Urban (VLA-Urb) condition. Though we were not able to equate these stimuli, we did not want to throw them out as we could still examine differences in preference within environment type.

Across all Aesthetic Value levels, images in the Nature conditions and Urban conditions were rated very differently on naturalness. Average naturalness ratings for VHA, HA, and LA-Nat were 6.72, 6.66, and 6.64, respectively. Average naturalness ratings for HA, LA, and VLA-Urb were 1.42, 1.44, and 1.39, respectively.

Table 2.1. Study 1 Aesthetic Value Ratings of Images in Each Condition
Ratings of all images in each condition from a previous validation study. Aesthetic value ratings are on a 1-7 scale (1 = strongly dislike, 7 = strongly like).

	Nature Conditions		**Urban Conditions**	
Aesthetic Value Condition	*M (SD)*	*Range* [Min, Max]	*M (SD)*	*Range* [Min, Max]
Very High Aesthetic Value	6.34 (0.13)	[6.19, 6.62]		
High Aesthetic Value	4.59 (0.17)	[4.28, 4.86]	4.58 (0.29)	[4.15, 5.29]
Low Aesthetic Value	3.78 (0.35)	[2.90, 4.21]	3.78 (0.23)	[3.30, 4.12]
Very Low Aesthetic Value			2.64 (0.31)	[1.77, 3.08]

Study 2: Examining Affect Change as a Function of Environment and Aesthetic Value

2.1. Study Intro

The primary question for Study 2 was whether changes in affect are due to naturalness, whether they are purely due to preference, or result from a combination of the two. All three of these possibilities were examined, using the image sets obtained from Study 1. A visual depiction of the hypotheses for Study 2, as well as the analyses chosen to address each hypothesis, are presented in **Figure 2.1**.

2.2. Method

Study 2 was preregistered on OSF prior to data collection: https://osf.io/tuezg

2.2.1. Participants

615 US-based adults (287 male, 324 female, 4 other) were recruited from Amazon Mechanical Turk. Age of the participants ranged from 20 to 76 years (*M* = 37.0, *SD* = 10.9). The full study procedures (including additional tasks after collecting the data for this study) were expected to take approximately 30 minutes and participants were compensated $3.00 for participating. Informed consent was administered by the University of Chicago Institutional Review Board (IRB). Sample size was decided prior to data collection and specified in the pre-registration. Each of the image conditions presented below was also broken into two groups in

subsequent tasks (unreported, see study preregistration for more details: https://osf.io/tuezg), so the sample was based on power for these subsequent tasks.

2.2.2. Experimental Conditions & Stimuli

Participants were randomly assigned to one of 6 image conditions based on Study 1: Very High Aesthetic Value Nature (VHA-Nat; n = 103), High Aesthetic Value Nature (HA-Nat; n = 103), High Aesthetic Value Urban (HA-Urb; n = 104), Low Aesthetic Value Nature (LA-Nat; n = 103), Low Aesthetic Value Urban (LA-Urb; n = 100), and Very Low Aesthetic Value Urban (VLA-Urb; n = 102). The full image sets and ratings (validated in Study 1) can be accessed at this link: https://osf.io/ehtk9/

2.2.3. State Affect Measures

To assess changes in state affect, two primary measures were used. The first consisted of a 6-item version of the State Trait Anxiety Inventory (STAI) (Marteau & Bekker, 1992) with 3 negative items (upset, tense, worried), and 3 positive low-arousal items (calm, relaxed, content). The presentation order of the 6 items in this scale was randomized across time points and participants. The second was a visual analog scale (VAS) for 4 emotion labels: happy, sad, inspired, and angry. The scale spanned from 1 to 100 at 1 unit intervals, and the order of the emotion labels was also randomized across time points and participants.

The composite STAI measures (STAI-Pos and STAI-Neg) were used in all reported analyses. A principal component analysis was also employed which allowed inclusion of all affect measures, and demonstrated very similar effects to the results of the STAI results reported. However, the data required standardization prior to performing the PCA, which affected the interpretability of changes in principal component values, so these analyses and results are detailed in the supplementary materials (Appendix C).

2.2.4. Procedure

Participants were randomly assigned to one of the 6 conditions (i.e., VHA-Nature, HA-Nature, LA-Nature, HA-Urban, LA-Urban, and VLA-Urban). After providing informed consent, participants filled out baseline measures of their affective state (T1). Subsequently, they viewed a series of 45 images and gave ratings of their aesthetic preference for the images on a 1-7 scale (1 = strongly dislike, 7 = strongly like). Each image was on the screen for a minimum of 7 seconds, and the next image would appear after a rating for the image was selected. After the image rating task, participants completed the same measures of their affective state a second time (T2).

2.2.5. Analysis

A visual depiction of the hypotheses for this study as well as the analyses chosen to address each hypothesis are presented in **Figure 2.1**. Before comparing affect change as a function of condition, preliminary analyses were conducted to examine whether the image sets used in the study elicited significant changes in emotional state. To this end, one sample t-tests were conducted on change scores for STAI-Pos and STAI-Neg, testing the null hypothesis that there was no change between T1 (pre) and T2 (post). This was performed separately for each condition (Analysis 1).

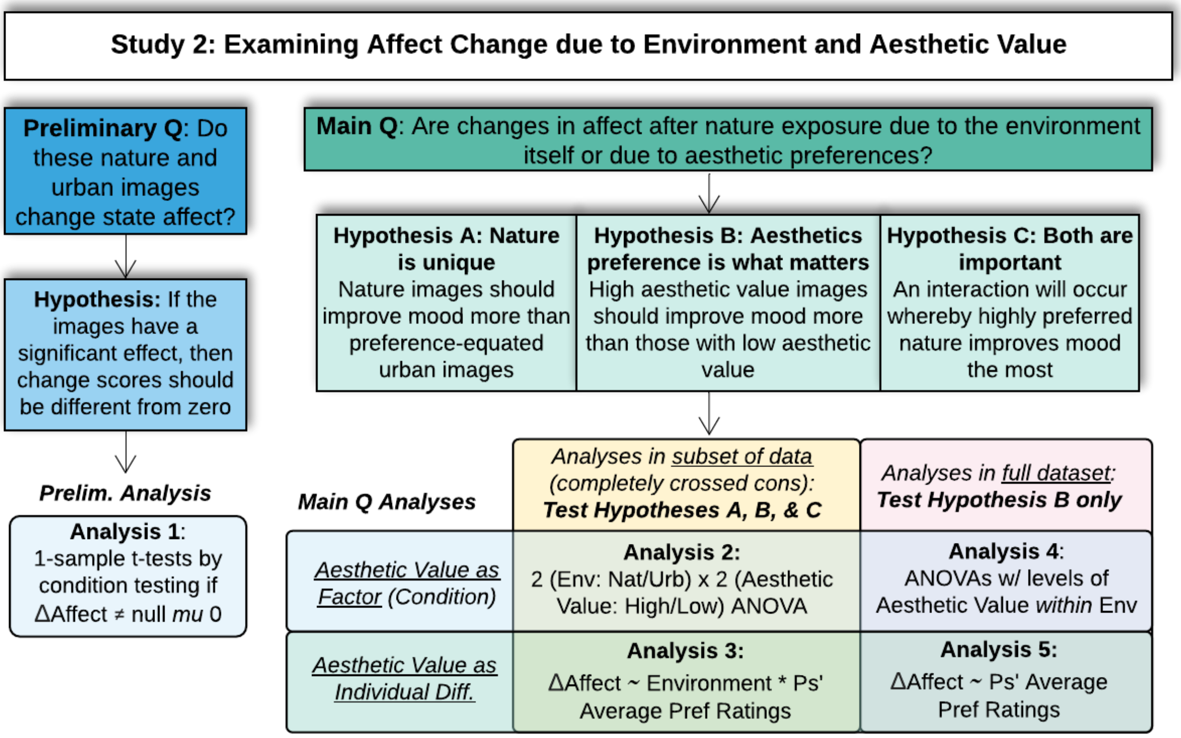


Figure 2.1. Hypotheses & Analyses Overview for Study 2

To test whether changes in affect are due to naturalness (Hypothesis A: *Nature is unique*), whether they are purely due to preference (Hypothesis B: *Aesthetic preference is what matters*), or a combination of the two (Hypothesis C: *Both are important*), four analytical approaches were employed. In those conditions which were completely crossed (i.e., HA-Nat and HA-Urb, LA-Nat and LA-Urb), both an ANOVA (Analysis 2: testing aesthetic preference conditions and environment type) and a linear regression (Analysis 3: using individual differences in preference and environment type) were conducted. In these completely crossed conditions, both main effects of environment and aesthetic preference could be examined, as could the interaction of the two. To accommodate the conditions that were not completely crossed, ANOVAs examining different aesthetic value conditions within an environment type were conducted (Analysis 4). Additionally, to analyze the role of individual differences in preference for the images, a linear regression predicting change in affect by participants' average preference rating (ignoring condition) was conducted (Analysis 5).

All statistical analyses were conducted in R v. 3.5.1 (R Core Team, 2019). ANOVAs were conducted using package 'ez' (v4.4-0, Lawrence, 2016), and post hoc comparisons were conducted using Tukey HSD multiple comparisons correction with a 95% family-wise confidence level. All between-subjects ANOVAs were first tested for significant heteroscedasticity using the Breush-Pagan test (function bptest in package 'lmtest' ; Zeleis & Hothorn, 2002). If identified, the ezANOVA option "white.adjust" was set to true, which uses a heteroscedasticity-corrected coefficient covariance matrix ('hccm' in package 'car'; Fox & Weisberg, 2019). Partial eta-squared effect size and 95% CIs were calculated using 'eta_sq' in package 'sjstats' (v0.17.6, Lüdecke, 2019). Cohen's d effect size and 95% CIs for comparison of group means were calculated using the 'apa.d.table' in package 'apaTables' (v.2.0.5, Stanley, 2018). Linear regressions were conducted using the 'lm' command in the 'stats' package (R Core Team, 2019).

Participants' preference for the images they viewed were determined by taking the average of their ratings for all 45 images they viewed. The average and standard deviation for each picture condition across all participants in that condition are presented in **Table 2.2.** As in Study 1, the average preference ratings for the equated image sets (HA-Nat and HA-Urb, LA-Nat and LA-Urb) were not significantly different from one another.

Table 2.2. Study 2 Aesthetic Value Ratings of Images in Each Condition
Ratings of all images in each condition from participants in Study 2. Aesthetic value ratings are on a 1-7 scale (1 = strongly dislike, 7 = strongly like).

Aesthetic Value Condition	**Nature Conditions**	**Urban Conditions**
	M (SD)	*M (SD)*
Very High Aesthetic Value	5.65 (0.84)	
High Aesthetic Value	4.78 (0.96)	4.55 (0.90)
Low Aesthetic Value	4.14 (1.18)	4.07 (0.79)
Very Low Aesthetic Value		3.26 (0.92)

2.3. Results

2.3.1. Analysis 1: Preliminary Tests for Significant Affect Change by Condition

The results of these preliminary tests can be found in **Table 2.3**. Overall, the VHA-Nat, HA-Nat, HA-Urb, and VLA-Urb elicited a significant change in positive and negative state affect. However, this was not generally true for the LA images, with the exception of a modest increase in STAI-Pos in LA-Nat. All changes were in a positive direction (affect improvement) with the exception of the VLA-Urban images, which had a deleterious effect on affective state.

Table 2.3. STAI Results: Change in affective state relative to baseline
Results of 1-sample t-tests comparing STAI positive and negative change to zero in each of the 6 conditions. **Significant p-value with Bonferroni family-wise multiple comparisons correction (α = 0.008) *Significant p-value uncorrected (α = 0.05)

		STAI-Positive			STAI-Negative		
Condition	**n**	***t-statistic***	***p-value***	***Cohen's d***	***t-statistic***	***p-value***	***Cohen's d***
VHA-Nat	103	5.46**	<0.001	0.54	-4.28**	<0.001	0.42
HA-Nat	103	4.22**	<0.001	0.41	-3.90**	<0.001	0.38
HA-Urb	104	2.42*	0.017	0.24	-2.83**	0.006	0.28
LA-Nat	103	2.54*	0.012	0.25	-0.70	0.486	0.07
LA-Urb	100	1.86	0.065	0.19	-1.84	0.069	0.18
VLA-Urb	102	-2.14*	0.034	0.21	2.41*	0.018	0.24

2.3.2. Analysis 2: Affect Change in Completely Crossed Data Subset (Aesthetic Value as Factor)

This analysis was conducted using only those conditions which were completely crossed (HA-Nat, HA-Urb, LA-Nat, LA-Urb), excluding the VHA-Nat and VLA-Urb conditions. To test the effect of environment and aesthetic value on affect change, 2 (Nat vs. Urb) x 2 (High vs. Low Aesthetic Value) factorial ANOVAs were conducted on change scores (T2 minus T1) for STAI positive and STAI negative scores. Results of these analyses did not yield significant effects of environment, aesthetic value condition, or the interaction for either STAI-Pos or STAI-Neg (all $p > 0.1$) [**Figure 2.2**].

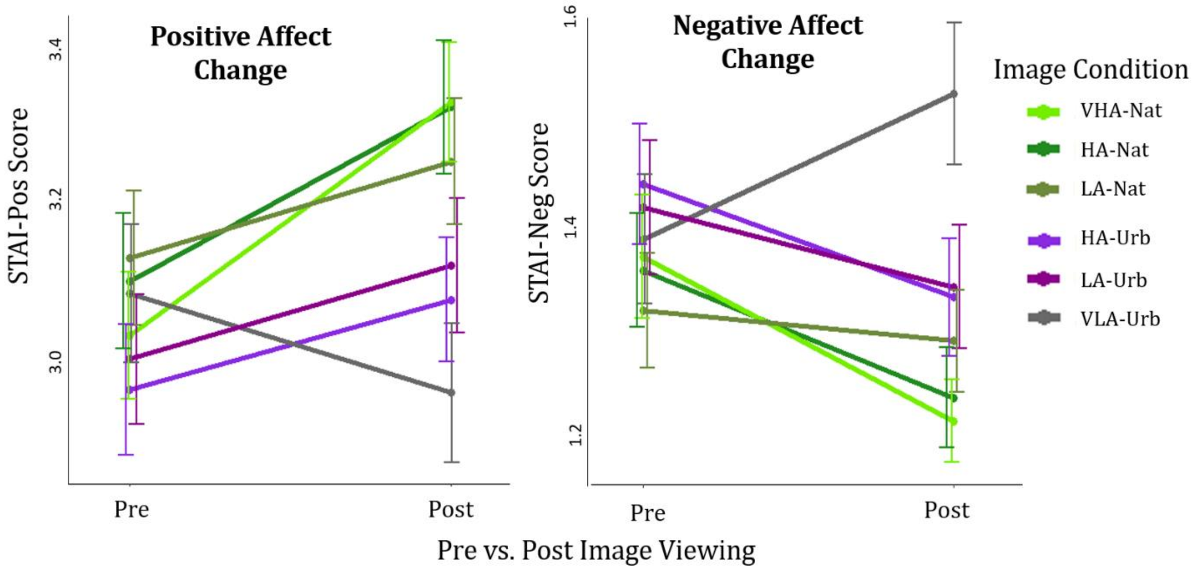


Figure 2.2. Study 2 Change in Affect by Image Condition
Average change in STAI-Pos (Left Panel) and STAI-Neg (Right Panel) between baseline (pre/T1) and after image viewing (post/T2) for each of the 6 image conditions. Error bars represent SEM.

2.3.3. Analysis 3: Affect Change in Completely Crossed Data Subset (Aesthetic Value as Individual Difference Measure)

Analysis 3 was also conducted using only the data with completely crossed conditions, but using participants' own average ratings for the images they viewed, rather than treating aesthetic value as a factor. To do this, two multiple regressions were performed predicting STAI-Pos and STAI-Neg by environment type, average image rating, and their interaction.

For STAI-Pos the overall model was not significant ($R^2 = 0.015$, $F(3,406) = 2.07$, $ps = 0.10$), though participants' average preference ratings were significantly predictive of change in STAI-Pos ($B = 0.07$, 95% CI [0.01, 0.14], $p = 0.03$). However, neither environment type nor the interaction of environment and individual preference ratings were significant (all $p > 0.41$). For STAI-Neg, the overall model was also not significant ($R^2 = 0.005$, $F(3,406) = 0.70$, $ps = 0.55$) and none of the predictors had a significant effect on the outcome variable (all $p > 0.49$).

2.3.4. Analysis 4: Affect Change between Aesthetic Value Conditions within an Environment

To handle the conditions which were not completely crossed in our design (VHA-Nat and VLA-Urb), separate factorial ANOVAs were conducted on each of the three aesthetic value levels for nature and urban images.

Nature Conditions

A one-way ANOVA with the 3 aesthetic value levels (Very High, High, Low) in the nature condition was performed for change in STAI-Pos and STAI-Neg. Results of this ANOVA for STAI-Pos yielded a trending effect of aesthetic value level $F(2, 306) = 2.88$, $p = 0.057$, $\eta_p^2 = 0.018$, 95% CI [0.0, 0.054]. The partial eta-squared indicates a small effect size. Post hoc comparisons were conducted and family-wise error corrected using Tukey's HSD, which showed a significant difference between the VHA-Nat ($M = 0.29$, $SD = 0.54$) and LA-Nat conditions ($M = 0.12$, $SD = 0.52$, $p = 0.045$, $d = 0.34$, 95% CI [0.06, 0.61]), indicating a greater increase in STAI-Pos for those in the VHA-Nat condition relative to those in the LA-Nat condition. However, there were no differences between VHA-Nat and HA-Nat (*M* = 0.22, *SD* = 0.48, *ps* = 0.55) or between HA-Nat and LA-Nat (*ps* = 0.37).

Results of this analysis for STAI-Neg showed a significant effect of aesthetic value, *F*(2,306) = 3.27, *p* = 0.039, η_p^2 = 0.021, 95% CI [0.0, 0.058]. The partial eta-squared indicates a small effect size. Post hoc comparisons showed a significant difference between VHA-Nat (*M* = -0.16, *SD* = 0.38) and LA-Nat (*M* = -0.03, *SD* = 0.42, *p* = 0.036, *d* = 0.32, 95% CI [0.05, 0.60]). This difference indicates a greater reduction in STAI-Neg for participants in the VHA-Nat condition relative to LA-Nat. No significant difference was found between VHA-Nat and HA-Nat (*M* = -0.12, *SD* = 0.32, *ps* = 0.77) or between HA-Nat and LA-Nat (*ps* = 0.17) [**Figure 2.2**].

Urban Conditions

For the urban images, a one-way ANOVA with the 3 aesthetic values (High, Low, Very Low) was also performed for change in STAI-Pos and STAI-Neg. Results of the first ANOVA indicated a significant effect of aesthetic value level on STAI-Pos, *F*(2, 303) = 6.08, *p* = 0.003, η_p^2 = 0.039, 95% CI [0.005, 0.086]. Here, the value of the partial eta-squared indicated a small-to-medium effect size. Post hoc comparisons were conducted and family-wise error corrected using Tukey's HSD, which showed a significant difference between the VLA-Urb (*M* = -0.12, *SD* = 0.59) and LA-Urb conditions (*M* = 0.12, *SD* = 0.63, *p* = 0.007, *d* = 0.40, 95% CI [0.12, 0.68]) as well as between VLA-Urb and HA-Urb (*M* = 0.11, *SD* = 0.47, *p* = 0.008, *d* = 0.44, 95% CI [0.17, 0.72]), indicating that participants in the VLA-Urb condition showed less of an improvement in STAI-Pos compared to HA-Urb or LA-Urb. No difference was found for STAI-Pos change between HA-Urb and LA-Urb (*ps* = 0.99).

Results of this analysis for STAI-Neg also showed a significant effect of aesthetic value, $F(2,303) = 8.37$, $p < 0.001$, $\eta_p^2 = 0.052$, 95% CI [0.012, 0.104]. The partial eta-squared indicates a medium effect size. Post hoc comparisons yielded a significant difference between VLA-Urb ($M = 0.14$, $SD = 0.59$) and LA-Urb ($M = -0.08$, $SD = 0.42$, $p = 0.003$, $d = 0.42$, 95% CI [0.15, 0.7]), as well as between VLA-Urb and HA-Urb ($M = -0.11$, $SD = 0.39$, $p < 0.001$, $d = 0.5$, 95% CI [0.22, 0.78]), but not between LA-Urb and HA-Urb (*ps* = 0.88). These results suggest a larger reduction in STAI-Neg for participants in the HA-Urb and LA-Urb conditions compared to those in the VLA-Urb condition. [**Figure 2.2**]

2.3.5. Analysis 5: Affect Change as Predicted by Individual Preference Ratings in full Dataset

To examine whether individual differences in participants' preference ratings for the images were related to changes in affect, two linear regressions were conducted to predict change in STAI-Pos and STAI-Neg by individuals' average preference rating.

Analyses conducted on the full dataset (not only on the completely-crossed conditions) showed that average image preference rating explained 5.5% of the variance in STAI-Pos change ($R^2 = 0.055$, $F(1,613) = 35.92$, $p < 0.001$). In this case, a higher average preference rating for the images viewed significantly predicted a greater increase in STAI-Pos ($\beta = 0.24$, 95% CI [0.16, 0.31], $p < 0.001$). Additionally, average image preference explained 2.4% of the variance in STAI-Neg change ($R^2 = 0.024$, $F(1,613) = 14.76$, $p < 0.001$). Here, higher preference ratings significantly predicted a greater decrease in STAI-Neg ($\beta = -0.15$, 95% CI [-0.23, -0.07], $p < 0.001$).

2.4. Discussion

Study 2 failed to demonstrate a significant effect of environment type on changes in either positive (STAI-Pos) or negative (STAI-Neg) affect for stimuli where preference was equated. That is, the HA-Nat and HA-Urb were not significantly different from one another, nor were the LA-Nat and LA-Urb (Analysis 2). There was modest evidence for individuals' own preference ratings as a predictor of positive but not negative affect change in these completely crossed conditions (Analysis 3). However, the overall experimental design included conditions which were not completely crossed (i.e. VHA-Nat and VLA-Urb). When examining differences in aesthetic value within an environment type, these more extreme aesthetic value conditions (VHA and VLA) yielded significantly larger changes in affective state compared to the HA or LA conditions. Specifically, VHA-Nat lead to greater improvements in both positive and negative affect relative to LA-Nat, and VLA-Urb lead to worsened positive and negative affect relative to both LA-Urb and HA-Urb (Analysis 4). Furthermore, participants' own preference ratings of the images were significantly predictive of change in both the positive and negative affect change when analyzed in the full dataset (Analysis 5).

While the greatest positive affect changes due to image condition were found in the VHA-Nat condition, Study 1 did not yield a sample of urban images that were as highly preferred as these nature stimuli to create a comparable condition in Study 2. Studies 3 and 4 were designed to address this problem by finding other image types that were as preferred as the VHA-Nat condition, but qualitatively different in context from the nature scenes.

Study 3: Very High Aesthetic Images Validation

3.1. Study Intro

To examine whether the significant changes in state affect found for the Very High Aesthetic value Nature condition (VHA-Nat) were due to the images being very highly preferred or due to something else specific to nature (over and above aesthetic preference), additional image sets were required that would include multiple categories of images also rated very highly. Study 3 was conducted to generate the stimulus sets needed to compare against the VHA-Nat condition. Another environmental condition (i.e. very high aesthetic urban) was the most relevant category, but highly preferred images in other, qualitatively different contexts were also examined in Study 3: Animals, Space, and Abstract Art.

3.2. Method

3.2.1. Participants

409 US-based adults (202 male, 206 female, 1 not reported) were recruited from Amazon Mechanical Turk. Age of the participants ranged from 19 to 74 years (*M* = 36.9, *SD* = 10.7). The full study procedures were expected to take approximately 15 minutes and participants were compensated $1.50 for participating. Informed consent was administered by the University of Chicago Institutional Review Board (IRB).

3.2.2. Procedure

Participants were randomly assigned to rate 75 images on a 1-7 Likert scale evaluating each photo on preference (anchors of 1 = "strongly dislike" and 7 = "strongly like"; n = 206) or naturalness (anchors of 1 = "very man-made" and 7 = "very natural"; n = 203). Each participant saw 15 images across five categories of images: animals, space, abstract patterns, natural environments, and urban environments. Images were pulled randomly, and each image received a minimum of 25 ratings on each attribute with an average of roughly 40 ratings per image.

3.3. Results

Average preference ratings for the 75 images in each category are plotted in **Figure 2.3 (Panel A)**. Based on these ratings, only two categories of images (Animals and Urban) had sufficiently overlapping distributions to extract preference-matched image sets with the VHA-Nat images. The images sets used for

Study 4 involved 30 images each of nature, urban, and animal images with similar average preference ratings and standard deviations (**Figure 2.3; Table 2.4**).

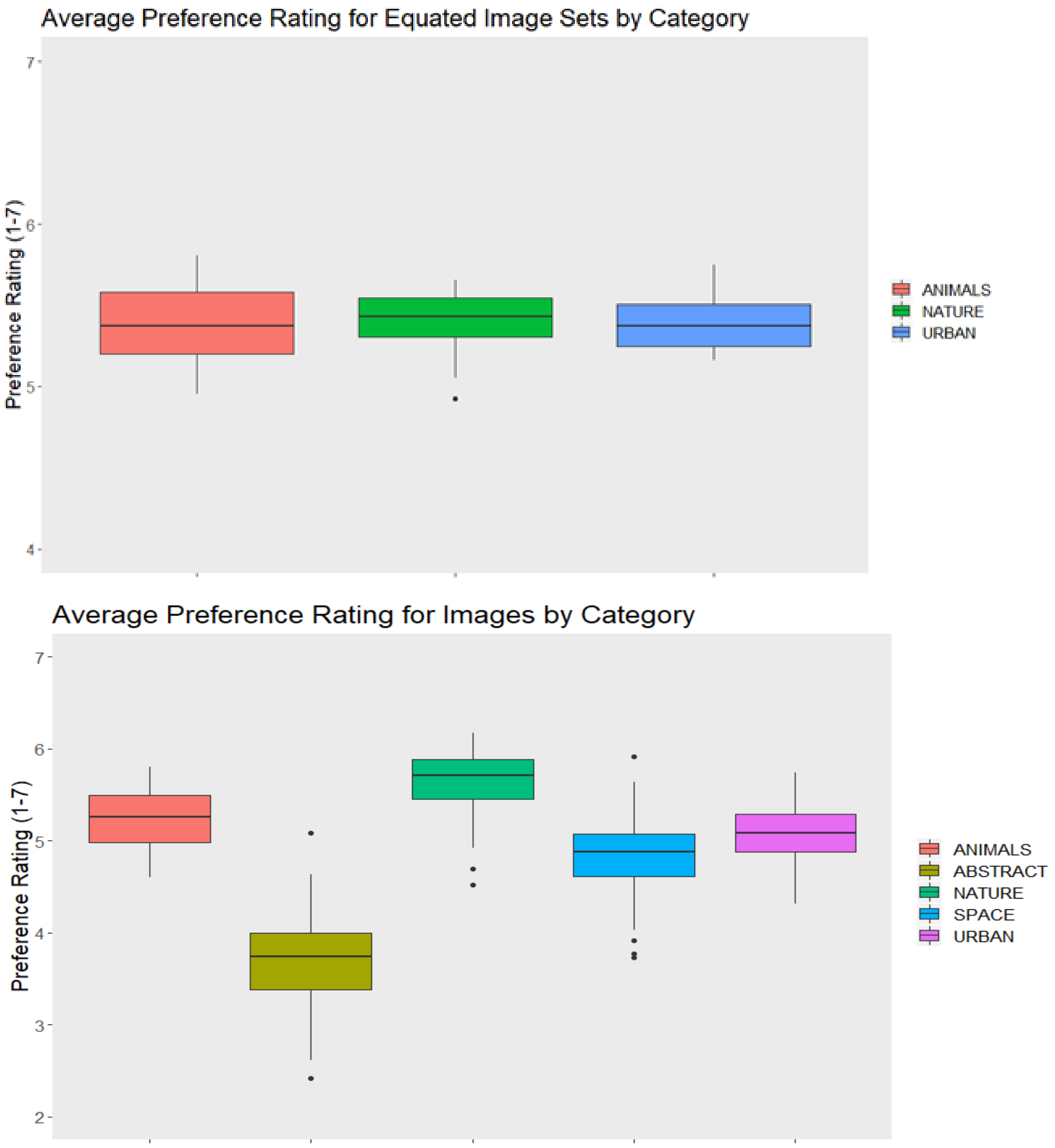


Figure 2.3. Image preference results from Study 3
Top Panel. Boxplots of preference-equated image sets (30 images each) in the Nature, Urban, and Animal image categories. Abstract images and images of space were significantly lower in preference and did not yield enough preference-equated images to create another stimulus set.
Bottom Panel. Boxplots of preference ratings for all 75 images in each category examined.

Table 2.4. Study 3 Ratings for the Preference-equated Image Sets
Based on the participants' ratings in Study 3, three categories of preference-equated images were used in Study 4. Average preference and naturalness ratings for these image sets (30 images each), presented below.

	Nature	*Urban*	*Animals*
Preference M (SD)	5.41 (0.18)	5.39 (0.16)	5.39 (0.26)
Naturalness M (SD)	6.52 (0.21)	2.73 (0.36)	6.47 (0.23)

Study 4: Examining Affect Change in Very Highly Preferred Stimuli

4.1. Study Intro

Study 4 sought to determine whether very highly preferred nature images would cause the same improvement in state affect as equally preferred images from other categories. That is, Study 4 tested whether the nature category itself was an additional source of affective benefit above and beyond aesthetic preference. It's worth clarifying that though animals are 'natural', the previous literature examining nature's cognitive and affective benefits focuses on natural environments/scenes, and therefore animals (that are not part of a natural scene) are categorically different from the types of nature of interest here.

In addition, Study 4 employed a negative mood induction procedure (MIP) in half of the participants to examine if baseline mood might impact how effective the VHA images are at improving affective state. For example, it might have been the case that as participants did not have high baseline negative affect in Study 2, they may have been less affected by the stimuli. In Study 2, the average STAI-Neg rating at baseline was 1.3 (on a 1-4 scale), the baseline average score for VAS Sad was 11, and for VAS Angry it was 6 (both on a 1-100 scale). Positive affect also started out relatively high in Study 2 (average STAI-Pos was 3.0, average VAS Happy was 62, and average VAS Inspired was 45). As such, in Study 4, the Negative MIP was included to ensure that any effects (or lack thereof) were not simply due to ceiling/floor effects of baseline affect.

Two empirical questions were examined in Study 4. The main question was whether the improvement in affect found for VHA-Nat in Study 2 was due to nature itself or simply due to preference. The secondary question was whether or not baseline affect influenced whether nature had an additional emotional benefit above and beyond preference. A visual depiction of the hypotheses for this study as well as the analyses chosen to address each hypothesis are presented in **Figure 2.5**.

4.2. Method

Study 4 was pre-registered on OSF prior to data collection: https://osf.io/u5r4c. The pre-registration included a PCA for data reduction (results of which can be found in Supplementary materials in Appendix C), but due to lessened interpretability of this approach (as in Study 2), the analyses presented diverge from the pre-registered ones. The full data for Study 4 are publicly available at: https://osf.io/ehtk9/.

4.2.1. Participants

602 US-based adults (271 male, 327 female, 1 other, 3 not reported) were recruited from Amazon Mechanical Turk. Age of the participants ranged from 19 to 69 years (*M* = 37.5, *SD* = 11.0). The full study procedures were expected to take approximately 15 minutes and participants were compensated $1.50 for participating. Informed consent was administered by the University of Chicago Institutional Review Board (IRB). The sample size was determined prior to data collection (see pre-registration) and was partially based on recreating the conditions of Study 2.

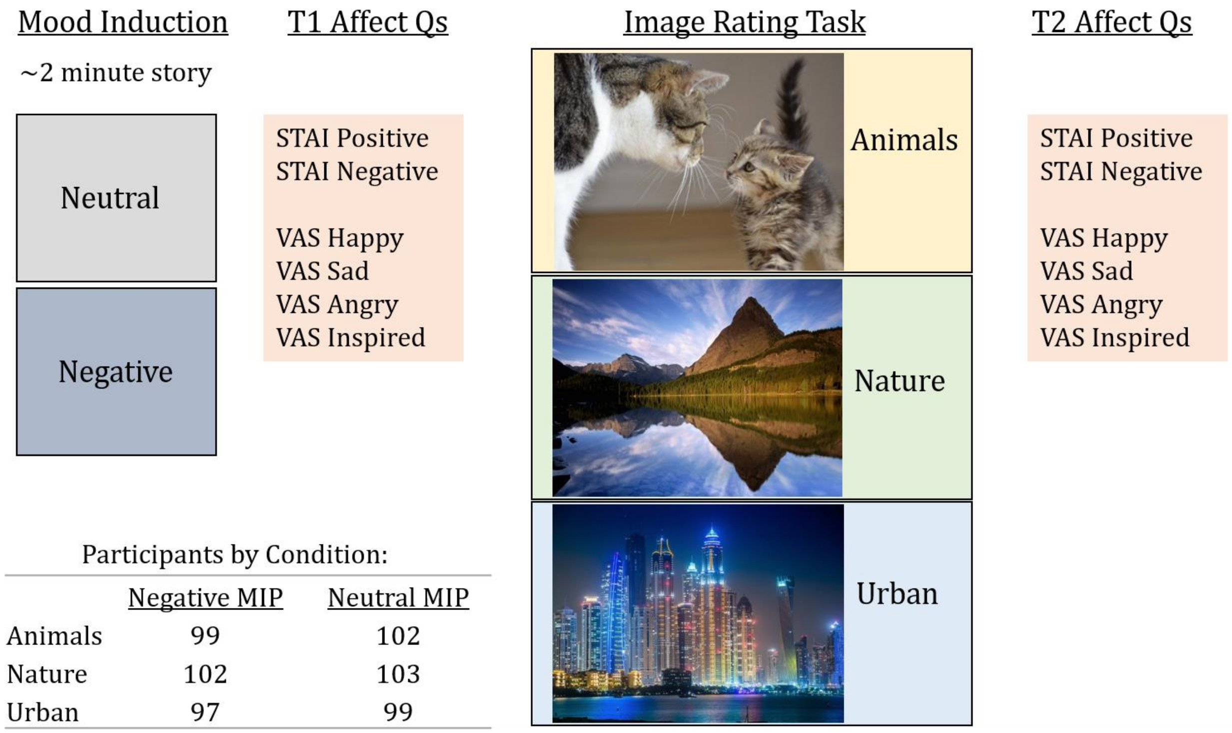
4.2.2. Experimental Conditions

Participants were randomly assigned to one of 3 image conditions based on the results of Study 3 (Anim, Nat, Urb) and one of 2 mood induction procedure groups (Negative, Neutral). The study design was fully between subjects with 6 conditions: Anim-Negative (n = 99), Anim-Neutral (n = 102), Nat-Negative (n = 102), Nat-Neutral (n = 103), Urb-Negative (n = 97), Urb-Neutral (n = 99). The full image sets and ratings (validated in Study 3) can be accessed at this link: https://osf.io/ehtk9/

4.2.3. Procedure

After informed consent was obtained, participants first completed the mood induction procedure (MIP). Participants were provided with a brief description prior to reading a short story: “In this study we would like to examine how different types of storytelling influence, thoughts, feelings, and judgments. On the next page we have a short story for you to read. The story will be on screen for a minimum of 2 minutes. Please try to minimize distraction and take your time reading the story completely before continuing." The negative story was taken from a report of a young woman whose father died after suffering from Alzheimer's dementia, and the neutral story was a short excerpt from the introductory chapter of *A Brief History of Time* (Hawking, 1988). The two stories were validated in a previous study evaluating the efficacy of several text-based online mood inductions (Verheyen & Göritz, 2009).

Following the MIP, participants filled out baseline measures of their affective state (T1). The measures used were identical to those collected in Study 2. Participants then viewed a series of 30 images and gave ratings of their aesthetic preference for the images on a 1-7 scale (1 = strongly dislike, 7 = strongly like). Each image was on the screen for a minimum of 7 seconds, and the next image would appear after a rating for the image was selected. After the image rating task, participants completed the same measures of their affective state a second time (T2). [See **Figure 2.4** for study design]



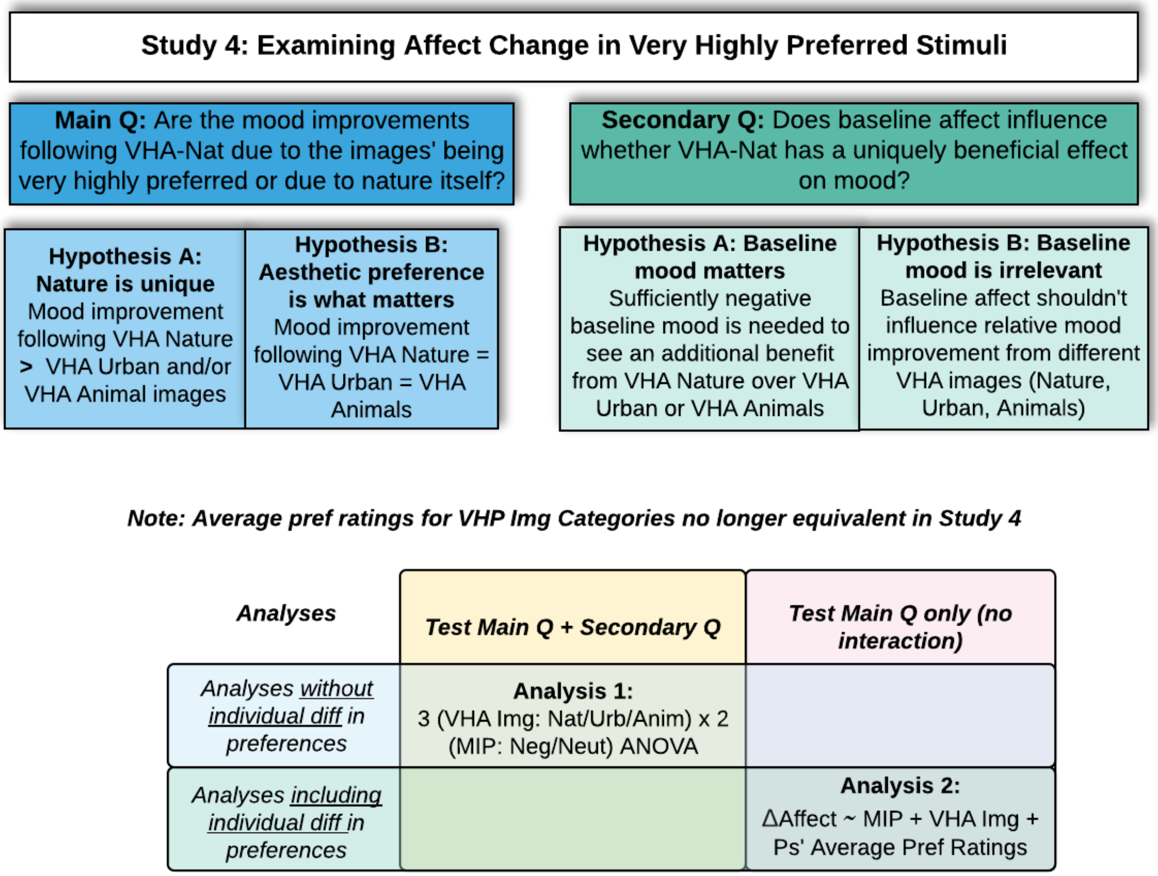
Participants by Condition:

	Negative MIP	Neutral MIP
Animals	99	102
Nature	102	103
Urban	97	99

Figure 2.4. Study 4 design
Mood induction type (negative vs. neutral) and Image category condition (animals, nature, or urban) were between-subjects factors.

4.2.4. Analysis

A visual depiction of the analytic approach for this study is presented in **Figure 2.5**. The main question of Study 4 was whether the improvement in affect found for VHA-Nat in Study 2 was due to nature itself (Hypothesis A: *Nature is unique*) or simply due to preference (Hypothesis B: *Aesthetic Preference is what matters*). The secondary question was whether baseline affect influenced whether nature has an additional emotional benefit above and beyond preference (Hypothesis A: *Baseline mood matters*) or isn't important (Hypothesis B: *Baseline mood is irrelevant*). The secondary question examined an interaction between preference-equated categories and mood induction. However, in Study 4, participants’ ratings diverged somewhat from 3, resulting in significantly different preference ratings between categories (described below, see **Table 2.5** and **Figure 2.6**). This issue prevented a strong test of the main effect of preference-equated category, as well as the interaction between MIP and image category in the ANOVA used in Analysis 1. Therefore, linear regression (Analysis 2) which tested the independent predictive value of mood induction procedure, image category, and participants’ own preference ratings was conducted as well. The same analysis tools in ‘R’ and relevant statistical procedures used in Study 2 were also employed in Study 4.



Analyses	*Test Main Q + Secondary Q*	*Test Main Q only (no interaction)*
Analyses without individual diff in preferences	**Analysis 1:** 3 (VHA Img: Nat/Urb/Anim) x 2 (MIP: Neg/Neut) ANOVA	
Analyses including individual diff in preferences		**Analysis 2:** ΔAffect ~ MIP + VHA Img + Ps' Average Pref Ratings

Figure 2.5. Hypotheses & Analyses Overview for Study 4

Participants’ preferences for the images they viewed were determined by taking the average of their ratings for all 30 images they viewed. The average and standard deviation for each picture condition across all participants in that condition are presented in **Table 2.5.** Ratings from participants in this study (Study 4) were somewhat different from those in image validation Study 3 [**Figure 2.6**], and our image categories were no longer fully equated on preference.

Table 2.5. Study 4 participants’ preference ratings.
Mean & SD by picture condition, collapsed across MIP conditions.

	Nature	*Urban*	*Animals*
Preference M (SD)	5.69 (0.85)	5.16 (0.96)	5.89 (0.88)

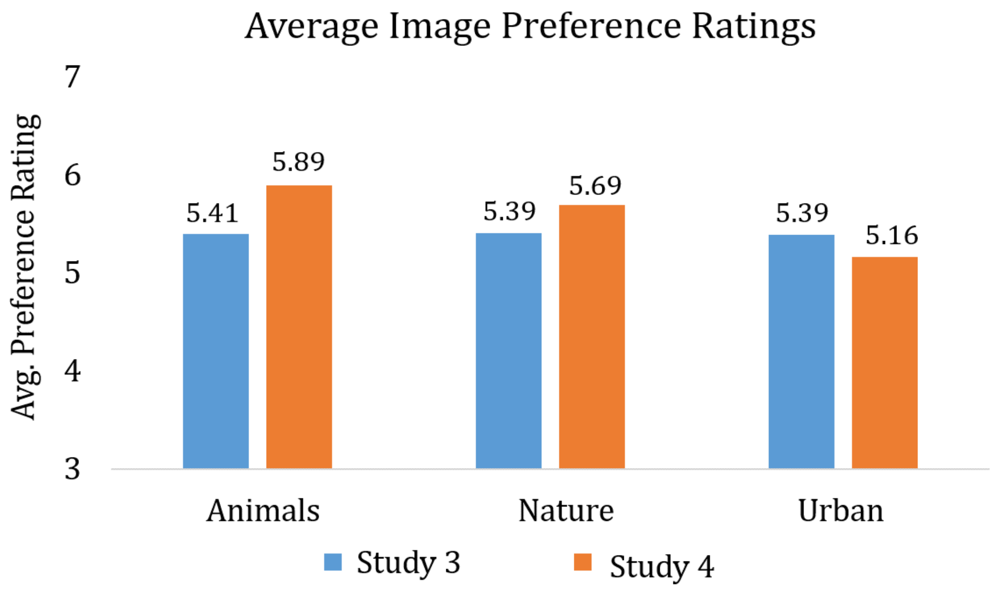


Figure 2.6. Average preference ratings by category for Study 3 and 4

4.3. Results

4.3.1. Analysis 1: Affect Change by Image Condition and MIP

To test the effect of image category and mood induction on affect change, a factorial ANOVA with VHA Image Category (Nature, Urban, Animal) and Mood Induction (Negative, Neutral) as between-subjects variable were conducted on change (T2 minus T1) for STAI-Neg and STAI-Pos.

Results of this ANOVA for STAI-Pos yielded a main effect of Mood Induction $F(1,596) = 152.6$ $p < 0.001$, $\eta_p^2 = 0.21$, 95% CI [0.15, 0.26]. The partial eta-squared indicates a large effect of MIP Condition. Participants who were first inducted into a negative mood showed a larger increase in STAI-Pos ($M = 0.85$, $SD = 0.81$) relative to those in the neutral MIP ($M = 0.16$, $SD = 0.51$). There was not a significant effect of VHA Image Category ($ps = 0.34$) or an interaction of VHA Image Category and MIP ($ps = 0.9$).

Similar results were found for STAI-Neg. A main effect of MIP was found ($F(1,596) = 246.0$, $p < 0.001$, $\eta_p^2 = 0.30$, 95% CI [0.24, 0.35], where again, participants first induced into a negative affective state showed a greater reduction in negative affect ($M = -0.81$, $SD = 0.76$) relative to those in the neutral mood induction ($M = -0.04$, $SD = 0.37$). The partial eta-squared indicates a large effect of mood induction type on change in STAI-Neg. No significant effect of VHA Image Category ($ps = 0.87$) or interaction of VHA Image Category and MIP ($ps = 0.39$) was found.

4.3.2. Analysis 2: Affect Change as Predicted by Image Condition, Mood Induction, & Individual Preference Ratings

To see whether individual differences in participants' preference ratings influenced change in affect, multiple regressions were conducted. With STAI-Pos and STAI-Neg as outcome variables, the regression analyses examined the respective contributions of mood induction procedure, VHA image category, and individuals' average preference ratings.

Results of this analysis on STAI-Pos are presented in **Table 2.6 (Panel A)**. The overall model was significant and explained 23.6% of the variance in change in positive affect ($R^2 = 0.236$, $F(4,597) = 46.11$, $p < 0.001$). In this case, a higher average preference rating for the images viewed significantly predicted a greater increase in STAI-Pos ($B = 0.14$, 95% CI [0.08, 0.20], $p < 0.001$). Being induced into a negative mood was also a significant predictor ($B = -0.67$, 95% CI [-0.78, -0.56], $p < 0.001$), but image category did not significantly contribute to positive affect change [**Figure 2.7**].

Results of this analysis on STAI-Neg are presented in **Table 2.6 (Panel B).** As in positive affect, the overall model was significant and explained 30.8% of the variance in STAI-Neg change ($R^2 = 0.308$, $F(4,597) = 66.36$, $p < 0.001$). A higher average preference rating for the images viewed significantly predicted a greater reduction negative affect ($B = -0.09$, 95% CI [-0.14, -0.03], $p = 0.001$), as did being in the negative mood induction group ($B = 0.77$, 95% CI [0.67, 0.86], $p < 0.001$).

Importantly, image category did not significantly contribute to change in STAI-Neg [**Figure 2.7**].

Table 2.6. STAI change regressions

Panel A. Regression results using STAI-Pos Change as the criterion

Predictor	*b*	*b* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	Fit
(Intercept)	0.05	[-0.32, 0.42]			
Avg_Pic_Rate	0.14**	[0.08, 0.20]	.03	[.00, .05]	
Pic_Con_Nat	-0.02	[-0.15, 0.11]	.00	[-.00, .00]	
Pic_Con_Urb	0.01	[-0.13, 0.14]	.00	[-.00, .00]	
MIP_Con	-0.67**	[-0.78, -0.56]	.19	[.14, .25]	
					R^2 = .236** 95% CI [0.18, 0.29]

Panel B. Regression results using STAI-Neg Change as the criterion

Predictor	*b*	*b* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	Fit
(Intercept)	-0.29	[-0.62, 0.04]			
Avg_Pic_Rate	-0.09**	[-0.14, -0.03]	.01	[.00, .03]	
Pic_Con_Nat	-0.03	[-0.14, 0.09]	.00	[-.00, .00]	
Pic_Con_Urb	-0.09	[-0.22, 0.03]	.00	[-.00, .01]	
MIP_Con	0.77**	[0.67, 0.86]	.29	[.23, .35]	
					R^2 = .308** 95% CI [0.26, 0.36]

A significant b-weight indicates the semi-partial correlation is also significant. b represents unstandardized regression weights. sr2 represents the semi-partial correlation squared. LL and UL indicate the lower and upper limits of a confidence interval, respectively.

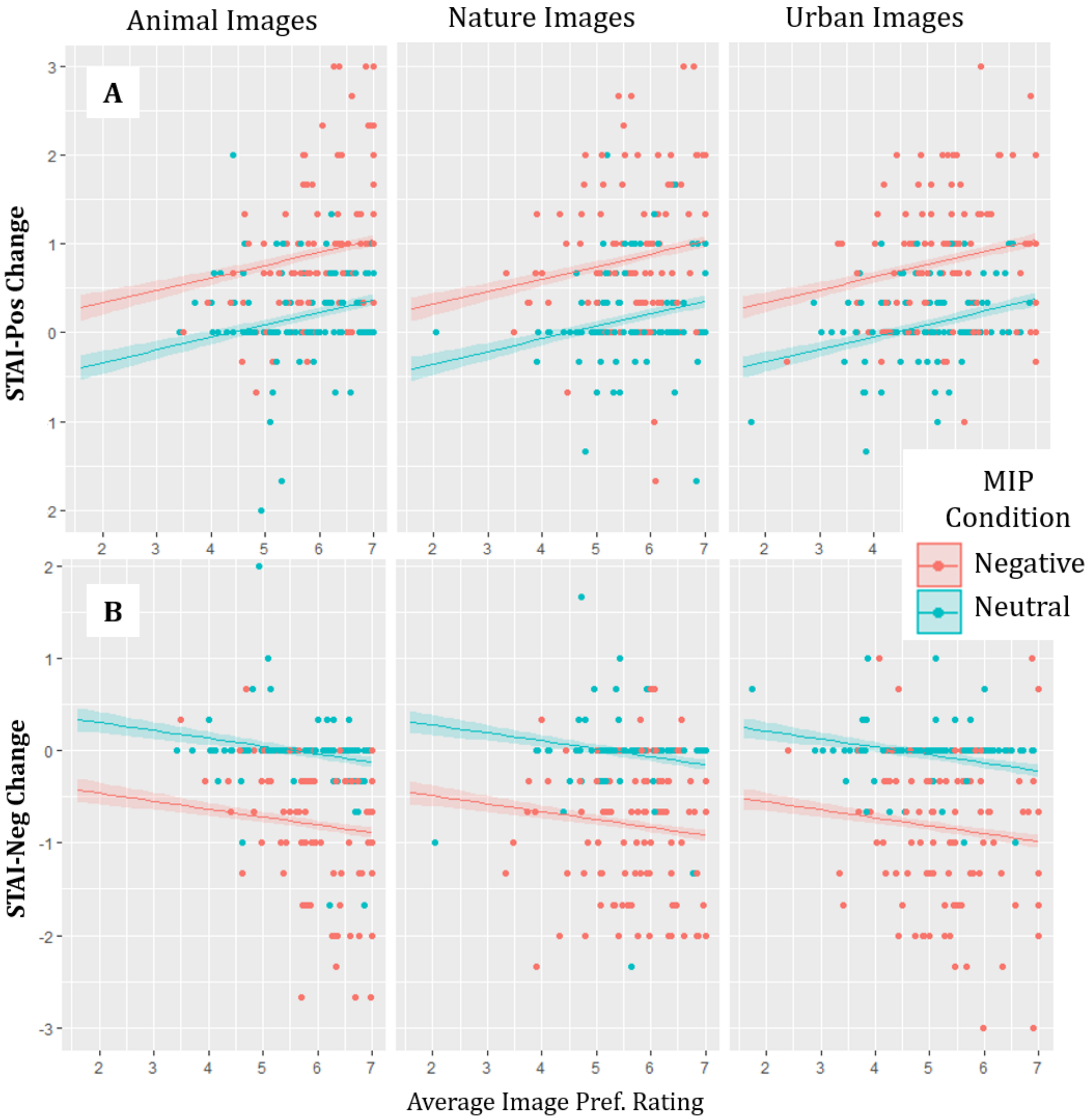


Figure 2.7. Multiple Regression Plots for Affect Change
Regression plots predicting change in STAI-Pos (**A**) and STAI-Neg (**B**) by VHA Image Category + Participants' Average Image Rating + Mood Induction (MIP) Condition

4.4. Discussion

The results of Study 4 were consistent with those of Study 2. Study 4 found that the very highly preferred image category (in this case, nature, urban, or animals) did not have a differential impact on affect change, i.e., they all improved affect to the same degree. Individual differences in how much participants liked the images they saw did predict improvement in both STAI-Pos and STAI-Neg which was also found in Study 2. Participants who were first induced into a negative mood showed larger changes in affect but this did not interact with VHA image category. Thus, the results of both studies provide support for a preference-based account of mood change rather than the effects being specific to nature stimuli (i.e., nature-based mood effects do not differ from other very highly preferred stimulus categories).

Study 5: Examining the Relationship between Beauty and Affinity

5.1. Study Intro

In Studies 2 and 4, affect change was driven by differences in preference rather than environment/category. In each of these studies, participants had evaluated preference on a scale that assesses affinity for the images (i.e., how much do you like/dislike the image). Up to this point, we have been assuming that preference (affinity) and aesthetics (beauty) are the same construct. Indeed, the terms aesthetics, affinity, and preference, are often used interchangeably (Staats et al., 2003; Ulrich, 1983; van den Berg et al., 2003). However, it is still possible that there is something special (i.e., rewarding, pleasing, or affinity-inducing) about natural environments above and beyond aesthetics (beauty) that causes them to be preferred (liked). For example, Valtchanov and Ellard (2014) propose that natural stimuli are endogenously visually rewarding. If this is the case, ratings of beauty and ratings of affinity (liking) may not be identical overall or may be different when examined in nature scenes versus in urban scenes. To test this, all 375 images rated on affinity (Study 1) were also rated on aesthetics/beauty in a new sample (Study 5). If the affinity and beauty ratings are not identical in these images, this would suggest that participants' affinity ratings in Study 1 are due to something other than aesthetics (i.e. endogenous visual reward not captured by perceived beauty). Further, if the nature images and urban images differ in how correlated affinity and beauty are, this might suggest something categorically different in how participants evaluate nature images and urban images. However, if the two are highly correlated across all images and within category (nature vs. urban) this would suggest that the preference ratings are primarily evaluations of aesthetics.

5.2. Method

Study 5 was pre-registered on OSF prior to data collection: https://osf.io/u2e6n, though the analyses reported in this paper were not initially detailed in this pre-registration. The full data for Study 5 are publicly available at: https://osf.io/ehtk9/

5.2.1. Participants

194 US-based adults (94 male, 100 female) were recruited from Amazon Mechanical Turk. This sample does not include 9 workers who provided feedback to indicate they encountered technical problems or were distracted during the study, or showed no variation in responding (i.e., gave the same rating to every image). Age of the participants ranged from 19 to 72 years (M = 38.3, SD = 12.4). The full study procedures (including additional tasks after collecting the data for this study) were expected to take approximately 15 minutes and participants were compensated $1.50 for participating. Informed consent was administered by the University of Chicago Institutional Review Board (IRB).

5.2.2. Procedure

Study 5 was designed to match the conditions of the original image rating procedure in Study 1 as closely as possible. Therefore, all 375 images from Study 1 were used in Study 5. Each image in Study 1 was rated on preference (affinity, i.e. ”How much do you like or dislike this image?”) by roughly 51 individuals, and each participant saw 100 images in a session. As in Study 1, participants in Study 5 saw 100 images pulled pseudorandomly, attempting to show a relatively equal number of natural and urban images as well as images that varied on aesthetic value. In this study, each image received a minimum of 35 beauty ratings (i.e., “How ugly or beautiful is this image?”) with an average of roughly 51 ratings per image.

5.2.3. Analysis

To test whether participants' ratings of aesthetics (beauty) in Study 5 were similar to affinity ratings from Study 1, correlations between images' average beauty ratings (Study 5) and affinity ratings (Study 1) were conducted using the 'cor' function in R. The analyses specified in the pre-registration (testing for differences in beauty in image sets that were matched on preference) were also conducted initially and are reported in the supplementary materials, but ultimately the correlation analyses in the full image dataset were more informative, so they are the primary results reported.

5.3. Results

When examined across all 375 images, the correlation between beauty ratings and affinity ratings was $r = 0.97$ [**Figure 2.8**]. Additionally, when examined within environment type, the correlations were similar in magnitude. For nature images this correlation was $r = 0.96$ and for urban images the correlation was $r = 0.95$.

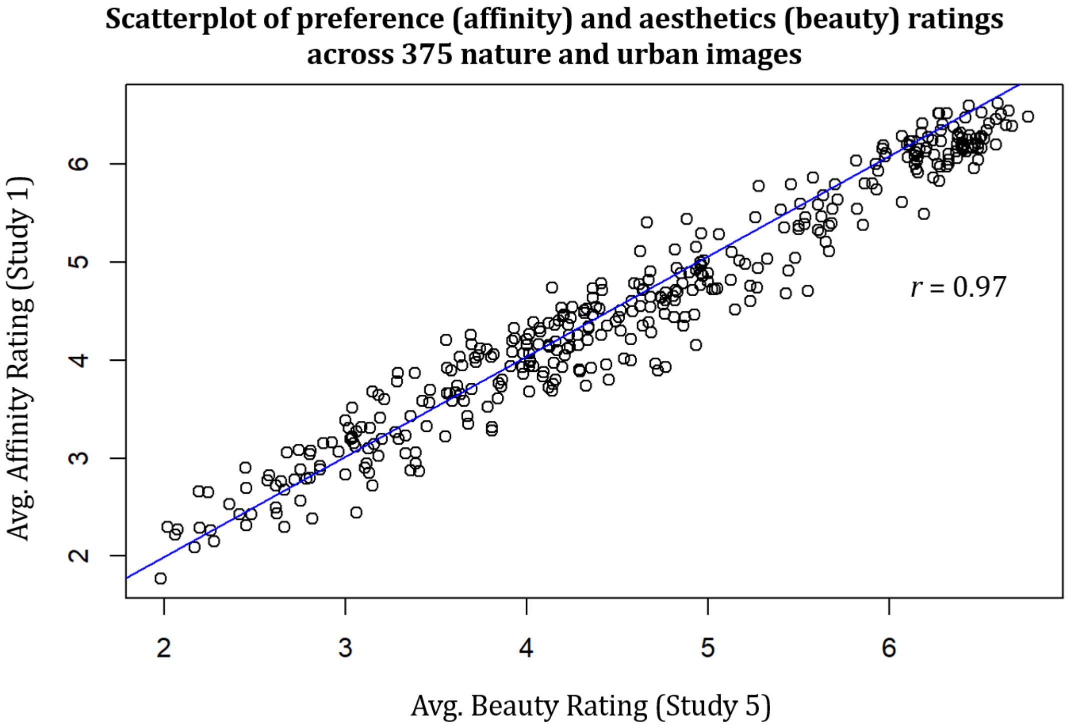


Figure 2.8. Correlation between preference and aesthetics

5.4. Discussion

The results of Study 5 failed to find support for the hypothesis that there is something special, unique, or "rewarding" (Valtchanov & Ellard, 2015) about the nature scenes in our study which make them preferred (liked) above and beyond aesthetics. These results demonstrated a near perfect correlation between affinity and aesthetics for the full sample of images, as well as when broken up by environment type. Therefore, our data do not suggest that there is some missing factor that is specific to natural environments which makes them preferred (liked) in our study.

General Discussion

Nature interactions reliably elicit positive changes in affect (Berman et al., 2012; Bratman, Daily, et al., 2015; MacKerron & Mourato, 2013; McMahan & Estes, 2015), however, the underlying mechanism remains unknown. Though many researchers have demonstrated the robust impact of nature on emotions, much of this previous work has not controlled for preference when examining mood effects. The goal of the present project was to clarify whether there is something unique about the affective benefits of nature stimuli over and above individuals' preference for these stimuli. Across several studies, consistent evidence for a preference-based account of affect change was found. That is, nature seems to have a positive effect on emotional state because it is highly preferred. If nature is sufficiently low on aesthetics or compared to an equally preferred urban image, there is no additional benefit of nature on affective state. Additionally, even nature very high on aesthetics does not elicit larger emotional responses than other equally preferred stimuli, suggesting there is not an additional benefit to affective state of viewing natural scenery per se.

Study 2 demonstrated that, once equated on preference, there were no significant differences in affect change between nature and urban environments. However, aesthetic value, as measured by participants' own ratings or by pre-established conditions within an environment type, did predict the extent to which participants' affective state improved post-picture viewing. In Study 2, the largest condition-level effects were found for images in the most extreme aesthetic value conditions (VHA-Nat and VLA-Urb), which were not completely crossed with environment type.

To overcome this, Study 4 used very high aesthetic value images in categories other than natural environments, and tested whether the improved effect on affect after VHA-Nat in Study 2 was due to the high aesthetic value or to the environment category nature itself. For these purposes, urban scenes and animal images with very high preference ratings were utilized. When comparing change in affect before and after image viewing, very high aesthetic nature did not have a larger effect than the animal or urban images. This finding was unaffected by whether participants had been induced to a negative mood state at baseline. Further, the results of multiple regression analyses, which examined both participants' average preference ratings and image category, showed that while rating the images as more highly preferred was significantly predictive of affect change, image category did not have a significant effect.

Lastly, Study 5 was conducted to address the possibility that the preference measure used in Study 1 and 2 (affinity) captured something unique about nature above and beyond aesthetic preferences in our stimuli. This idea was not supported by the data, as explicit ratings of beauty were almost perfectly correlated with affinity ratings across all images.

The focus of the present research was to address whether there is something unique about natural environments that can lead to changes in affect even when preference is taken into account. Interestingly, the results of this research do not support that viewing nature scenes has an acute effect on affect that can be attributable to something beyond preference. However, it is important to note that the current study does not shed light on why, in general, natural environments are so highly preferred to begin with.

Decades of research have spawned theoretical accounts of the origins of nature preferences. These include evolutionary theories such as Biophilia (Kellert & Wilson, 1995), and Stress Reduction Theory (Ulrich et al., 1991), which propose that because our evolutionary history took place in predominantly natural environments, humans therefore feel an innate affinity towards nature. Other theories propose that the ease of processing visual features often found in natural environments (e.g., fractalness) causes nature to be preferred and causes a positive affective response (Perceptual Fluency Account; Joye et al., 2016). Still others propose that we prefer natural environments due to their potential restorative value (Hartig & Staats, 2006).

The current research cannot address why people have a preference for nature stimuli over other types of stimuli. Indeed, this preference might be a vital part of why nature is viewed as unique. From the image preference ratings obtained in Study 1, it was challenging to find urban environments that were as preferred as the high aesthetic nature images (HA-Nat) to be used in Study 2, and was not possible to find urban images to match the very high aesthetic (VHA) nature stimuli. Similarly, in Study 3, only two (animals, urban scenes) of four other categories examined yielded sufficiently overlapping preference distributions to use in Study 4. Additionally, it's noteworthy that though the content of the animal images are quite different from nature scenery, this category is indeed comprised of natural things. It is likely that, because of this lack of preference overlap, many studies examining the effects of nature interventions have used images or videos that were not similarly preferred. For example, in the 2008 study conducted by Berman and colleagues (Berman, Jonides, & Kaplan, 2008), the average preference rating for the nature images used (on a 1-7 Likert scale) was approximately 5.5, whereas for urban images it was approximately 2.8. Therefore, it is worth noting that though the current data suggest that the affective benefits are only due to preference and not due to anything unique to nature scenes, in some sense, nature is a "special" kind of stimulus due to it being so overwhelmingly preferred (at least among adults, please see Meidenbauer et al., 2019 for research examining environmental preferences in children).

The primacy of aesthetics in nature-elicited affect change has a number of notable implications. Perhaps one of the most important relates to the ongoing debate of the role of nature preferences in cognitive restoration. In particular, SRT (Ulrich et al., 1991) posits that the cognitive benefits of nature interactions occur due to changes in affective state and reductions in stress. The SRT framework would therefore predict that if one experienced natural and urban environments which were equally preferred, superior changes in cognitive performance would not be expected after nature exposure. In comparison, Attention Restoration Theory, or ART (Kaplan, 1995; Kaplan & Berman, 2010), does not assume that restoration relates to affect in any way. ART theorists instead focus on features of natural environments which restore directed attention resources while softly capturing involuntary attention, and do not propose that the extent to which a natural environment is preferred matters. Support for this comes from recent evidence that affective and cognitive benefits of nature are dissociable (Stenfors et al., 2019). Though this study demonstrates the difficulty in finding preference-equated environments that match nature preferences, a strong test of whether preference plays a role in the cognitive benefits of nature would be to compare objective performance on cognitive tasks before and after exposure to preference-equated nature and urban images, videos, or walks.

Another implication relates to the use of biophilic design in architecture and urban planning (Joye, 2007). Though the idea of designing buildings to contain nature-like features is not new (Alexander, 2002; Kellert, 2012; Salingaros, 1998), recent research has generated compelling evidence for the overlap between architectural aesthetics and naturalness (Coburn et al., 2019). Broadly speaking, there are many visual features common in natural environments which are also highly aesthetically preferred, such as fractalness or recursive complexity (Van den Berg et al., 2016), density of curved edges (Berman, Hout, et al., 2014), or color-related properties such as blue-green hue and high saturation diversity (Kardan, Demiralp, et al., 2015). The results of this study would suggest that, if one goal of biophilic architecture is to promote positive affective responses, design ought to prioritize inclusion of natural features which provide the most aesthetic value over those which may appear natural but not highly predictive of beauty, such as visual disorder (Kotabe et al., 2017). Furthermore, from an urban planning perspective, this research suggests one clear way to improve city residents' affective well-being is through the incorporation of aesthetically pleasing urban green infrastructure (UGI). Implementing beautiful UGI would be a more feasible way to improve the aesthetic value of currently developed spaces. In addition, nature exposure is associated with improved attentional resources (Berman et al., 2008; Schertz & Berman, 2019), improved mental health (Bratman et al., 2019), reduced crime (Kuo & Sullivan, 2001a, 2001b; Schertz et al., 2019) and greater neighborhood social cohesion (de Vries et al., 2013; Kuo et al., 1998). Therefore, it is likely that the benefits obtained from this UGI would not be limited to residents' emotional functioning.

This study contains a few notable limitations. Though changes in affect have been documented across both real and simulated nature interventions (McMahan & Estes, 2015), these data do not directly speak to whether the results would be different after real life environmental experiences. Given the difficulty in finding urban images that were sufficiently preferred to be able to perform this research, conducting a similar study in preference-equated real environments would likely be very challenging, if not altogether impossible. However, we cannot rule out the possibility that there are qualities of natural environments which contribute to affect change that cannot be captured in images (including additional sensory stimuli), or that the results would not be different with longer term exposures. Another limitation is that we have focused this work primarily on nature's benefits rather than examining the detriments related to urban environments. There is some evidence for a preference effect here as well. In Study 2 the very low aesthetic value urban images were the only category to induce negative affective responses, and in Study 4 very highly preferred urban images elicited a positive affective response. However, Study 2 did not have an equally 'un-preferred' nature condition, so this is still an open question.

In summary, the present research suggests that there is nothing unique about nature beyond preference when it comes to improving affective state, and that viewing anything that a person prefers will have a positive effect. Yet it remains important to emphasize the difficulty in finding stimuli that were as highly preferred as nature scenes. Thus, while there may not be anything unique about nature for affect change above and beyond aesthetics, the observation that natural environments, as well as scenes containing nature-related stimuli, are preferred remains a significant one. Overall, the results of this research contain not only important implications for the research of other environmental psychologists, but also provide insights which may be useful in domains such as architecture, urban design, and nature-based clinical interventions to improve the well-being of residents.

CHAPTER 3: Preference and the Cognitive Benefits of Nature Exposure

Abstract

Research on the psychological benefits of natural environments over urban ones routinely shows that spending time in nature or viewing natural images can improve cognitive functioning, restore attention, and improve mood. While there is evidence that the cognitive and mood benefits are independent effects (i.e., you can see the cognitive benefits without the mood benefits, and one does not always predict the other), there is the additional factor of nature preference that makes dissociating these two effects difficult. Generally, adults prefer natural environments to urban ones, and highly preferred stimuli are also related to mood improvements. The current study, through using preference matched nature and urban stimuli across two levels of aesthetic value, attempts to separate what we propose are the cognitive effects of nature and the mood effects of preference. To this end, 300 participants completed a two-session study online via Amazon's Mechanical Turk, where they performed the dual n-back task and reported their mood before and after an image intervention (nature, urban, or control). Participants always saw images of the same aesthetic value (high, low, or control) but environment type was counterbalanced between sessions within a participant. Neither environment type nor aesthetic value had a significant influence on change in cognitive performance (operationalized by d' on the DNB task). State affect was

influenced by aesthetic preference, but affect was unrelated to cognitive performance.

1. Introduction

The benefits of nature exposure on cognitive processes have been documented extensively in empirical and observational research, across multiple modalities of nature exposure (Bratman et al., 2012; Ohly et al., 2016; Schertz & Berman, 2019; Stevenson et al., 2018). Recent meta-analyses (Ohly et al., 2016; Stevenson et al., 2018) have shown that the effects of nature interventions are strongest on attention tasks that place high demands on working memory, such as the backwards digit span task (BDS; Berman et al., 2008, 2012; Jung et al., 2017; Van Hedger et al., 2018) and tasks requiring cognitive flexibility, such as the trail-making task B (Cimprich & Ronis, 2003; Jung et al., 2017; Shin et al., 2011). Less consistent effects have also been documented in tasks requiring sustained attention and concentration (Berto, 2005; Hartig et al., 2003; Tennessen & Cimprich, 1995).

These results have been situated in a variety of theoretical accounts of nature's cognitive benefits, such as Attention Restoration Theory (ART; Kaplan, 1995; Kaplan & Berman, 2010), Stress Reduction Theory (SRT; Ulrich et al., 1991), and the Perceptual Fluency Account (PFA; Joye et al., 2016; Joye & van den Berg, 2011). Though the ultimate benefits for cognitive function after a nature experience is similarly construed across these theoretical frameworks, the underlying mechanisms differ substantially.

Attention restoration theory suggests that natural environments are one type of environment that can place low demands on directed (voluntary) attention and also gently capture involuntary attention, a concept referred to as 'soft fascination' (Kaplan, 1995). As described by ART, nature is a place where finite directed attention resources can replenish. Three other attributes of an attention-restoring-environment suggested by Kaplan (1995) are 1) providing a sense of 'being away' from a place or mindset where directed attention demands are high, 2) having enough detail or scope to engage the mind, and 3) being compatible with the current goals of the individual. Thus, for the purpose of nature's benefits according to ART, whether an environment is preferred or improves mood is unrelated to the mechanisms underlying nature's cognitive benefits. In contrast, stress reduction theory proposes that the mechanism for nature's salubrious effect on cognition is through nature-elicited increases in positive affect, which subsequently reduces stress and frees up mental resources (Ulrich, 1983; Ulrich et al., 1991). Lastly, the perceptual fluency account suggests that natural environments are more readily processed by our visual systems, due in part to features such as increased self-similarity, or fractalness (Joye & van den Berg, 2011). According to PFA, effortless processing improves affect, which in turn, improves cognition.

As nature preferences are linked to mood improvements (see Chapter 2; van den Berg et al., 2003; White et al., 2017), and decades of research have documented the link between nature exposure and improved affect (see McMahan & Estes, 2015 for a meta-analysis), it's difficult to know what benefits are simply due to nature itself or result from experiencing a preferred environment. Though there is some evidence that the affective and cognitive benefits of nature exposure are independent (Stenfors et al., 2019), examination of the cognitive effects of preference-equated nature and urban environments has thus far been lacking.

This study was designed to test whether cognitive restoration after nature exposure was due to nature itself (in line with Attention Restoration Theory), due to differences in the Aesthetic Value of an environment (which may or may not occur as a result of improved affect, potentially in line with Stress Reduction Theory and Perceptual Fluency), or due to the interaction of each. To this end, participants completed a cognitive task before and after exposure to natural and urban environments equated on aesthetic value across two sessions approximately 1 week apart. If the previously documented cognitive restoration effects post-nature exposure are the result of nature and not preference/aesthetic value, then participants' performance on the cognitive task should improve to a greater degree after viewing nature images compared to preference-equated urban images. However, if the results are not due to nature and instead, result from the typically confounded factor of aesthetic value, then participants' change in performance on the cognitive task should be overall greater in those who viewed higher aesthetic value images than participants who viewed lower aesthetic value images. If both factors are important, then an interaction would be expected, whereby the greatest improvement in cognitive performance will occur for those viewing highly preferred nature images and the worst for those viewing less preferred urban images. All of these possibilities were tested. Lastly, as the three presented theories differ on the importance placed on nature's effects on positive affect, this study will test whether cognitive and affective changes are independent or related.

2. Method

2.1. Participants

308 US-based adults (133 male, 174 female, 1 other) were recruited from Amazon Mechanical Turk for a two-part study. Age of the participants ranged from 19 to 71 years (M = 36.0, SD = 11.0). Six participants in session one encountered technical difficulties, and their data were not analyzed. Of the 302 participants asked to participate in session two, 275 completed the full study procedures for the second session. The full study procedures for session one were expected to take approximately 30 minutes and participants were compensated $3.00 for participating. The full study procedures for session two were also expected to take 30 minutes, and participants were compensated $6.00 for participating. Informed consent was administered by the University of Chicago Institutional Review Board

(IRB). Sample size was decided prior to data collection and specified in the pre-registration (https://osf.io/vn4sh).

2.2. Experimental Design

The main experimental design included both within- and between-subjects factors. The between-subjects independent variables are Aesthetic Value (High vs. Low) and order (Nature or Urban session first), and the within-subjects independent variable is Environment Type (Nature vs. Urban). In addition, a control condition was employed to examine the effect of practice alone. Participants in the control condition also completed two sessions, but were shown half of the neutral images in each session.

2.3. Conditions & Stimuli

Participants were randomly assigned to one of the 4 experimental conditions at session 1: High Aesthetic Value-Nature (HA-Nat; n = 49), High Aesthetic Value-Urban (HA-Urb; n = 55), Low Aesthetic Value-Nature (LA-Nat; n = 51), Low Aesthetic Value-Urban (LA-Urb; n = 46), or were assigned to the control condition (Con; n = 101). For session 2, participants in experimental groups were exposed to the opposite Environment Type and matched Aesthetic Value level from session 1 (i.e., participants who were assigned to HA-Nat in session 1 were assigned to HA-Urb in session 2).

The HA- and LA- Nature and Urban images were the same 45 image sets used in Chapters 1 and 2 (see **Table 3.1** for preference ratings). The control images used were primarily pictures of household items in indoor settings, such as a desk, hand towels, headphones, light fixtures, etc. These 90 images were chosen based on a separate image validation study conducted on Mturk and had an average aesthetic value rating of 4.0.

Table 3.1. Aesthetic Value Ratings of Images in Each Condition
Ratings of all images in each condition from a previous validation study. Aesthetic value ratings are on a 1-7 scale (1 = strongly dislike, 7 = strongly like).

	Nature Conditions		**Urban Conditions**	
Aesthetic Value Condition	*M (SD)*	*Range* [Min, Max]	*M (SD)*	*Range* [Min, Max]
High Aesthetic Value	4.59 (0.17)	[4.28, 4.86]	4.58 (0.29)	[4.15, 5.29]
Low Aesthetic Value	3.78 (0.35)	[2.90, 4.21]	3.78 (0.23)	[3.30, 4.12]

2.4. Image Rating Task

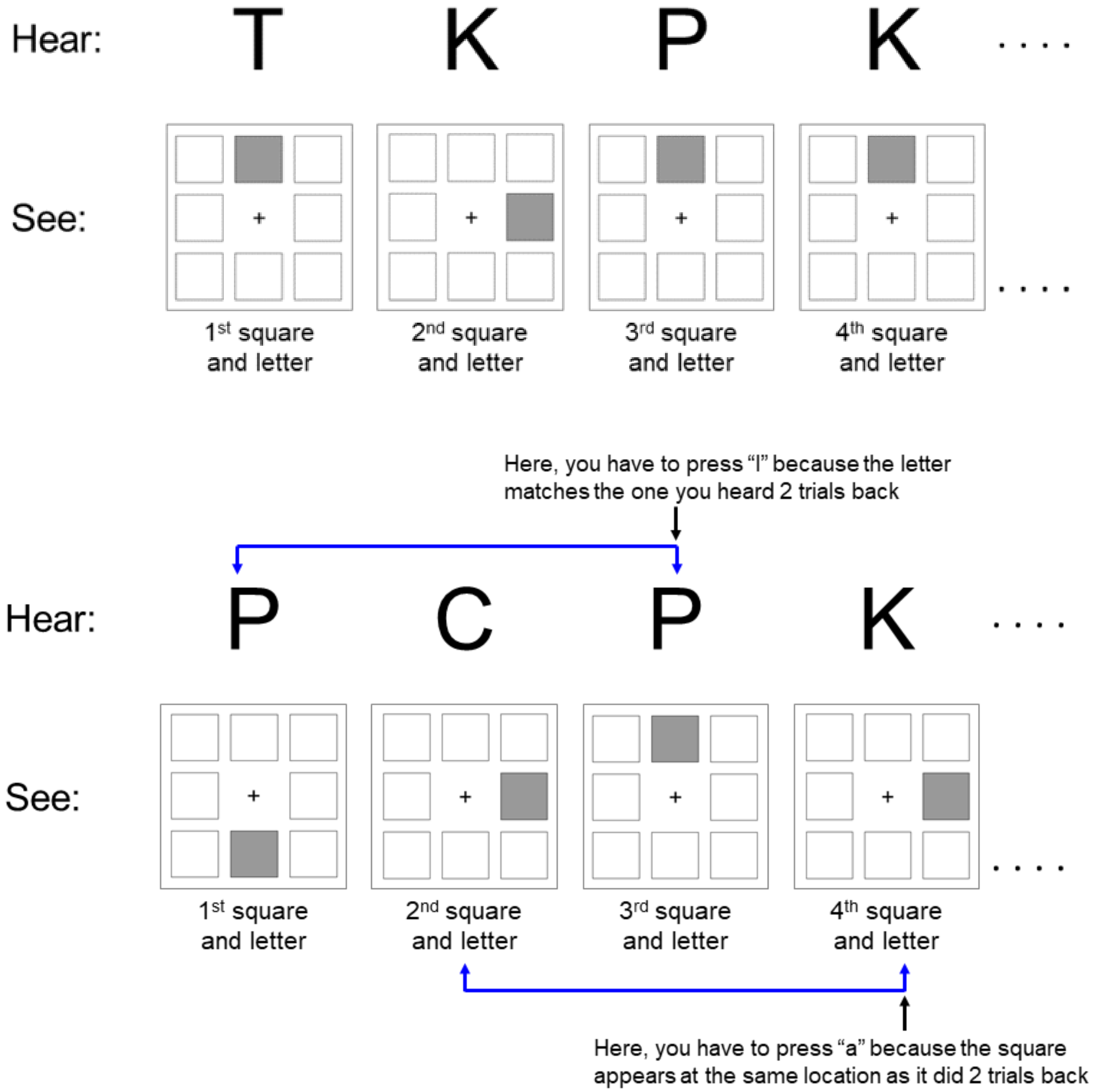
In each session, participants completed an image rating task (IRT) as part of the picture intervention. Each IRT involved 45 images which were presented on the screen for a minimum of 7 seconds. For each image, participants rated their preference for the image on a 7-point Likert scale (anchors: 1 = Strongly Dislike, 7 = Strongly Like). The next image appeared after 7 seconds or after participants made their responses if longer than 7 seconds. Average preference for each of the image sets was calculated by taking the mean of all 45 image ratings.

2.5. Affect Measures

To examine changes in state affect, the short form of the Positive and Negative Affect Schedule (PANAS; (Watson et al., 1988) was employed. The shortened PANAS is comprised of 10 positive and 10 negative adjectives, and participants are asked to rate to what extent they feel this way at the moment on a 5-point Likert scale (anchors: 1 = Very Slightly or Not at All, 5 = Extremely). Composite positive (PA) and negative affect (NA) scores were calculated by adding all of the PA and NA items together.

2.6. Cognitive Measures

To examine changes in directed attention, the dual n-back (DNB) task was used. Though a large number of studies examining Attention Restoration Theory primarily use the backwards digit span (BDS) task, recent research looked at the DNB task and found a larger effect of intervention on DNB performance relative to BDS (Van Hedger et al., 2018). The DNB task places large demands on directed attention and working memory (Lilienthal et al., 2013), making it well suited to examine the changes expected according to ART. In this study, participants performed 2 blocks of dual 2-back and 2 blocks of dual 3-back at each assessment time point. As in Van Hedger et al. (2018), the dual n-back consisted of an auditory (spoken letter) and a visuospatial (square moving around a 3x3 grid) component. Participants were told to press “A” if the spoken letter matched the letter located *n* trials back (i.e. 2 or 3 trials back, depending on n-back level). At the same time, participants were asked to press the "L" key if the square was in the same location presented *n* trials back. If both square and letter matched, participants were asked to press both the "A" and "L" keys, and if neither matched, participants did not need to press any key (see **Figure 3.1**).



Here, you have to press "l" because the letter matches the one you heard 2 trials back

Here, you have to press "a" because the square appears at the same location as it did 2 trials back

Figure 3.1. Dual N-back Task.
Dual N-back overview (Top) and Dual 2-back example (Bottom)

At the first time point in each session, participants were taken through the full instructions for completing the DNB task, followed by a practice round of dual 2-back with trial-by-trial accuracy feedback, then two blocks of scored dual 2-back without trial-level feedback. They were then taken through a practice round of dual 3-back and two blocks of scored dual 3-back. In each case, practice rounds consisted of 10+n trials and scored rounds consisted of 20+n trials. For each round, there was a fixed ratio of trial types: 50% of trials were no match, 20% were auditory match only, 20% were visuospatial match only, and 10% were both auditory and visuospatial matches. For each participant, a d' score was calculated for each n-back level (Macmillan & Creelman, 2004) at each assessment time. d' for each DNB level was defined as the z-scored Hit rate - the z-scored False Alarm rate across both main blocks of task.

At the second assessment within a session (post-image viewing), an abridged version of the instructions was used, but the same sequence of tasks (practice dual 2-back, main dual 2-back, practice dual 3-back, main dual 3-back) was used.

2.7. Statistical Analysis

As outlined in the pre-registration, our confirmatory analyses were to conduct 2 (within-subjects factor: Environment Type) x 2 (between-subjects factor: Aesthetic Value Level) x 2 (between-subjects factor: Order) mixed model ANOVAS on each of the outcome measures: change in dual n-back performance, change in positive affect (PA), and change in negative affect (NA). All analyses were conducted in R v. 3.5.1 (R Core Team, 2019). ANOVAs were conducted using package ‘ez’ (v4.4-0, Lawrence, 2016), and post hoc comparisons were conducted using Tukey HSD multiple comparisons correction with a 95% family-wise confidence level.

3. Results

3.1. Quality Check/Participant Attrition Information

Based on Mturk worker feedback (email or comment in survey) regarding technical issues, 6 participants were excluded from session 1 and 4 were excluded from session 2, yielding an initial sample of 302 participants for session 1 and 271 participants with both session 1 and 2 data. Of the 271 subjects with data for both sessions, 54 participants were excluded due to bad DNB data. Bad DNB data was defined as meeting either of the following conditions: 1) Worse than chance performance (d’ < 0) on more than half of the dual 2-back and 3-back blocks, 2) Too many key presses (more responses than there were trials), or 3) Too few key presses (responded only a couple times per block). Additionally, 15 participants were excluded due to Mturk worker feedback indicating they were distracted by something during the study that prevented them from focusing on the task. Another 56 participants were excluded due to questionable timing in the study, defined as either 1) long RTs during the image rating task or a long duration of the IRT, indicating lack of attention or interruption during the IRT or 2) long delays between important parts of the study (between image intervention and the second round of

DNB, in particular). After removal of these participants, 158 participants remained. Attrition by condition is presented in **Table 3.2.** (Note: a small number of participants fit into one or more of these categories, and therefore the total number of removed participants is not equal to the sum of the categories).

Table 3.2. Usable Participants

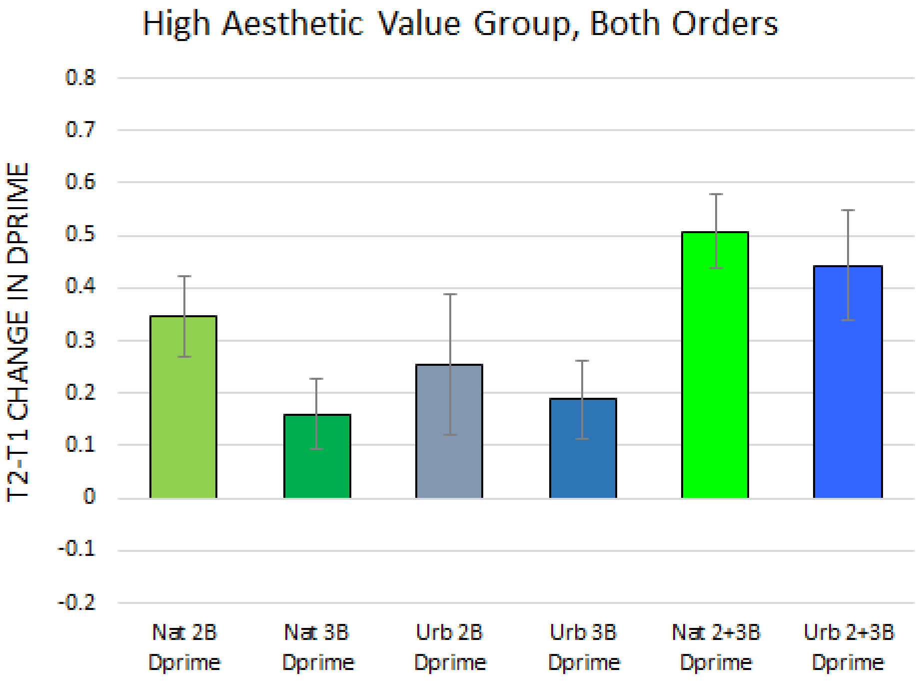
Participants with usable DNB data in bold out of total participants (in parentheses)

Condition	N for both orders	N for Nature 1st, Urban 2nd order	N for Urban 1st, Nature 2nd order
High Aesthetic Value	**57** (94)	**25** (44)	**32** (50)
Low Aesthetic Value	**50** (89)	**26** (48)	**24** (41)
Control	**51** (88)	N/A	N/A
TOTAL	**158** (271)		

3.2. Confirmatory Analyses on Cognitive Restoration Change

To analyze changes in Dual n-back performance as a function of Environment type, Aesthetic Value Level, and Environment order, 2x2x2 mixed-effects ANOVAs were performed separately on d’ change in dual 2-back, in dual 3-back, and in a composite of both dual 2- and 3-back. None of the main effects or interactions in these omnibus ANOVAs reached statistical significance (all $p > 0.11$) [**Figure 3.2**].

Figure 3.2. Change in d' by Environment Type
Shown separately by Aesthetic Value Level Group and Control Group, for dual 2-back (2B), dual 3-back (3B), or a composite of dual 2 and 3 back (2+3B)



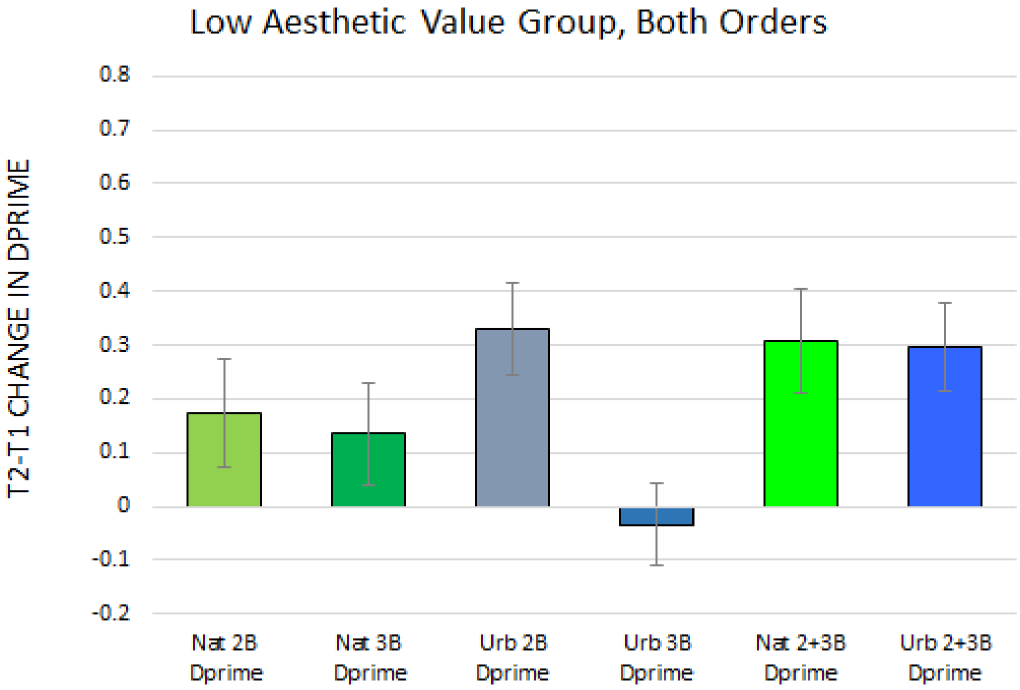
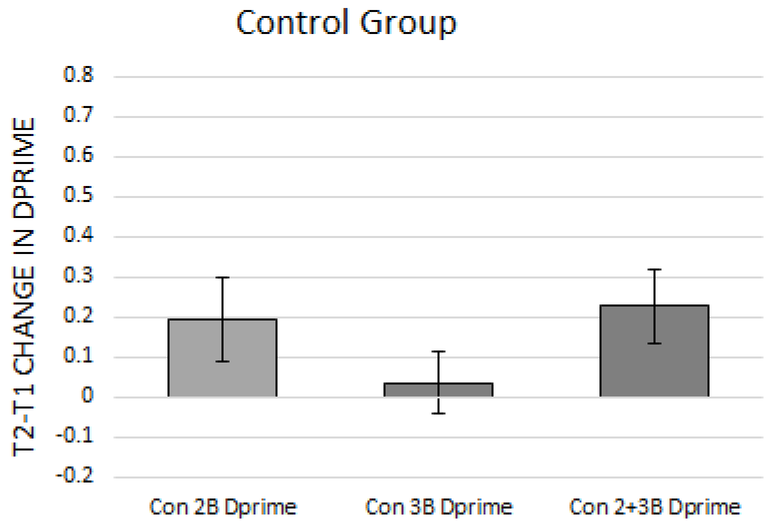


Figure 3.2, continued.



3.3. Exploratory Analyses on Cognitive Restoration Change

3.3.1. Analyses within Aesthetic Value Level

As previous research has demonstrated that cognitive restoration after nature is strongest when a within-subjects design is employed, exploratory analyses looking only at the within-subjects effects were conducted. As Environment type was a within-subject manipulation and Aesthetic Value Level was a between-subjects manipulation, analyses of the effect of Environment type within an Aesthetic Value Level were conducted. Results of these 2 (Environment type) x 2 (Order) ANOVAs for each dual n-back level in each Aesthetic Value Level yielded no significant main effects or interactions (all uncorrected *p* > 0.08).

3.3.2. Analyses Incorporating Individual Differences in Image Preference Ratings

Though there were no effects due to Aesthetic Value Level in the omnibus ANOVA, it remained possible that aesthetic value at the level of individual differences may have a greater impact than at the level of pre-specified Aesthetic Value condition. As this was the case in Chapter 2, ANCOVAs and regressions were conducted on each level of dual n-back.

In this analysis, 2 (Environment Type) x 2 (Order) mixed-model ANCOVAs were used with participants' average preference ratings for the images in each environment as covariates. The results of these ANCOVAs on change in dual n-back were not significant (all uncorrected $p > 0.14$).

In the regressions, average image preference rating was used to predict change in dual n-back performance, irrespective of condition. The results of this regression did not yield any significant relationships between d' change and average preference rating (all $p > 0.25$). Together these results suggest that image preference did not relate to change in cognitive performance.

3.4. Confirmatory Analyses on Affect Change

To analyze changes in affect as a function of Environment type, Aesthetic Value Level, and Environment Order, 2x2x2 mixed-effects ANOVAs were performed separately on change in positive affect (PA) and negative affect (NA).

The results of this analysis on PA elicited a main effect of Environment Type, $F(1,103) = 4.0$, $p = 0.005$, $\eta_p^2 = 0.032$, 95% CI [0.0, 0.12], whereby participants had a larger decrease in PA after viewing the natural environments ($M = -0.18$, $SD = 0.38$) compared to the urban ones ($M = -0.08$, $SD = 0.47$). No other main effects or interactions were found for positive affect (all $p > 0.09$). For NA, the omnibus ANOVA did not yield any significant effects or interactions (all $p > 0.05$). [**Figure 3.3**]

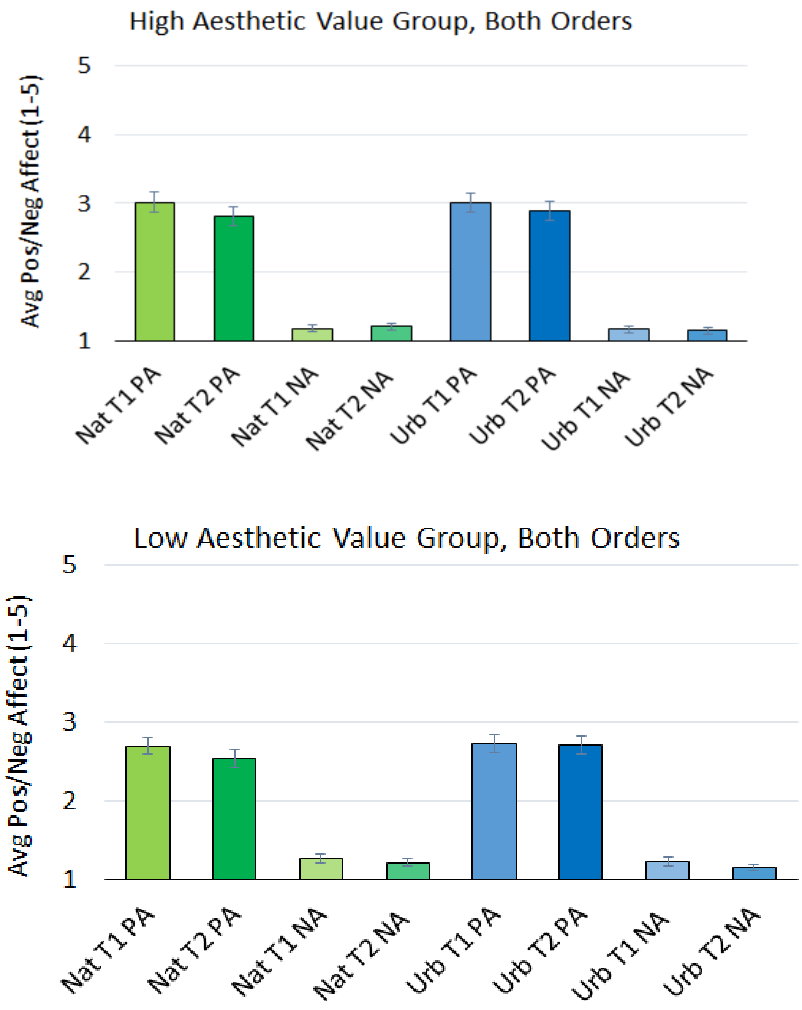


Figure 3.3. Positive (PA) and Negative Affect (NA) at T1 (pre) and T2 (post)
Shown separately by Aesthetic Value Level Group

3.5. Exploratory Analyses on Affect Change

3.5.1. Analyses Incorporating Individual Differences in Image Preference Ratings

To examine whether individual differences in preference ratings had an effect on affect change (as in Chapter 2), ANCOVAs using individual differences in preference ratings were analyzed. In these analyses, 2 (Environment Type) x 2 (Order) mixed-model ANCOVAs were used with participants' average preference ratings for the images in each environment as covariates. Indeed, upon examination of the average preference ratings for each of the image types, it became apparent that while the nature and urban images in the High Aesthetic Value category were equated ($ps = 0.5$), the images in the Low Aesthetic Value category were not ($p = 0.02$) [See **Figure 2.4**].

Though there was a significant effect of Environment type on positive affect in the confirmatory analyses, once individual differences in preference were included as covariates, this effect of environment was no longer significant ($ps = 0.17$). No significant effects were found in this ANCOVA for PA. When conducted on change in negative affect, the ANCOVA yielded no significant effects (all $p > 0.09$).

Lastly, as with change in cognitive performance, regressions were conducted to examine whether average image preference was significantly related to change in affect, irrespective of condition. The results of these regressions showed a marginal effect of image preference on positive affect ($R^2 = 0.01$, $F(1,212) = 3.4$, $p = 0.067$), where higher preference was associated with more positive affect change. This analysis was not significant for negative affect ($ps = 0.6$).

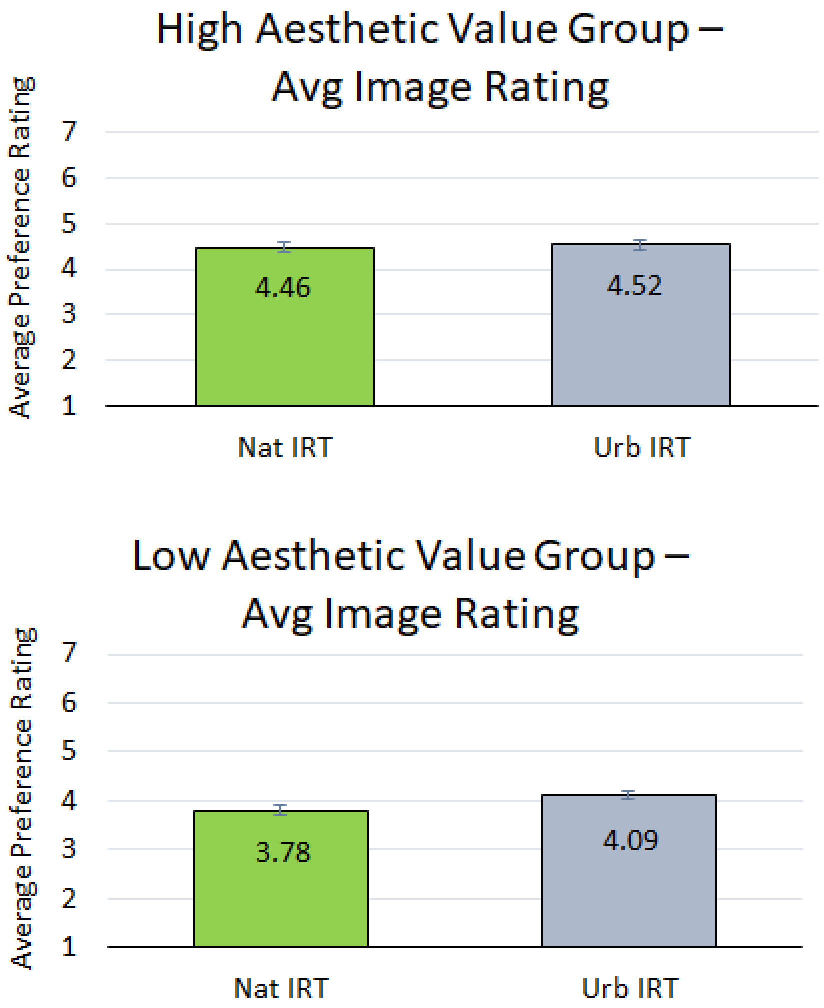


Figure 3.4. Average Image Preference Ratings by Condition

4. Discussion

The primarily null results of this study did not provide much evidence for a nature-based or preference-based account of nature's effects on cognitive restoration. Change in dual n-back performance was not affected by aesthetic value level, environment type, order, nor any interaction of these three factors. They were also not predicted by individual differences in how much participants preferred the images they saw. Consistent with the results of Chapter 2, state affect was impacted by individuals' own preference ratings and not affected by environment type once individual differences in aesthetic value were accounted for.

Unfortunately, with these null results gathered from an online platform where participant compliance and distractions cannot be fully accounted for, it is difficult to say whether these results reflect a 'true' null effect or are impacted by extraneous factors. One possibility regarding these results is that the image interventions were not potent enough to elicit meaningful changes in cognitive restoration. As the effect sizes for virtual nature exposure are generally smaller than those obtained with real nature exposure (i.e. nature walks), it is entirely possible that the intervention used here was not immersive enough or long enough to have an actual effect. However, given the questionable quality of the collected data, it is also very possible that in a more controlled setting where noise/distractions can be minimized, the interventions may have an effect. As a consequence, in chapter 4, an in-lab experiment using a 10-minute video intervention was used.

CHAPTER 4: Neural Correlates of Cognitive Effort and Restoration from Equally Preferred Nature and Urban Environments: an fNIRS Investigation

Abstract

Though behavioral research on the cognitive benefits of nature exposure is abundant, work examining the neural underpinnings of nature's salubrious effects is relatively scarce. Technical limitations of functional MRI prevent the investigation of how the brain may differentially process real natural and urban environments, but advances in functional near infrared spectroscopy (fNIRS) have demonstrated its potential for the monitoring of cortical hemodynamics outside of the laboratory. The current study was designed with two aims: 1) to validate the efficacy of fNIRS as a measure of cognitive effort using a standard n-back task, large sample size, and more robust statistical analysis, and 2) to measure whether behavioral or fNIRS measures of cognitive restoration can be elicited from preference-equated videos of natural vs. urban environments. fNIRS and behavioral data from 70 participants were collected and analyzed. No differences in cognitive restoration (behaviorally or via fNIRS activation) were found between the two video types. However, results from the fNIRS validation showed reliably greater neural activity for the 2-back task relative to the 1-back task, suggesting frontal and parietal fNIRS measurements are sensitive to differences in cognitive load. Multivariate analyses (PLS) demonstrated differences in the relation between performance (accuracy) and change in the deoxyhemoglobin fNIRS signal as a function of n-back level. Future studies to

examine the outstanding questions of the role of virtual nature exposure vs. preference and fNIRS measurements during virtual and real environmental exposures are discussed.

1. Introduction

Though behavioral research on the cognitive benefits of natural environments is abundant (Ohly et al., 2016; Schertz & Berman, 2019; Stenfors et al., 2019; Stevenson et al., 2018), very few studies have examined the direct effect of natural environments on the neurobiological processes which underlie behavioral changes. Several studies using psychophysiological metrics (e.g., heart-rate, blood pressure, cortisol, skin conductance) have tested responses before, during, and after exposure to different environments and typically, related them to a behavior of interest (Beute & de Kort, 2014; Brown et al., 2013; Kondo et al., 2018; Laumann et al., 2003; Ulrich et al., 1991). These peripheral nervous system measures provide useful information about physiological states, particularly as they relate to stress and arousal, but provide a limited proxy for the neural activation elicited by natural or urban environments. To gain direct insight into the neural mechanisms which allow for cognitive restoration after nature exposure, neuroimaging techniques must be used.

A handful of research studies using functional magnetic resonance imaging (fMRI) and electroencephalography (EEG) have been conducted in recent years to address this question. Due to the physical constraints of the method, most fMRI studies have related nature exposure outside the scanner to functional brain activity collected in a separate fMRI session. For example, Bratman and colleagues collected resting state fMRI from participants before and after a 90-minute nature and urban walk, and found decreases in self-reported rumination and subgenual prefrontal cortex activation for the nature walk but not the urban one (Bratman, Hamilton, et al., 2015). Similarly, a recent study (Tost et al., 2019) gathered ecological momentary assessments and geolocation data to measure nature exposure and emotional well-being over one week, then related this to neural activity during an emotion regulation task. This study found that increased exposure to greenspace as assessed by GPS data was related to reduced activity in the dorsolateral prefrontal cortex (DLPFC) when participants were asked to down-regulate negative emotions.

Additionally, a few recent neuroimaging studies have looked at cognitive changes and neural responses after viewing virtual natural environments, and found some evidence for reduced attention demands in processing certain types of natural environments over urban ones. Both electrophysiological measures (i.e. alpha power from EEG; Grassini et al., 2019) and metabolic measures (fMRI activation in posterior cingulate cortex; Tang et al., 2017) were found to differ between nature and urban environments. However, none of these studies established a relationship between measures of behavior and brain activity, and neural activity was not measured during exposure to real natural environments.

Advancements in mobile EEG technology are creating opportunities to study neural activity in realistic, outdoor environments (Debener et al., 2012; Piñeyro Salvidegoitia et al., 2019). While this a promising avenue of research, there are still some shortcomings of this approach with regards to data quality, ability to access raw data, and difficulty disentangling the effects of physical movement from the EEG components of interest (Ries et al., 2014; Zink et al., 2016). Importantly, another neuroimaging modality that has gained traction in recent years, functional near-infrared spectroscopy (fNIRS), is very well-suited to studying brain activity in realistic natural environments. Compared to EEG and fMRI, fNIRS is robust to motion artifacts and environmental noise, making it an increasingly popular method for studying neural activity outside of standard laboratory experimentation (Pinti, Tachtsidis, et al., 2018; Yücel et al., 2017).

fNIRS uses light spectroscopy at near-infrared wavelengths to measure the same cerebral metabolic changes that are measured using functional MRI (Buxton, 2010; Huppert et al., 2006). In both methods, the measurements taken are metabolic proxies for neuronal activity. When neural activity increases, so does the metabolic demand, leading to increased blood flow in the surrounding vasculature. This blood flow causes an increase in concentrations of oxygenated hemoglobin and a decrease in concentrations of deoxygenated hemoglobin (Buxton, 2013; Huppert et al., 2006). A key difference between the two methods is that the biological signal of interest in fMRI (the BOLD, or blood oxygenation level dependent response) specifically measures change in deoxyhemoglobin, which has paramagnetic properties that distort the magnetic field (Pauling & Coryell, 1936). In contrast, fNIRS measures concentration changes in both oxyhemoglobin (HbO) and deoxyhemoglobin (HbR) by shining light into superficial cerebral cortex, and taking advantage of the difference in light absorbing properties between oxy- and deoxy-hemoglobin at different optical wavelengths. Specifically, HbO absorbs NIR light better at wavelengths < 800 nm, and HbR absorbs light better for wavelengths > 800 nm. fNIRS sources emit NIR light at two wavelengths and nearby detectors measure the amount of light that remains after it has traveled through cortical tissue. This change in light intensity from source to detector at each wavelength is eventually converted into HbO and HbR concentration changes using the modified Beer-Lambert law (Jacques, 2013).

Though the same basic biological signal is measured in fMRI and fNIRS, there are a few key differences between the two. While the temporal resolution of BOLD response and changes in hemoglobin changes are still relatively slow, peaking around 5-6 seconds post-stimulus onset, the acquisition rate for fNIRS (anywhere from 4-500 Hz) is much higher than that for fMRI (usually 0.3-0.5 Hz), allowing for a more complete temporal modeling of the hemodynamic response (Huppert et al., 2006). However, while fMRI benefits from precise spatial resolution (1-3 mm) and whole brain coverage, fNIRS resolution is typically on the scale of 1.5 to 3 cm and can only measure activity in superficial neural cortex, at a depth of approximately 5-8 mm (Huppert, 2016).

Empirical work showing that BOLD signal change is driven by changes in deoxyhemoglobin (HbR)(Buxton et al., 1998) lead to the theoretical assertion that HbR concentration changes in fNIRS are more tightly coupled with the BOLD response than are HbO concentrations. Studies involving simultaneous fNIRS-fMRI recordings have provided empirical evidence for this theoretical link, demonstrating higher temporal correlation between HbR and BOLD relative to HbO and BOLD (Huppert et al., 2006; Kleinschmidt et al., 1996; Schroeter et al., 2006). However, the signal-to-noise ratio (SNR) of HbO is generally higher than that of HbR in fNIRS, and HbO still correlates highly with BOLD signal despite being a faster response (i.e., peaking at ~4 seconds) (Huppert et al., 2006; Strangman et al., 2002). Due to the improved SNR, many studies focus analyses of interest on changes in HbO, or total Hb (HbT = HbO - HbR). Ultimately, while the underlying neurobiological changes measured in fNIRS is the same as fMRI, there are some discrepancies between the signals of BOLD, HbO, and HbR that result primarily from differences in acquisition and analysis.

Historically, many of the typical fNIRS analysis techniques mirrored those of task-based fMRI. The vast majority of existing fNIRS studies involved initial data preprocessing (i.e., downsampling, bandpass or wavelet filtering, motion correction), conversion into HbO and HbR concentrations, and a general linear model to compare Hb changes between task conditions or between task and rest (Cooper et al., 2012; Pinti, Scholkmann, et al., 2018; Scholkmann et al., 2014). However, it was recently demonstrated that these typical approaches overall fail to account for specific statistical properties of the fNIRS signal, and in doing so, inflate the false positive rate of reported results (Barker et al., 2013; Huppert, 2016).

Specifically, due to the sampling rate of fNIRS being much faster than the physiological signals measured (0.01 to 0.1 Hz), the physiological noise is colored (i.e., has specific temporal or spatial frequencies in the noise spectrum) and contains high serial correlations.[1] This colored noise leads to correlations in the error terms of the linear model and violates the assumption of independent sample points. This lack of independence of measurement is further affected by high correlation between nearby source-detector pairs and between the HbO and HbR signals.

Additionally, the noise in fNIRS exhibits temporal heteroscedasticity when motion artifacts occur, and spatial heteroscedasticity when there are channels (source-detector pairs) with highly varied SNR. This latter point is particularly applicable when comparing SNR across channels that are partially obstructed by hair (with poor data quality) versus those which make direct contact with the scalp

---

[1] Though the noise in fMRI is also colored (due to a sampling rate of 0.5 Hz vs signal of 0.1 Hz), the effect is much smaller than fNIRS and the overall SNR in fMRI is better than fNIRS so the noise has much less of an effect on the fMRI signal.

such as the forehead (with much higher data quality). This can also occur when conducting group-level analyses where there may be substantial SNR differences between participants (Huppert, 2016).

Due to these issues, new analysis methods have been developed to account for these specific statistical properties of fNIRS. By applying pre-whitening to the linear model to reduce noise correlations and using robust regression to down-weight statistical outliers, these methods perform better on sensitivity-specificity analyses and show better control of type-I errors (Barker et al., 2013; Huppert, 2016; Santosa et al., 2018). With proper statistical analysis to account for these unique noise properties, fNIRS provides an increasingly rigorous method of portable and versatile neuroimaging.

An ultimate goal of the portable fNIRS system to measure the neural processing of real, outdoor natural and urban environments, and relate this to performance on cognitive tasks. Before applying fNIRS to this relatively complex situation, a necessary first step is to validate its efficacy in measuring cognitive effort and restoration in a controlled laboratory setting. To ensure that the results are not affected by the inflated false positive rate which can occur from standard GLM analysis, this study used the Brain AnalyzIR Toolbox (Santosa et al., 2018), a toolbox designed to provide solutions to the statistical issues mentioned above. This study also employed a well-validated working memory paradigm, the N-back task, to test for the expected increase in frontal and parietal neural activity with increasing cognitive load (Owen et al., 2005). Importantly, this study also examines exposure to videos of preference-equated natural and urban environments, and tests for behavioral changes in cognitive restoration between the two videos.

There were three primary aims of the current study. The first aim was to examine whether the nature video (a longer and more immersive intervention than the pictures used in Chapter 3) elicited greater cognitive restoration compared to a preference-equated urban video, as measured by behavioral performance (accuracy on the 3-back task). It was hypothesized that delta performance on the 3-back task would be greater for those who watched the nature video than for those who watched the urban video. (This aim was a pre-registered confirmatory analysis: https://osf.io/73evf)

A second, more fundamental aim was to validate the efficacy of fNIRS as a measure of cognitive effort to allow future investigation in both laboratory and more realistic environments (i.e., during nature walks). Cognitive effort here is quantified by load-dependent increases in oxygenated hemoglobin (HbO) and decreases in deoxygenated hemoglobin (HbR). It was hypothesized that during the initial n-back task, prefrontal and parietal cortical activity would be largest for the 3-back task (highest cognitive load), lessened for the 2-back task, and smallest for the 1-back task (lowest cognitive load). Additionally, to test whether fNIRS activity is dependent upon both task demands and performance, an exploratory behavioral partial least squares (PLS) analysis was conducted to examine the relationship between accuracy and neural activity as a function of n-back level.

Lastly, it is possible that there are indicators of cognitive effort in neural activity that are not visible in performance on the cognitive task. For example, with adequate practice, an individual might be able to achieve a consistent level of accuracy on a task, but the amount of mental effort required to achieve the same performance could vary depending on the conditions (Murata et al., 2015). As such, a final goal was to examine whether neural activation, quantified by changes in oxy- or deoxy-hemoglobin, would show condition-level (environment intervention type) differences in cognitive restoration that are not visible in performance alone. To test this, activation in the post-video 3-back task will be compared across video conditions. It was hypothesized that if the nature video had a restorative effect but change in performance was similar across groups, then participants who viewed the nature video would have lessened prefrontal and parietal activity (lower HbO and higher HbR) during the post-video 3-back block relative to those who viewed the urban video. This result would suggest that, though performance is similar across groups, those who had greater cognitive restoration after the nature video needed to exert less effort on the post-video 3-back task to achieve equivalent performance to those who viewed the urban video (Neubauer & Fink, 2009). (This third aim was a pre-registered exploratory analysis: https://osf.io/73evf).

2. Method

This study was pre-registered on OSF prior to data collection: https://osf.io/73evf but not all planned analyses are included in the chapter.

2.1. Participants

Seventy adults participated in this study. All participants had normal or corrected-to-normal visual acuity. Participants gave written informed consent before participation and experimental procedures were approved by the University of Chicago's Committee for Institutional Review Board (IRB). Participants were compensated at a rate of $26 or 2 units of course credit, plus a performance-based bonus of up to $10. The full study procedures typically took 75-90 minutes.

Two participants were excluded from all data analysis due to participant non-compliance with the study procedures. Six additional participants were excluded from fNIRS analysis due to technical issues (2 participants) or low quality fNIRS data (4 participants), leaving a final sample of 62 participants. Of the 62 participants with usable fNIRS data, 28 were male and 34 were female, and the mean age was 23.6 years (*SD* = 6.3 years).

Behavioral analysis of the n-back task was conducted on the total dataset (N = 68) and in a dataset excluding a subset of participants with overall poor performance (n = 47), defined by less than an average accuracy of 60% on any given n-back level.

2.2. Procedure

After providing informed consent, experimenters measured the participants' head to determine cap size and placement, then began to set up the cap while participants were taken through task instructions and given an opportunity to practice the n-back task. After the first round of practice, the cap was placed on the participants' head, alignment was checked based on the location of Cz relative to nasion and inion, then hair was moved as needed to provide clear access to the scalp for the sources and detectors. fNIRS data were then calibrated and checked for quality before proceeding. If any channels were not displaying sufficiently high quality data, placement and hair-clearing were performed again before continuing. Next, participants completed a short round of additional practice, and then filled out the baseline affect measures before continuing on to the main round of n-back task. Subsequently, participants viewed a 10-minute video of either a natural or urban environment. The videos were preference-matched based on a separate validation study. After the video ended, participants filled out the same affect questions as before, then completed an additional 6 blocks of the 3-back task. After this, the cap was removed and participants completed a demographics questionnaire and a post-study survey. All experimental procedures were coded and presented using PsychoPy.

2.3. N-back Task

The experimenter took participants through step-by-step instructions of the n-back task before participants began practice. Participants were told “In this task, you will see a sequence of short words, such as AXE and BOX, separated by brief fixations. Every 2 seconds, a word will be presented. Every time a word appears, you will need to indicate whether it matches the same word that was presented N trials ago. In this study, N will be 1, 2, or 3. If you are asked to do a 1-back task, you will have to press the “m” key every time the current word is the same as the one presented previously (1 trial back) in the sequence. Selecting “m” means that the current word matches the one presented 1 trial back. If the current word does not match the one presented 1 trial back, you will press the “n” key for non-match. In a 2-back task, you will compare the current word to the word 2 trials back and in a 3-back task you will compare the current word to the word 3 trials back.” [**Figure 4.1**]

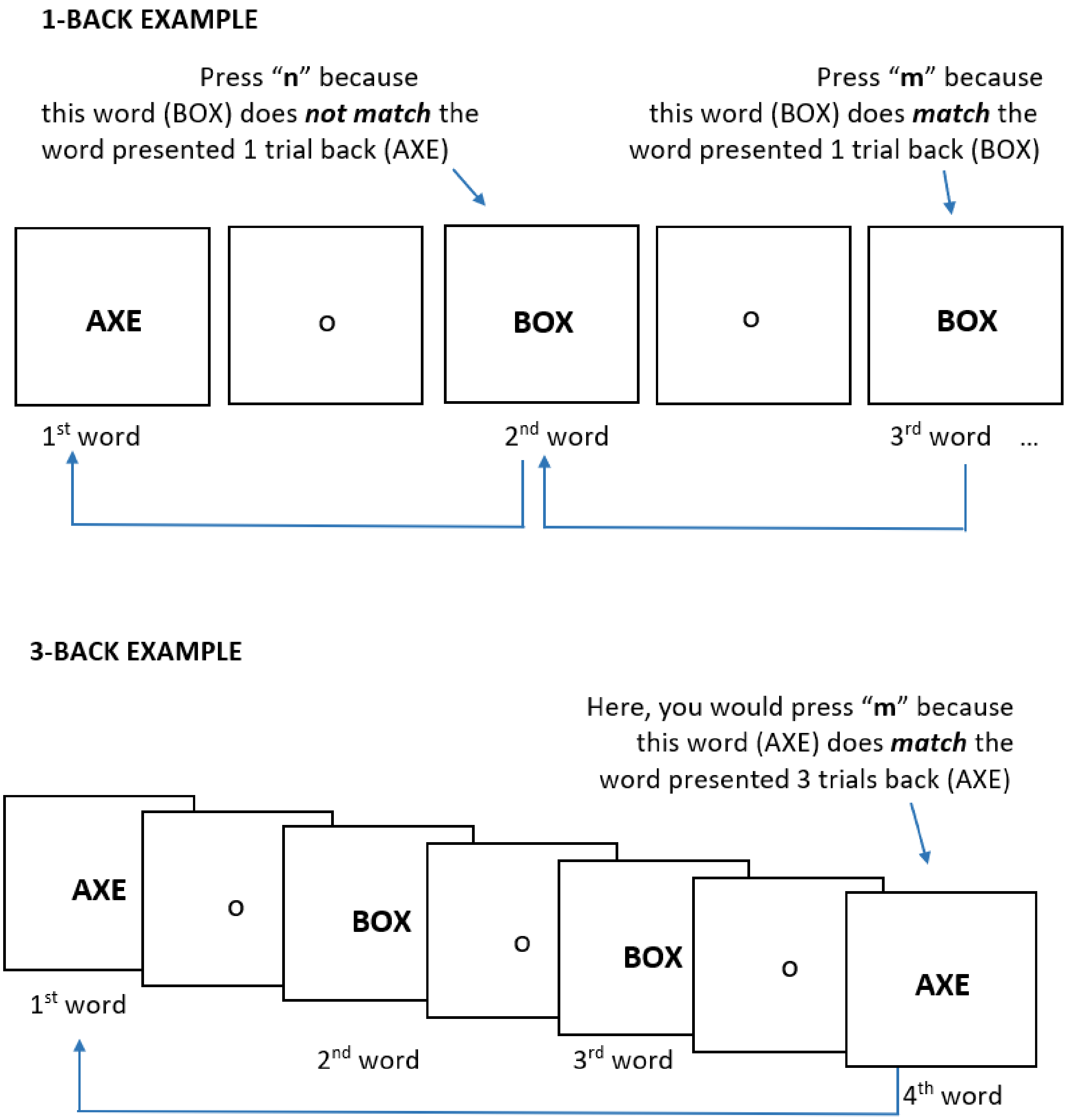


Figure 4.1. N-back Task
Example of 1-back task (Top) and 3-back (Bottom). 2-back task not shown.

After the experimenter took participants through the n-back instructions participants performed the first round of n-back practice, consisting of 9 blocks. In this first round, accuracy feedback was provided on a trial-by-trial level as well as at the end of each block. Participants completed 3 blocks of 1-back, 3 blocks of 2-back, then 3 blocks of 3-back. After the fNIRS cap was set up, participants began the second round of practice: a single block of 1-back, then 2-back, then 3-back, without trial-by-trial feedback.

The main N-back task involved 18 blocks, with 6 blocks of each n-back level, pseudorandomly presented. Each block began by displaying the n-back level and a fixation cross (5 seconds). Each task block contained 15 words, presented for 2 seconds each for a total of 30 seconds, followed by 20 seconds of rest. Therefore, the length of each block was 55 seconds, and a total of 16.5 minutes to complete the full main n-back round. Participants received a performance-based bonus during this round of n-back task, wherein performance > 90% on a block earned an additional 40 cents per block, > 80% earned an additional 30 cents per block, and > 60% earned an additional 20 cents per block. Performance under 60% did not yield a cash bonus in this study. The post-video 3-back task involved 6 sequential blocks of 3-back, with the same timing as the main n-back round and the same performance-based bonus.

2.4. Affect Measures

State affect was measured using a visual analog scale (1-7) for 5 items: happy, sad, angry, calm, and inspired. Items were presented randomly.

2.5. fNIRS Data Acquisition

fNIRS data were collected from a continuous-wave NIRSport2 device (NIRx Medical Technologies, LLC). The wavelengths of emitted light (LED sources) in this system were 760 nm and 850 nm, corresponding to oxygenated hemoglobin and deoxygenated hemoglobin concentrations, respectively. The data were collected at a sampling rate of 4.5 Hz using the NIRx acquisition software, Aurora fNIRS. The fNIRS cap contained a total of 16 sources and 16 detectors creating 43 total channels covering bilateral frontal cortex (33 channels) and right parietal cortex (10 channels).

2.6. fNIRS Optode Locations (Montage)

The montage was created using fOLD (fNIRS Optodes' Location Decider; Morais et al., 2018), which allows placement of optodes in the international 10-10 system to maximally cover anatomical regions of interest, as specified by one of 5 parcellation atlases. The AAL2 (Automated Anatomical Labeling; Rolls et al., 2015) parcellation was used to generate the montage, which was designed to gain as much coverage of the prefrontal cortex (PFC) as possible, covering bilateral superior and inferior frontal gyri. This emphasis on frontal cortical areas was decided based on evidence from other n-back studies using fMRI (see Owen et al., 2005 for a meta-analysis) and fNIRS, which have demonstrated that load-dependent changes in HbO and HbR are found across areas of the PFC (Ayaz et al., 2012; Fishburn et al., 2014; Herff et al., 2014; Sato et al., 2013). Additionally, PFC activity is associated with tasks requiring sustained attention (Bunce et al., 2011; Shimizu et al., 2009), which is necessary in successfully performing the n-back task and is relevant for hypotheses related to attention restoration after nature exposure (Kaplan & Berman, 2010; Schertz & Berman, 2019).

The right parietal region was selected as an additional ROI for this task due to evidence that parietal cortical regions are engaged during attention-demanding tasks in fNIRS (Hosseini et al., 2017; Murata et al., 2015) and for the purpose of measuring functional connectivity across the frontoparietal attention network, which we hypothesized to increase with n-back load and with practice (Fishburn et al., 2014; Thompson et al., 2016)[2]. As parietal data quality is usually less consistent than channels unobstructed by hair (such as the forehead), the majority of optodes (12 sources and 12 detectors) were focused on prefrontal regions, leaving only 4 sources and 4 detectors to cover parietal areas. Rather than sparsely covering bilateral parietal cortex, better coverage of right parietal cortex was examined in the current study. Though verbal working memory storage and rehearsal are more associated with left-lateralized regions of parietal cortex (Awh et al., 1996; Ravizza et al., 2004), meta-analyses demonstrate bilateral parietal activation across verbal and non-verbal n-back tasks (Mencarelli et al., 2019; Owen et al., 2005). Right parietal was chosen as participants would be required to use their right hand to respond during the task and our parietal montage overlapped with the standard

---

[2] Functional connectivity analyses are planned but have not yet been conducted

sensorimotor fNIRS montage. As we did not want to have the more anterior channels in our parietal montage to be affected by differences in contralateral sensory or motor-evoked activity (i.e. due to less or more responding based on task difficulty), we opted to focus on right parietal coverage. [**Figure 4.2**]

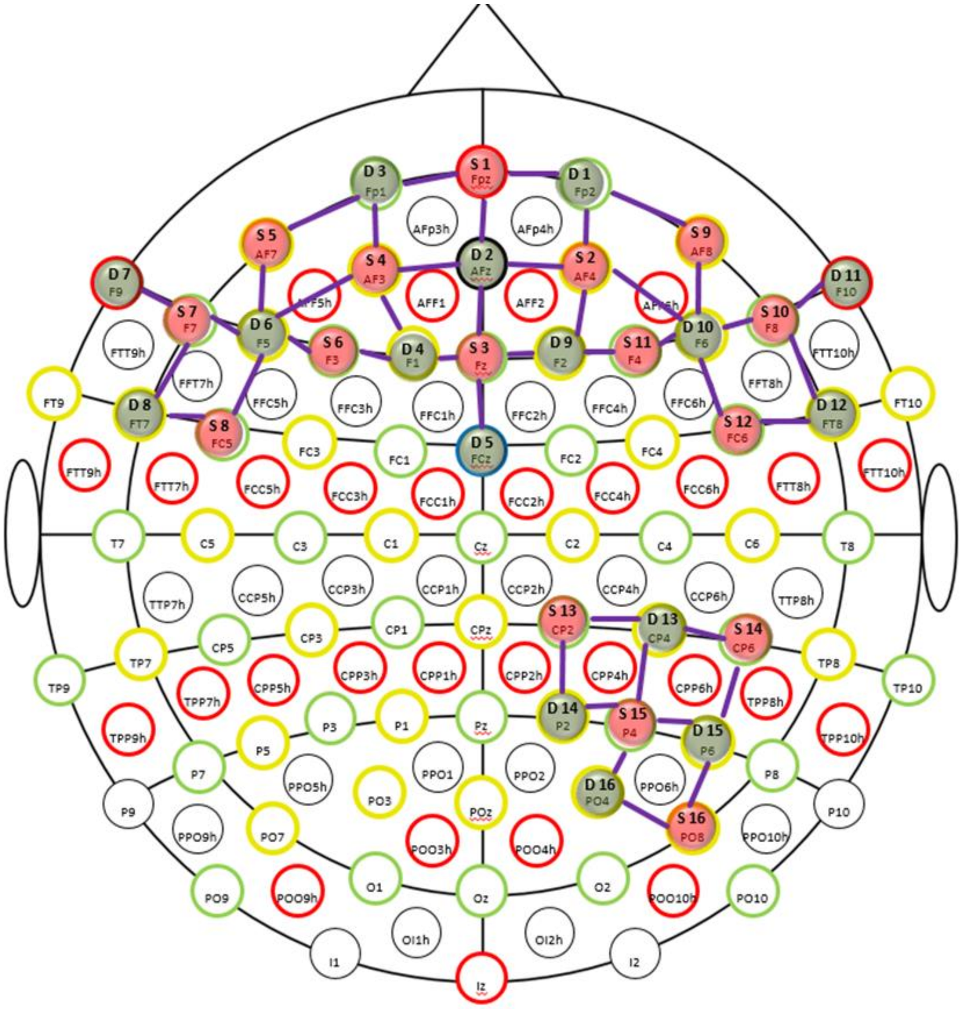


Figure 4.2. fNIRS Montage in international 10-10 coordinate space

Gross ROIs from the montage (used in subsequent figures) were defined based on the Brain AnalyzIR Toolbox's depth map function (Santosa et al., 2018). Depth maps show the distance from each fNIRS optode to the superficial cortex of several AAL2 labeled regions of the Colin27 atlas, which can be used to determine coverage of an ROI based on the montage used. As a topological fNIRS layout cannot access depths greater than approximately 30 mm, the channels (lines) projected over yellow or orange regions in **Figure 4.3** (representing depths > 30 mm) are ones that do not reach the specified ROI, whereas channels covering green or blue areas are within range of the nearest cortical point within the ROI.

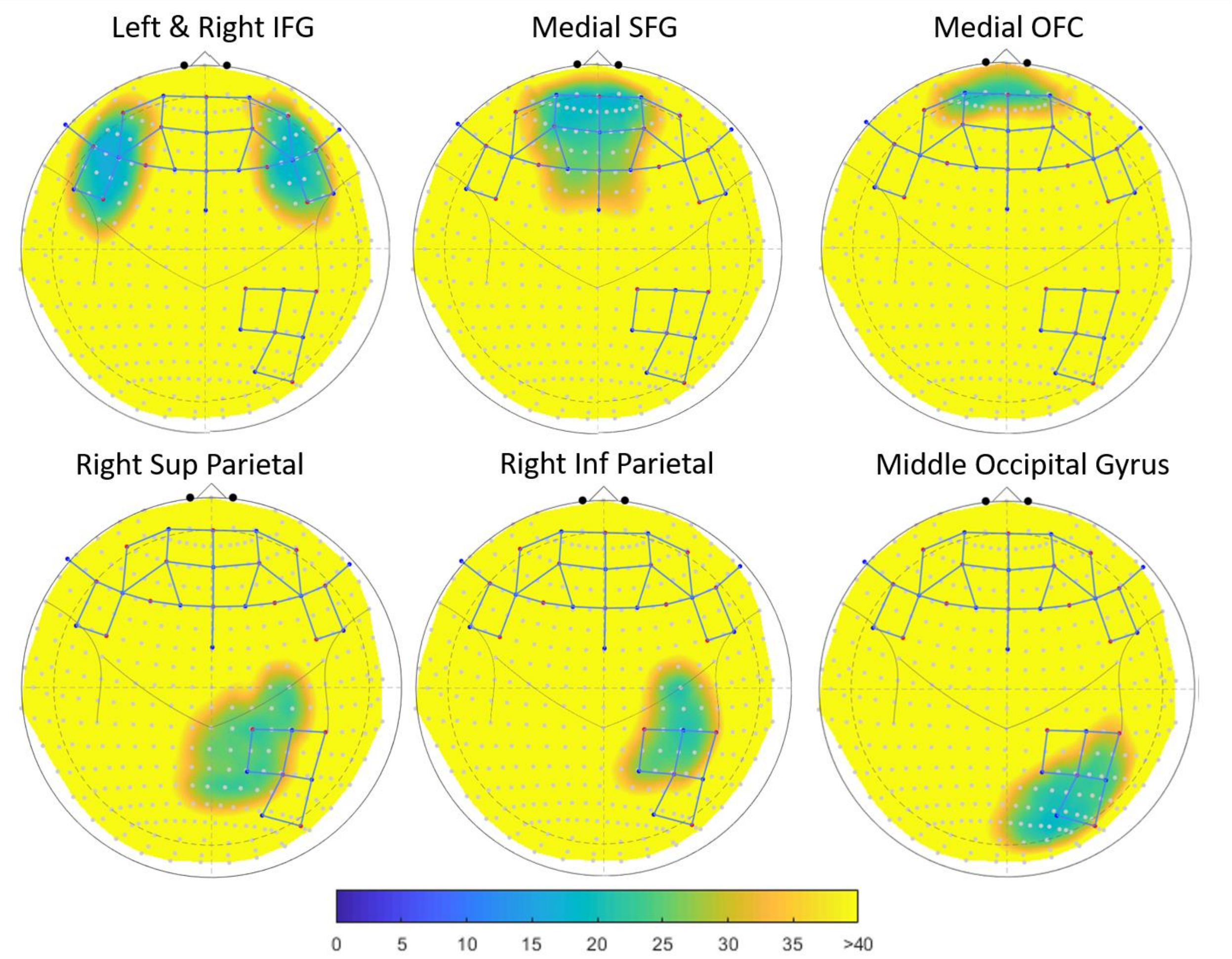


Figure 4.3. Gross ROI depth maps with superimposed montage
fNIRS montage (registered to Colin27 atlas) and depth map for 6 ROIs taken from the AAL2 parcellation: Left and Right Inferior Frontal Gyrus, Medial Superior Frontal Gyrus, Medial Orbitofrontal Cortex, Right Superior Parietal Gyrus, Right Inferior Parietal Gyrus, Middle Occipital Gyrus.

2.7. Analysis

2.7.1. Behavioral Analysis

Accuracy on the n-back task was calculated by taking the average accuracy over the 6 blocks of each n-back level (main round 1-back, 2-back, and 3-back, and post-video 3-back). Accuracy-level differences between levels of the main n-back task were analyzed using repeated measures ANOVAs (function 'ezANOVA' in package 'ez'(Lawrence & Lawrence, 2016), and post-hoc contrasts were conducted using paired t-tests in the 'stats' package in R. Analyses examining change in 3-back as a function of video condition were run using a mixed-effects ANOVA with time as a within-subjects effect and video as a between-subjects effect, again using ezANOVA.

A separate set of behavioral analyses were conducted on participants who scored higher than 60% on all n-back levels. Forty-seven of the 68 participants met this criterion and were included in the analysis.

To analyze change in affect, the average of VAS scores for "happy" and "calm" were used to create a measure of positive affect at baseline (after fNIRS setup and before the main n-back round) and after video viewing. The average of "sad" and "angry" was taken to make a composite measure of negative affect. Change in affect was analyzed as a mixed-effects ANOVA with time as a within-subjects effect and video as a between-subjects effect, also using ezANOVA.

2.7.2. fNIRS Data Analysis: Quality Check

fNIRS data were first loaded into the HOMER2 software package (Theodore J. Huppert et al., 2009) for visual inspection and segmentation into the main n-back task, video, and post-video 3-back task blocks. Visual inspection was done to examine overall data quality (at the level of the participant) and to assess the quality of the parietal data, which was much noisier and more variable than the frontal data. Visual inspection was performed by examining the power spectral density plots for all channels to identify the presence of a cardiac oscillation, which is typically around 1 Hz (Tong et al., 2011). The presence of this cardiac signal is a good indicator that the optical density signals are successfully coupled with a physiological hemodynamic response (Hocke et al., 2018). [See **Figure 4.4**] This method was used to do a first pass evaluation. Based on this visual inspection, 4 participants with unusable data (defined as 5 or fewer clean channels) were identified and excluded in further analysis. Parietal data quality was also examined and logged to determine whether analysis of this region would be fruitful. Of the 62 kept participants, 17 had fully usable parietal data, 20 had mostly usable parietal data (at least half of channels showing good physiological coupling), and 25 had unusable parietal data (only a few usable channels or none).

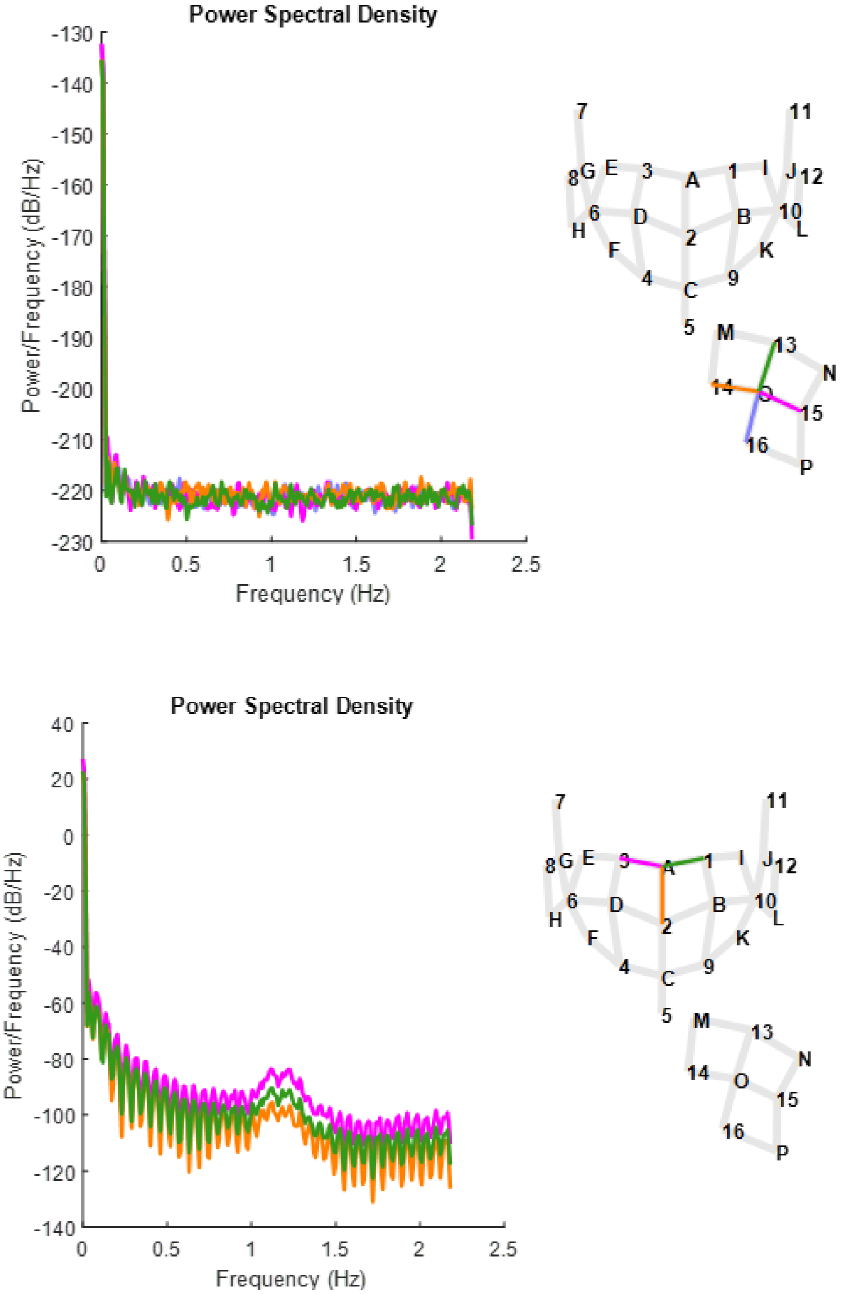


Figure 4.4. Power spectral density plots
Example of channels with good physiological coupling (top) and bad coupling (bottom).

2.7.3. fNIRS Data Analysis: Pre-processing Pipeline and Task-Based Activation

fNIRS data were then analyzed using the NIRS Brain AnalyzIR Toolbox (Santosa et al., 2018). Using this toolbox, the .nirs data (raw light intensity) were loaded into the program, converted into optical density, then converted to oxygenated (HbO) and deoxygenated (HbR) hemoglobin concentrations using the modified Beer-Lambert law (Jacques, 2013).

Once the data were in the form of HbO and HbR concentrations, first level (subject-level) statistics were calculated. As alluded to previously, fNIRS data have unique statistical properties that are not accounted for by typical fMRI-based analysis, and can inflate the type-I error rate (Huppert, 2016). In particular, unlike fMRI, fNIRS suffers from serially-correlated errors (due to a higher sampling rate than the physiological signal of interest) and heavy-tailed noise distributions (due to motion-related artifacts and often, large differences in SNR between channels and between participants; Huppert, 2016). To correct for these issues, the first level general linear model run on individual participants' data uses an autoregressive, iteratively reweighted least-squares model (AR-IRLS). The AR-IRLS model employs an auto-regressive filter (pre-whitening) to deal with the serially correlated errors and uses robust weighted regression to iteratively down-weight outliers due to motion artifacts (Barker et al., 2013). This model saves both the subject level regression coefficients and their error-covariance matrices to be used in statistical tests and contrasts for each subject, and eventually, for use in second-level (group-level) analyses.

In this first-level AR-IRLS model, a hemodynamic response function (HRF) basis set (e.g., canonical HRF, boxcar, full deconvolution) must be specified. Based on research investigating the sensitivity-specificity of basis sets in fNIRS as a function of signal quality and task period (Santosa et al., 2019), a canonical HRF basis was selected for this analysis. Work by Santosa et al. (2019) found that for tasks of sufficiently long durations (> 10 seconds, as in the current study), the canonical HRF performs best in a sensitivity-specificity (ROC) analysis. The canonical model has lower degrees of freedom than a full deconvolution of the raw hemodynamic response (finite impulse response, or FIR model), which improves performance on ROC analysis. This is true at durations of more than 10 seconds, even though there may be a mismatch between the shape of the canonical HRF and the actual hemodynamic response (Santosa et al., 2019).

Based on the output of the first level statistical models, subject-level leverage for the group analyses were calculated and outlier participants (those which contribute significant leverage towards the group results, defined by subject-level leverage of $p < 0.05$) were removed from group-level analyses[3]. Next, second-level

---

[3] In the main n-back analysis, 3 of the 62 participants with usable main n-back data were removed (#s P42, P67, and P70), yielding a final n of 59. In the main 3-back vs. post-video 3-back analysis, 2 of the 62 participants were removed (#s P67 and P70).

statistical models were calculated, which use the full covariance from the first-level models to perform a weighted least-squares regression. Robust regression was also applied to the second-level model to down-weight outliers at the group-level. The results of this analysis were used for group-level contrasts between N-back levels at each channel.

Group activation results are reported as statistical maps using Benjamini-Hochberg false-discovery rate-corrected p-values (e.g., q-values (Benjamini & Hochberg, 1995). This FDR correction is applied to all data in the second-level analysis, including 43 channels, oxy- and deoxy-hemoglobin, and 3 conditions, making the correction very conservative over all tests. The same correction is applied separately for group-level contrasts.

2.7.4. fNIRS Data: Behavioral PLS Analyses

Behavioral PLS analysis (Berman, Misic, et al., 2014; McIntosh & Lobaugh, 2004) https://www.rotman-baycrest.on.ca/index.php?%20section=84) was conducted to identify significant relationships between fNIRS activity and task performance as a function of n-back level. PLS (partial least squares) is a multivariate, data-driven approach often used to examine brain-behavior associations in neuroimaging research by relating two sets or “blocks'' of data to one another (Krishnan et al., 2011). In this study, the fNIRS data block consisted of the regression coefficients (β) from the first-level statistical model (AR-IRLS), corresponding to changes in HbO or HbR for each n-back level relative to baseline for each participant. The behavioral block consisted of average accuracy for each n-back level across blocks of the main n-back task for each participant. Thus, for each PLS (HbO or HbR), each participant had 129 values for the brain activity block (activation betas for each of 3 n-back levels for 43 channels) and 3 values for the behavioral block (average accuracy for each of 3 n-back levels). The goal of this analysis is to find weighted patterns of the input variables, referred to as latent variables or LVs, that maximally co-vary with each other. Therefore, in this behavioral PLS, the LVs represent a specific pattern of the brain-performance relationship, as well as the spatial pattern of activity across all channels that support the particular brain-behavior pattern.

Before running the PLS, histograms of fNIRS beta values were plotted to examine whether the brain data block contained any extreme outliers that may bias the PLS and would be removed in the AnalyzIR Toolbox's robust regression (Theodore J. Huppert, 2016). One participant contained extreme outliers at channel 29 (i.e. beta values < -100 and > 100), and was therefore excluded from PLS analysis.[4] Ten thousand permutation tests were performed to obtain p-values for each latent variable (LV) and 10,000 bootstrap samples with replacement were created to generate the 95% confidence intervals for mean correlation between fNIRS activity and performance for each channel. The bootstrap ratios (salience[weights]/SE[reliability]) measure the reliability of the brain-behavior

---

[4] Direction and significance of results did not change if this participant was included

relationship at each channel, and a larger bootstrap ratio indicates a strong contribution to the LV. In this study, channels with bootstrap ratios larger than +3 or smaller than -3 were determined to be statistically significant.

3. Results

3.1. Behavioral Results

3.1.1. Main N-back performance

Results of the repeated measures ANOVA examining accuracy as a function of n-back level in the main task yielded a significant effect of n-back level on accuracy, $F(2,134) = 93.4$, $p < 0.001$, $\eta_p^2 = 0.58$, 95% CI [0.47 0.66]. As expected, accuracy for the 1-back task ($M = 0.90$, $SD = 0.11$) was significantly better than accuracy for the 2-back task ($M = 0.77$, $SD = 0.17$, $p < 0.001$) and for the 3-back task ($M = 0.71$, $SD = 0.17$, $p < 0.001$). Accuracy for the 2-back task was also significantly higher than for the 3-back task ($p < 0.001$) [**Figure 4.5**].

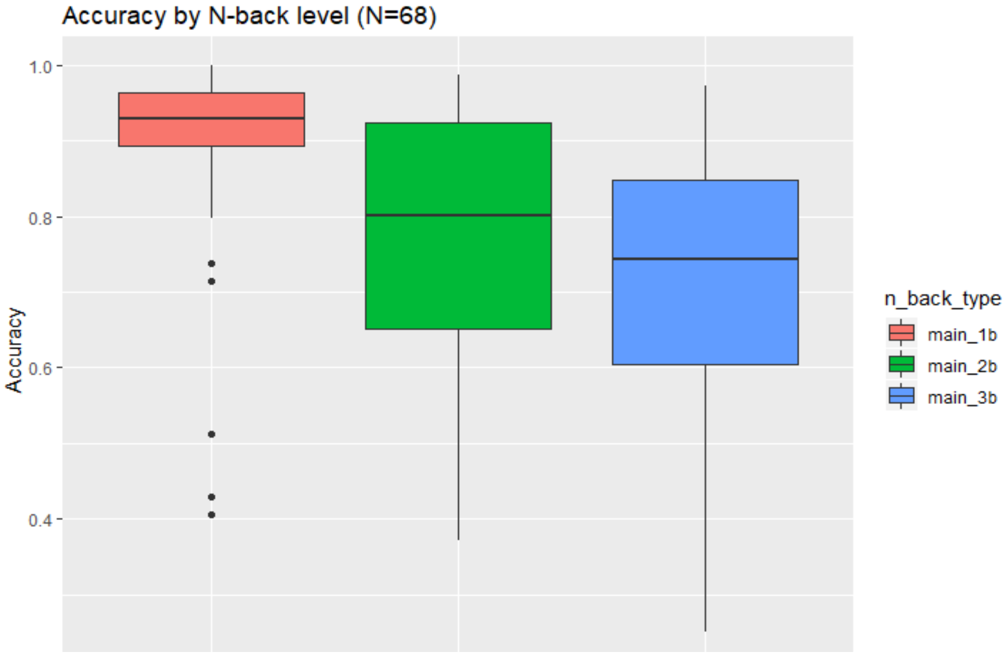


Figure 4.5. Average accuracy in Main N-back Round

The same analysis was run after removing participants who did not reach an average accuracy of 60% on any given n-back level. The results of this analysis did not diverge from those on all participants' data. (Details can be found in Appendix D).

3.1.2. Change in N-back Performance

Results of the ANOVA testing for an effect of environment type (nature/urban video) on change in 3-back performance over time (main round/post-video) yielded a significant main effect of time, $F(1,66) = 17.7$, $p < 0.001$, $\eta_p^2 = 0.21$, 95% CI [0.06 0.37], where participants performed significantly better in the post-video 3-back blocks ($M = 0.76$, $SD = 0.19$) relative to the pre-video 3-back blocks ($M = 0.71$, $SD = 0.17$). No main effect of video or an interaction of video and time was found (all $p >$

0.25) [**Figure 4.6**]. When examined as a change score (post-video 3-back — main 3-back), controlling for main 3-back performance, the results were the same (*p* = 0.85).

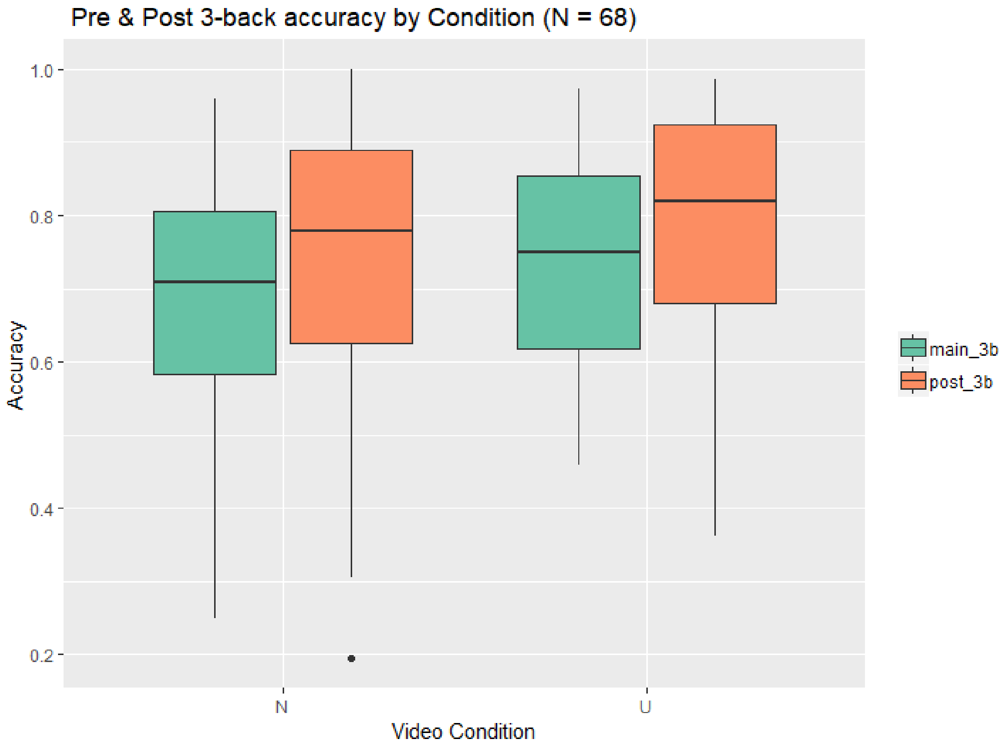


Figure 4.6. Main 3-back (Pre) and Post-Video 3-back Performance by Video Condition

When participants with poor performance (average accuracy for any n-back level < 60%) were removed, the results did not change. (See Appendix D for details of this analysis).

3.1.3. Affect Change

The ANOVA testing the effect of environment (nature/urban video) on change in positive affect over time (baseline/post-video) yielded a main effect of time, $F(1,66) = 8.64$, p = 0.004, $\eta_p^2 = 0.11$, 95% CI [0.01 0.26], with positive affect improving between first assessment (*M* = 4.6, *SD* = 1.12) and post-video viewing (*M* = 4.9, *SD* = 0.99). No effect of video condition or interaction between video and time was found (all *p* > 0.46). No effects were found for change in negative affect (all *p* > 0.46).

3.1.4. Video Preference

Though the videos used were preference-equated in a separate study, average preference ratings for the videos (collected at the very end of the study) were not fully equated in this sample. On average, video preference for participants who viewed the nature video (*M* = 4.76, *SD* = 1.48) was higher than those who viewed the urban video (*M* = 3.66, *SD* = 1.49; *d* = 0.74, *p* = 0.003).

3.1.5. Relationships between Video Preference, Affect, and Cognitive Performance

Correlation analyses were run to see whether individual differences in video preference were predictive of change in cognitive performance and in state positive and negative affect, as well as whether the changes in positive and/or negative affect and cognition were related. Video preference was negatively correlated with change in negative affect: *r* = -0.28, *p* = 0.026, but this relationship was not significant when correcting for the number of correlations examined ($\alpha = 0.05/5 = 0.01$). Video preference was not correlated with change in positive affect (*ps* = 0.78) or with change in 3-back performance (*ps* = 0.96). Counter to previous work, decrease in negative affect was correlated with improvement in 3-back performance, $r = -0.28$, $p = 0.022$, though this effect did not survive with the corrected alpha of 0.01. Positive affect change was not correlated with change in performance (*ps* = 0.18).

3.1.6. Interim Summary: Behavioral Results

Behavioral results showed significant differences in accuracy as a function of n-back level (1-back > 2-back > 3-back) and a significant improvement in accuracy for the post-video 3-back relative to the main blocks of 3-back. Unfortunately, the hypotheses regarding cognitive restoration by environment type were not supported, and change in 3-back accuracy before and after the video intervention was not different between those who watched the nature video versus the urban video.

Overall, participants' positive affect improved between baseline (before the main n-back task) and after video viewing, regardless of which video they watched. However, the videos were not fully equated for participants in this study, and preference for the nature video was higher than the urban video. Consistent with the results of other chapters, individual differences in participants' preference for the video predicted change in affect (specifically, negative affect) but did not relate to change in cognitive performance. In contrast to other work (Berman et al., 2008; Stenfors et al., 2019), we did find a relationship between affect change and cognitive change. However, both of these correlations failed to be significant once adjusting for the number of tests conducted.

3.2. fNIRS Results

3.2.1. Main N-back Task

Relative to baseline, significant increases in oxygenated hemoglobin (HbO) were found for 1 channel (medial SFG) for the 1-back task and for 5 frontal channels and 1 parietal channel for the 2-back task. No channels showed significant increases in HbO concentrations for the 3-back task. [**Table 4.1**] No channels showed significant decreases in HbR for any n-back level.

Table 4.1. Significant Activation by Channel & ROI for each n-back level. Significant channels identified as FDR-corrected $q < 0.05$. ROI defined by maximal coverage of AAL2 parcellation ROI. *p*-value listed is before FDR correction. Power listed is the estimated type-II power for that entry (calculated by computing the minimum detectable change).

	S	**D**	**ROI**	***t-stat***	***p***	***q***	***power***
1-back HbO	1	2	L Medial Superior Frontal Gyrus	3.24	0.001	0.041	0.77
2-back HbO	2	1	R Superior Frontal Gyrus	3.48	0.001	0.032	0.83
	4	2	L Superior Frontal Gyrus	3.28	0.001	0.042	0.78
	4	3	L Superior Frontal Gyrus	3.66	< 0.001	0.023	0.88
	2	10	R Middle Frontal Gyrus	3.3	0.001	0.042	0.79
	11	10	R Inferior Frontal Gyrus	3.86	< 0.001	0.017	0.91
	15	15	R Angular Gyrus	4.41	< 0.001	0.004	0.97
3-back HbO	-	-	-	-	-		

3.2.2. Main N-back Task Contrasts

2-back vs. 1-back

For HbO, 18 channels in bilateral frontal and right parietal cortex showed significantly larger ($q < 0.05$) increases during the 2-back task relative to the 1-back task. No channels yielded larger HbO increases for the 1-back task relative to 2-back.

For HbR, 7 channels, primarily in bilateral IFG, displayed larger decreases for 2-back over 1-back. Additionally, 5 channels, primarily in the right middle occipital gyrus, yielded larger decreases in HbR for 1-back relative to 2-back. [**Figure 4.7, Top Panel**]

3-back vs. 1-back

For HbO, 8 channels, primarily in left and right IFG, yielded significantly larger increases for 3-back relative to 1-back. Larger HbO increases for 1-back over 3-back were found in 7 channels, primarily located in right inferior parietal cortex and left SFG.

For HbR, 5 channels (4 prefrontal, 1 inferior parietal), demonstrated larger deactivation in the 3-back task compared to 1-back. Eight channels (4 frontal and 4 occipito-parietal) showed the opposite pattern. [**Figure 4.7, Middle Panel**]

3-back vs. 2-back

For HbO, 22 channels showed significantly larger increases during the 2-back task compared to the 3-back task. These channels covered bilateral frontal and right parietal areas. Only one frontal channel was greater for the 3-back relative to the 2-back task.

For HbR, 9 channels distributed across bilateral frontal and right parietal cortex showed larger decreases for the 2-back task relative to the 3-back, and 4 channels (2 in medial SFG and 2 in inferior parietal cortex) showed the inverse pattern. [**Figure 4.7 Bottom Panel**]

Figure 4.7. N-back level contrasts for HbO (left) and HbR (right).
Only significant channels ($q < 0.05$) are shown. Channels are displayed on top of 10-20 coordinates and depth maps for left and right Inferior Frontal Gyri, medial Superior Frontal Gyri, and right Superior and Inferior Parietal Gyri. For HbO contrasts, positive t-values (red) correspond to relatively larger activity for the first term in the contrast, and negative t-values (blue) correspond to larger activity for the second term. The opposite pattern applies to HbR contrasts.

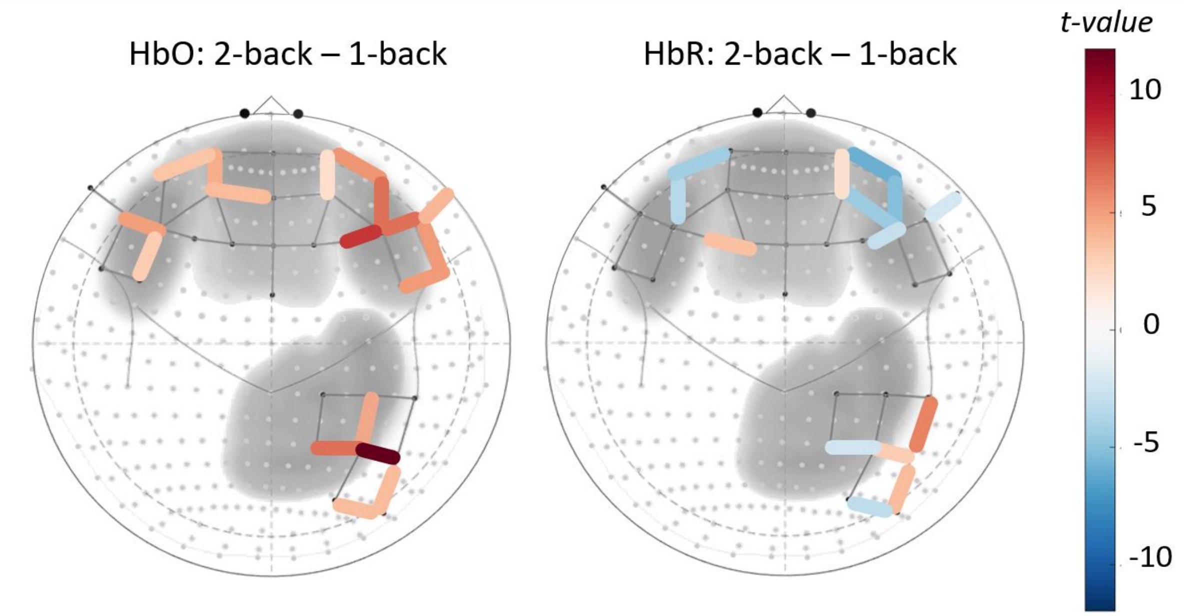
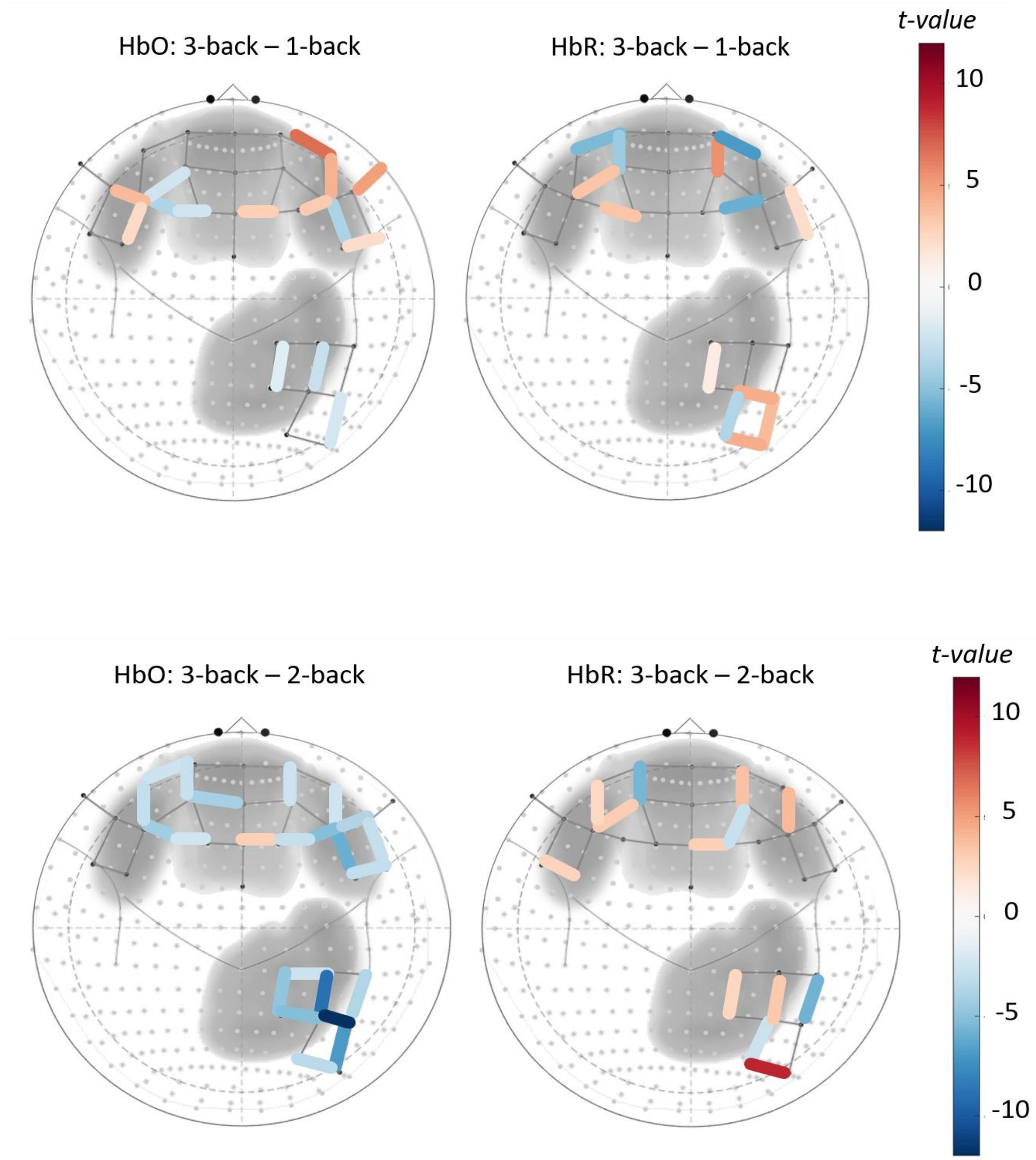


Figure 4.7, continued



3.2.3. Main 3-back vs. Post-video 3-back Task Contrasts

Analyzed across all participants (regardless of video condition), significantly larger increases in HbO were found for post-video 3-back relative to main round 3-back in 9 channels, primarily in left IFG and SFG and right parietal cortex. No channels yielded significantly larger HbO concentration changes for main 3-back relative to post-video 3-back. For HbR, two left IFG channels showed larger decreases for post-video 3-back relative to main 3-back. No channels showed the inverse pattern. [**Figure 4.8**]

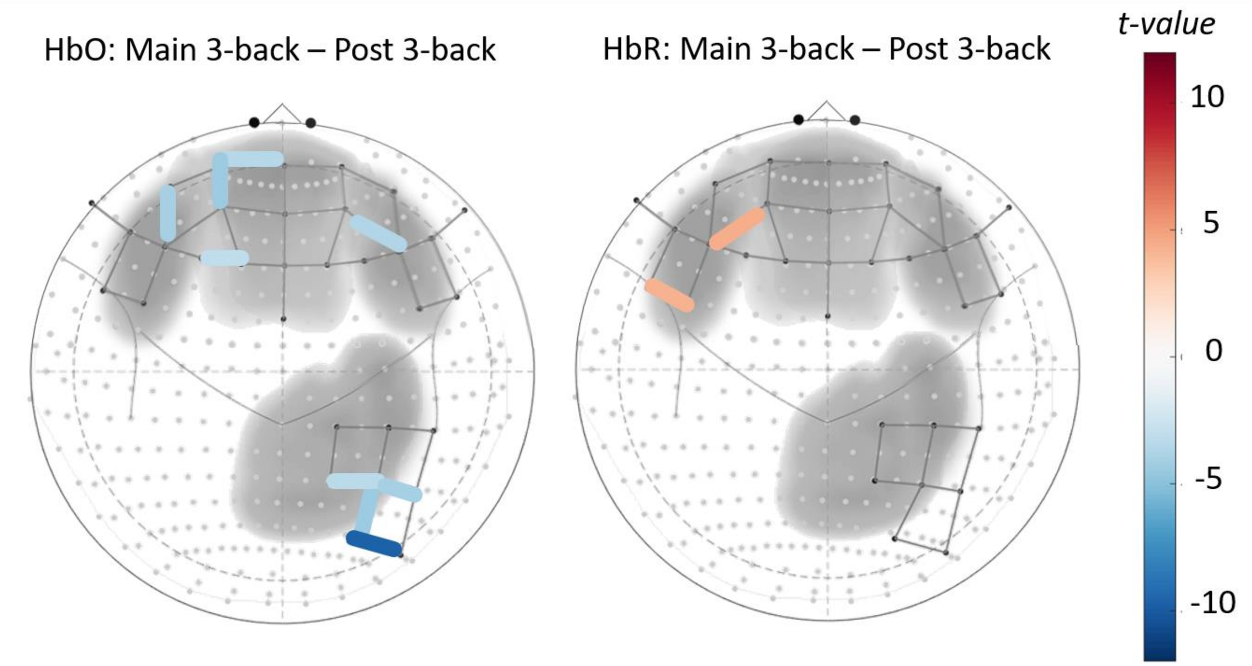


Figure 4.8. Main 3-back vs. Post-video 3-back contrasts for HbO (left) and HbR (right).
Only significant channels ($q < 0.05$) are shown.

3.2.4. fNIRS Activation: Interim Summary

In summary, group level activation maps and contrasts between n-back conditions showed the most consistent results in the 2-back task relative to baseline and comparing activity during the 2-back task relative to the 1-back task. The consistently higher HbO and lower HbR concentration changes during the 2-back task, but not 3-back task, suggest that a minimum level of accuracy may be needed to elicit reliable activation in the fronto-parietal cortical regions examined. Participants overall performed relatively poorly on the main 3-back task. For the 59 participants used in group-level analysis of the main n-back task[5], the average accuracy was 73.6% for the 3-back task. In comparison, average accuracy for these 59 participants was 80% for the 2-back task and 92.3% for the 1-back task.

Further evidence that there is a performance threshold required before significant, reliable activity can be detected by fNIRS activation in these regions comes from the results of the main 3-back task relative to the post-video 3-back task. For the 60 participants included in this analysis[6], accuracy in the post-video 3-back task was significantly higher (78.5%) than the main round 3-back (73.2%). Thus, while our initial hypothesis regarding load-dependent activation (i.e., 3-back >

---

[5] Three of the 62 usable fNIRS participants (#s P42, P67, and P70) were removed due to undue group-level leverage, see section on *fNIRS Data Analysis: Pre-processing Pipeline and Task-Based Activation.*

[6] Two of the 62 usable fNIRS participants (#s P67 and P70) were removed due to undue group-level leverage, see section on *fNIRS Data Analysis: Pre-processing Pipeline and Task-Based Activation.*

2-back > 1-back) was not fully supported, it appears that this effect is dependent upon participants' abilities to perform well on the task.

3.3. Behavioral PLS Analysis - fNIRS Activity ~ Task Performance

Separate behavioral PLS analyses were run to relate performance to concentration changes in HbO and HbR. Though no statistically significant LVs were found for oxyhemoglobin (HbO), the first latent variable from the analysis with deoxyhemoglobin concentrations (HbR) was significant and explained 51% of the crossblock covariance (p = 0.024). Four SFG channels (#4, #8, #12, and #25) showed n-back level dependent changes in HbR as a function of task accuracy [**Table 4.2**]. All of these significant channels had bootstrap ratios < -3 (none were > 3), indicating the direction of the brain-behavior relationship was the same across all four channels. Specifically, for these channels, a larger reduction in HbR (equivalent to increased neural activity) was positively associated with better performance on the 3-back task, unrelated to activity on the 2-back task, and negatively related to performance on the 1-back task. This pattern of results suggests that the neural computations required for high levels of accuracy vary as a consequence of task difficulty/cognitive load. **[Figure 4.9]**

Table 4.2. Significant Channels for LV 1

Channel #	S	D	ROI	Bootstrap Ratio
4	2	1	R Superior Frontal Gyrus	-3.3
25	9	1	R Middle/Superior Frontal Gyrus	-4.3
8	3	2	Medial Superior Frontal Gyrus	-3.8
12	4	2	L Superior Frontal Gyrus	-3.5

Figure 4.9. Results from first latent variable for HbR
LV 1 demonstrated an n-back load-dependent relationship between changes in deoxyhemoglobin concentrations (HbR) and performance. (A) The left panel shows correlation between accuracy and HbR concentration change separately by N-back level. Error bars are 95% confidence intervals around the mean correlation. The right panel shows significant channels (labeled by number), which had bootstrap ratios (BSR) < -3. (B) Scatterplots showing the correlation between HbR (β for task-evoked change from baseline) and performance (accuracy) at each channel, separated by n-back level.

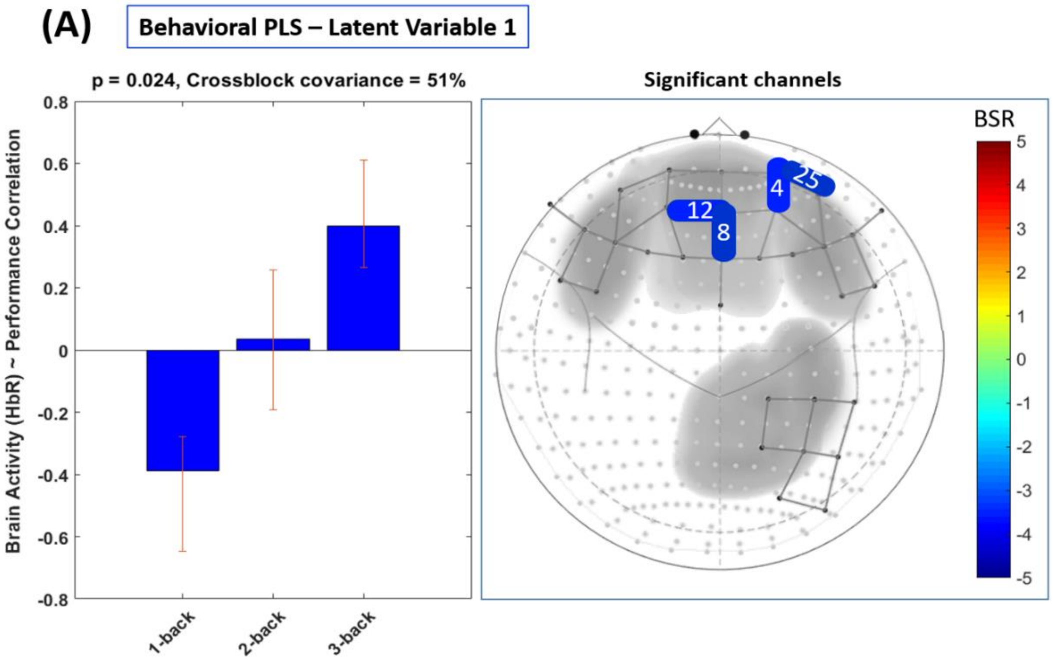
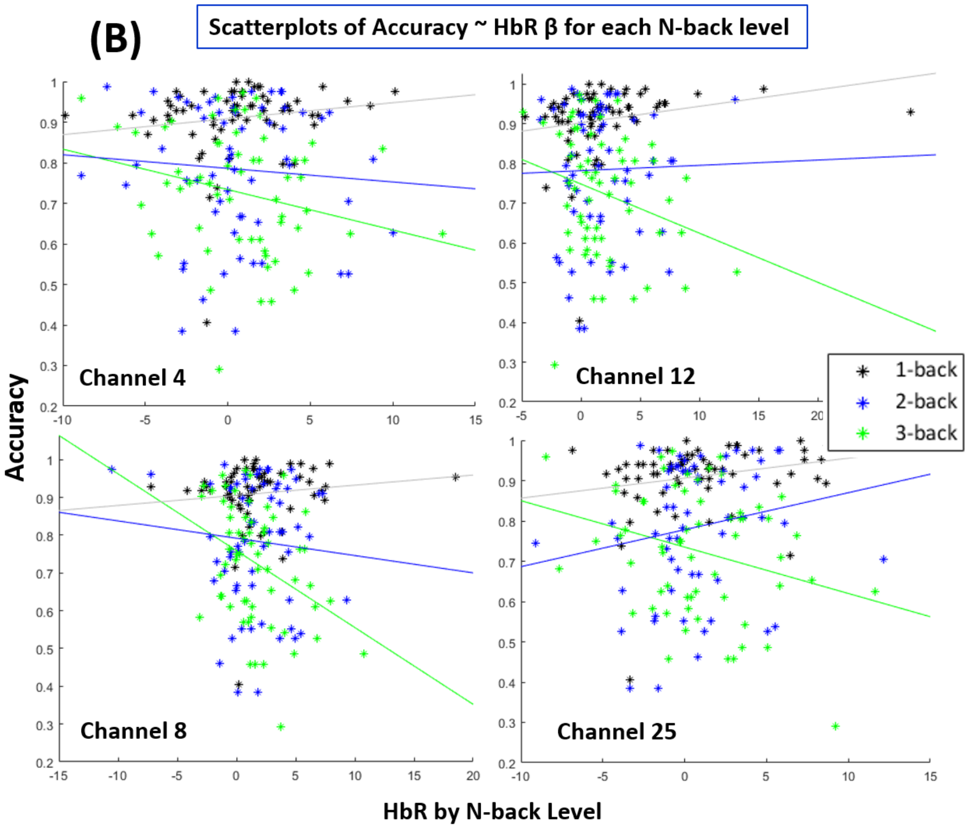


Figure 4.9 continued



3.4. fNIRS differences by Environment Intervention

As there were no behavioral differences between those who viewed the nature video versus those who viewed the urban video, it was hypothesized that there still may be differences in fronto-parietal neural activity between the two groups during the post-video 3-back task. Using the same significance threshold ($q < 0.05$), no baseline differences (main 3-back) were found between the two groups.

Unfortunately, no differences were found in the post 3-back contrast in either HbO or HbR concentration changes[7].

4. Discussion

4.1. Cognitive Restoration in Preference-Matched Environments

This study was designed, in part, to test whether preference-equated nature and urban videos would have distinctive effects on cognitive restoration, as measured by either change in performance on a 3-back task or by hemodynamic signal changes in frontal and parietal neural cortex. Unfortunately, none of the hypotheses regarding cognitive restoration due to nature exposure were supported by the current study.

There are several possibilities as to why the nature exposure did not improve performance to a greater extent than the urban one. One possibility is tied to the virtual nature of the intervention. Though this study used 10 minute videos, which were presumably more immersive environmental exposures than pictures, it is possible that this intervention was still too short or dissimilar to realistic nature interventions (e.g., a nature walk), which more reliably show these cognitive benefits (Berman et al., 2008, 2012; Bratman, Daily, et al., 2015). Though some

---

[7] The AnalyzIR toolbox does not allow for contrasts of contrasts (i.e., does not enable simple subtraction so comparing difference scores is not possible), so the (Post-Main:Nature) - (Post-Main:Urban) contrast was not calculated. However, to examine this, the contrast betas from Post - Main for each condition were included in a Task PLS. This Task PLS was not significant for either HbO or HbR.

studies do elicit cognitive restoration with virtual environmental exposure, the results are usually weaker and less consistent than real nature (Stenfors et al., 2019). Thus, one possibility is simply that the intervention “dose” wasn’t potent enough to elicit such effects.

Another possibility relates to the features of the chosen video stimuli in the framework of Attention Restoration Theory (Kaplan, 1995). In particular, as both videos primarily involved wide, spanning shots of nature or cities and neither video contained any audio, it is plausible that neither video placed significant demands on directed attention resources. In the context of ART, one reason for urban environments’ deleterious effects on cognition is that voluntary attention must be employed to a larger degree in cities (i.e. avoiding traffic and passersby, ignoring advertisements and other irrelevant stimuli) than in unthreatening natural environments (Kaplan & Berman, 2010). Another key element of ART is that for an environment to be restorative, it must have elements that softly capture involuntary attention without overwhelming it. Though the nature video contained many of the elements that are typically thought of as fulfilling this ‘softly fascinating’ criterion, such as waterfalls, trees, and streams, it is possible that some of the elements in our urban video were similarly softly fascinating. For example, the urban video contained several scenes involving churches, town squares, or castles in European cities, which might be interesting or novel enough to gently capture this involuntary attention without being too attention demanding. It remains an open question whether the environments used in this study are sufficiently dissimilar in the features needed to elicit the effects that would be proposed by Attention Restoration Theory.

A third, complementary possibility for these results, relates to the importance of preference in virtual nature exposure. As it is exceptionally difficult to find real preference-equated nature and urban environments, the closest approach that could be taken here was by using virtual exposures (e.g. photos and videos) that separate samples of participants rated on aesthetic preference. However, as some studies using virtual nature exposure with non-preference-equated environments have found improvements in cognitive performance (Berman et al., 2008; Bourrier et al., 2018; Van Hedger et al., 2018), it might simply be that preference is a meaningful ingredient to achieve the benefits of simulated nature exposure. This too remains an open question that demands future investigation.

4.2. The Utility of Functional Near-Infrared Spectroscopy as a Measure of Cognitive Effort

The other key aim of this study was to validate the use of fNIRS for measuring cognitive effort in a laboratory setting, with the ultimate goal of using the device in more ecologically valid, naturalistic environments (e.g., outdoors in nature or cities). Though a number of previous fNIRS studies have examined prefrontal activity using attention demanding working memory tasks such as the N-back (Aghajani et al., 2017; Ayaz et al., 2012; Fishburn et al., 2014; Kuruvilla et al., 2013; Sato et al., 2013), recent work has demonstrated that due to the unique statistical properties of fNIRS, the standard analysis approach (based on fMRI) can severely inflate the false positive rate (Huppert, 2016). Therefore, to provide convergent evidence for previous studies examining load-dependent changes in PFC and parietal cortex, a standard n-back task was employed with a larger sample of participants, and using the recently developed Brain AnalyzIR Toolbox (Santosa et al., 2018) to deal with these fNIRS-specific statistical properties.

Overall, the fNIRS results were consistent with the general hypothesis that tasks placing higher demands on attention and working memory would lead to increased frontal and parietal activation as measured by HbO and HbR concentration changes. This was most evident by the widespread frontoparietal activation elicited by the 2-back task relative to the 1-back task.

Notably, activity in the 3-back task did not follow the hypothesized pattern. This non-linear load effect has been demonstrated in other fNIRS studies (Aghajani et al., 2017; Mandrick et al., 2013, 2016), and seems likely due to poor performance on the task. These results are consistent with the idea that when task demands exceed the current mental capacity of participants, they may disengage from the task and potentially, fail to recruit the necessary cognitive resources (Mandrick et al., 2013). As performance and frontoparietal activity in the main 3-back task was significantly lower than that for the post-video 3-back task, it appears that the post-video task (benefitting from both practice and a 10-minute restoration video) was overall more feasible for participants, leading to appropriate recruitment of relevant cortical areas.

Interestingly, results of the PLS analysis that incorporated individuals' accuracy by n-back level demonstrated evidence for an interaction of load and performance in recruitment of the PFC. Specifically, this multivariate approach showed that changes in deoxyhemoglobin concentrations (HbR) in the medial SFG did not simply decrease with load or performance, but the brain-behavior relationship differed by n-back level. Here, greater reduction in HbR (i.e. more activation) was positively related to performance on the 3-back task, unrelated to accuracy in the 2-back task, and negatively related to accuracy in the 1-back task. This pattern of results suggests more automaticity during the 1-back task (less activation) led to better performance on this relatively easy task, and extensive recruitment of the PFC was required for high accuracy on a more difficult, cognitively demanding 3-back task.

This effect may reflect what has been proposed by the neural efficiency hypothesis: that participants with overall greater cognitive processing ability will show less activation during easy tasks and more during difficult tasks (Dunst et al., 2014; Neubauer & Fink, 2009). This is thought to result from the lower metabolic demands that a "more efficient" brain requires during cognitive tasks. Though the neural efficiency hypothesis is often framed as reflecting individual differences in intelligence, there is also evidence that this effect occurs as a result of more efficient strategies after adequate practice on a specific task (Sayala et al., 2006). Thus, one possibility for this interaction of task difficulty and prefrontal activation is that this reflects individual differences in the learning and adoption of effective strategies during practice. Interestingly, recent neuroimaging work has shown that individuals whose brains are in a more scale-free or fractal state tend to reap the benefits of practice to a greater degree than do those starting in a less scale-free state (Kardan, et al., in prep). Though scale-free neural dynamics have been demonstrated in fMRI and EEG (Churchill et al., 2016; Kardan et al., 2020), whether this signal can be extracted from fNIRS data remains an open question.

Future Directions

It remains important to test whether the null results achieved during virtual environmental exposure with preference-equated stimuli were driven by a non-immersive, generally weaker nature experience versus an important role of preference in eliciting the benefits of (virtual) nature exposure. As such, a subsequent study will be conducted to replicate the cognitive benefits of simulated nature with non-preference equated stimuli (Berman et al., 2008; Bourrier et al., 2018; Van Hedger et al., 2018) with a well-powered sample. If these effects are indeed found with images of nature that are dissimilar in their overall aesthetic preference, this would suggest that at least for virtual nature, preference plays an important role. However, if no changes are found between non-equated environment types, this would suggest that the virtual nature exposure may be insufficient to reliably elicit cognitive restoration.

This study provides important new evidence for the validity of fNIRS as a metric of cognitive effort during a well-used task in a controlled environment, using both a large sample size and appropriate statistical analysis. Though the hypotheses regarding our nature interventions were not supported, the most powerful tests of nature's restorative effects have been found in real nature walks. Therefore, one obvious future study would be to examine cognitive restoration both behaviorally and using fNIRS activity incorporating real walks in nature and urban environments.

Additionally, though not reported here, quantifying the neural activity during the walk itself (not just during the pre- and post-walk cognitive tasks) is a necessary element of the overall goal of this work. Two possibilities for quantifying neural activity during nature or urban environmental exposure would be to examine functional connectivity (Fishburn et al., 2014) and scale-free neural dynamics (Churchill et al., 2016; Kardan et al., 2020) during walks in realistic environments and relate these measures to change in cognitive performance. Before scaling up to an fNIRS study of nature and urban walks, however, it would be important to validate the utility of these two measurements in this more controlled environment. As such, these two analyses on the current data will be conducted to ensure that functional connectivity and scale-free dynamics can be reliably measured, and ideally, predict change in performance on the 3-back task (irrespective of which video was viewed).

General Discussion

The overarching question examined by this dissertation is whether nature preferences account for the affective and cognitive benefits of exposure to natural environments. Several theories have been presented for why natural environments engender these benefits, though they often diverge on the role of nature preferences in the cognitive and affective outcomes. Attention Restoration Theory (ART; Kaplan, 1995), focusing on the cognitive benefits, posits that it is the features of natural environments which may softly capture involuntary attention, while simultaneously placing minimal demands on finite, voluntary attention resources. In this framework, the affective benefits and nature preferences are unrelated, and non-threatening natural environments are not the only type of exposure that could lead to cognitive restoration. For example, cognitive restoration could occur after exposure to a stimulus with features that might meet the softly-fascinating criterion (such as abstract art containing fractal or curved edges) as long as this exposure doesn’t tax directed attention resources. This is in contrast to Stress Reduction Theory (Ulrich, 1991), which posits that an evolutionarily-based affinity for nature underlies the affective improvements and stress reduction, which in turn, frees up cognitive resources.

The results obtained in Chapter 2 suggest that the affective benefits are indeed driven by our aesthetic preference for nature, and there is no added benefit of the naturalness of nature scenes for improving mood. Though not the primary aim of the work, the studies in Chapters 3 and 4 also provided evidence for a link between individuals' preference ratings and the extent to which their affective state improved. A limitation of these studies is that the exposure to nature was in the form of images and videos. This was necessary in order to find preference-equated environments, but the extent to which the effects fully mimic what would be experienced with realistic nature remains an open question.

The primarily null results of Chapters 3 and 4 did not provide evidence to support or negate a role of preference in cognitive restoration after nature exposure. Follow up studies using preference-equated real environments—where the cognitive restoration effects are more reliable—would be ideal to more definitively answer this question. As it is unclear whether the null results stem from something important about preference or the relative weakness of a picture intervention, another approach to test this question is to replicate the original studies which have found effects using non-equated stimuli (Berman et al., 2008) with a sufficiently powered sample. This would provide evidence for whether, at least in non-real nature exposure, preference does matter.

It's noteworthy that Chapter 1 provides some correlational evidence that preference is not important for the cognitive benefits. Children overall did not prefer the natural environments to the same extent as adults, but children with greater exposure to more natural features near their home and/or school scored lower on parent-reported inattentiveness. This was true when controlling for income, and this relationship was unrelated to children's preferences. Further, in this sample, other measures of child behavior such as emotional and peer problems, were unrelated to any measure of nature exposure. These observations, in addition to a large body of literature demonstrating the cognitive benefits of children's interactions with nature (Wells, 2000; Dadvand et al., 2015; Faber Taylor & Kuo, 2015), are suggestive that cognitive restoration in children is independent of preference, in line with what is proposed by Attention Restoration Theory (Kaplan, 1995; Kaplan & Berman, 2010).

References

Aghajani, H., Garbey, M., & Omurtag, A. (2017). Measuring mental workload with EEG+ fNIRS. *Frontiers in Human Neuroscience*, *11*, 359.

Alexander, C. (2002). *The Phenomenon of Life: The Nature of Order: An Essay on the Art of Building and The Nature of the Universe*.

Amoly, E., Dadvand, P., Forns, J., López-Vicente, M., Basagaña, X., Julvez, J., Alvarez-Pedrerol, M., Nieuwenhuijsen, M. J., & Sunyer, J. (2014). Green and blue spaces and behavioral development in Barcelona schoolchildren: the BREATHE project. *Environmental Health Perspectives*, *122*(12), 1351–1358.

Awh, E., Jonides, J., Smith, E. E., Schumacher, E. H., Koeppe, R. A., & Katz, S. (1996). Dissociation of Storage and Rehearsal in Verbal Working Memory: Evidence From Positron Emission Tomography. *Psychological Science*, *7*(1), 25–31.

Ayaz, H., Shewokis, P. A., Bunce, S., Izzetoglu, K., Willems, B., & Onaral, B. (2012). Optical brain monitoring for operator training and mental workload assessment. *NeuroImage*, *59*(1), 36–47.

Balling, J. D., & Falk, J. H. (1982). Development of Visual Preference for Natural Environments. *Environment and Behavior*, *14*(1), 5–28.

Barker, J. W., Aarabi, A., & Huppert, T. J. (2013). Autoregressive model based algorithm for correcting motion and serially correlated errors in fNIRS. *Biomedical Optics Express*, *4*(8), 1366–1379.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, *57*(1), 289–300.

Berman, M. G., Hout, M. C., Kardan, O., Hunter, M. R., Yourganov, G., Henderson, J. M., Hanayik, T., Karimi, H., & Jonides, J. (2014). The perception of naturalness correlates with low-level visual features of environmental scenes. *PloS One*, *9*(12), e114572.

Berman, M. G., Jonides, J., & Kaplan, S. (2008). The Cognitive Benefits of Interacting With Nature. *Psychological Science*, *19*(12).

Berman, M. G., Kross, E., Krpan, K. M., Askren, M. K., Burson, A., Deldin, P. J., Kaplan, S., Sherdell, L., Gotlib, I. H., & Jonides, J. (2012). Interacting with nature improves cognition and affect for individuals with depression. *Journal of Affective*

Disorders, *140*(3), 300–305.

Berman, M. G., Misic, B., Buschkuehl, M., Kross, E., Deldin, P. J., Peltier, S., Churchill, N. W., Jaeggi, S. M., Vakorin, V., McIntosh, A. R., & Jonides, J. (2014). Does resting-state connectivity reflect depressive rumination? A tale of two analyses. *NeuroImage*, *103*, 267–279.

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

Beute, F., & de Kort, Y. A. W. (2014). Natural resistance: Exposure to nature and self-regulation, mood, and physiology after ego-depletion. *Journal of Environmental Psychology*, *40*, 167–178.

Beute, F., & de Kort, Y. A. W. (2018). The natural context of wellbeing: Ecological momentary assessment of the influence of nature and daylight on affect and stress for individuals with depression levels varying from none to clinical. *Health & Place*, *49*, 7–18.

Bourrier, S. C., Berman, M. G., & Enns, J. T. (2018). Cognitive Strategies and Natural Environments Interact in Influencing Executive Function. *Frontiers in Psychology*, *9*, 1248.

Bowler, D. E., Buyung-Ali, L. M., Knight, T. M., & Pullin, A. S. (2010). A systematic review of evidence for the added benefits to health of exposure to natural environments. *BMC Public Health*, *10*, 456.

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*, 118–136.

Bratman, G. N., Hamilton, J. P., Hahn, K. S., Daily, G. C., & Gross, J. J. (2015). Nature experience reduces rumination and subgenual prefrontal cortex activation. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(28), 8567–8572.

Brown, D. K., Barton, J. L., & Gladwell, V. F. (2013). Viewing nature scenes positively affects recovery of autonomic function following acute-mental stress. *Environmental Science & Technology*, *47*(11), 5562–5569.

Browning, M. H. E. M., Mimnaugh, K. J., van Riper, C. J., Laurent, H. K., & LaValle, S. M. (2019). Can Simulated Nature Support Mental Health? Comparing Short, Single-Doses of 360-Degree Nature Videos in Virtual Reality With the Outdoors. *Frontiers in Psychology*, *10*, 2667.

Bunce, S. C., Izzetoglu, K., Ayaz, H., Shewokis, P., Izzetoglu, M., Pourrezaei, K., & Onaral, B. (2011). Implementation of fNIRS for Monitoring Levels of Expertise and Mental Workload. *Foundations of Augmented Cognition. Directing the Future of Adaptive Systems*, 13–22.

Buxton, R. B. (2010). Interpreting oxygenation-based neuroimaging signals: the importance and the challenge of understanding brain oxygen metabolism. *Frontiers in Neuroenergetics*, *2*, 8.

Buxton, R. B. (2013). The physics of functional magnetic resonance imaging (fMRI). *Reports on Progress in Physics*, *76*(9), 096601.

Buxton, R. B., Wong, E. C., & Frank, L. R. (1998). Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. *Magnetic Resonance in Medicine: Official Journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, *39*(6), 855–864.

Cantor, J. H., & Cantor, G. N. (1964). Observing behavior in children as a function of stimulus novelty. *Child Development*, *35*, 119–128.

Christensen, R. H. B. (2018). *Package "ordinal."*

Churchill, N. W., Spring, R., Grady, C., Cimprich, B., Askren, M. K., Reuter-Lorenz, P. A., Jung, M. S., Peltier, S., Strother, S. C., & Berman, M. G. (2016). The suppression of scale-free fMRI brain dynamics across three different sources of effort: aging, task novelty and task difficulty. *Scientific Reports*, *6*, 30895.

Cimprich, B., & Ronis, D. L. (2003). An environmental intervention to restore attention in women with newly diagnosed breast cancer. *Cancer Nursing*, *26*(4), 284–294.

Coburn, A., Kardan, O., Kotabe, H., Steinberg, J., Hout, M. C., Robbins, A., MacDonald, J., Hayn-Leichsenring, G., & Berman, M. G. (2019). Psychological responses to natural patterns in architecture. *Journal of Environmental Psychology*, *62*, 133–145.

Cooper, R. J., Selb, J., Gagnon, L., Phillip, D., Schytz, H. W., Iversen, H. K., Ashina, M., & Boas, D. A. (2012). A systematic comparison of motion artifact correction techniques for functional near-infrared spectroscopy. *Frontiers in Neuroscience*, *6*, 147.

Cope, M., & Delpy, D. T. (1988). System for long-term measurement of cerebral blood and tissue oxygenation on newborn infants by near infra-red transillumination. *Medical & Biological Engineering & Computing*, *26*(3), 289–294.

Corraliza, J. A., Collado, S., & Bethelmy, L. (2012). Nature as a Moderator of Stress in Urban Children. *Procedia - Social and Behavioral Sciences*, *38*, 253–263.

Crompton, J. L. (2001). The Impact of Parks on Property Values: A Review of the Empirical Evidence. *Journal Of Leisure Research*, *33*(1), 1–31.

Dadvand, P., Nieuwenhuijsen, M. J., Esnaola, M., Forns, J., Basagaña, X., Alvarez-Pedrerol, M., Rivas, I., López-Vicente, M., De Castro Pascual, M., Su, J., Jerrett, M., Querol, X., & Sunyer, J. (2015). Green spaces and cognitive development in primary schoolchildren. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(26), 7937–7942.

Debener, S., Minow, F., Emkes, R., Gandras, K., & de Vos, M. (2012). How about taking a low-cost, small, and wireless EEG for a walk? *Psychophysiology*, *49*(11), 1617–1621.

de Vries, S., van Dillen, S. M. E., Groenewegen, P. P., & Spreeuwenberg, P. (2013). Streetscape greenery and health: stress, social cohesion and physical activity as mediators. *Social Science & Medicine*, *94*, 26–33.

Dunst, B., Benedek, M., Jauk, E., Bergner, S., Koschutnig, K., Sommer, M., Ischebeck, A., Spinath, B., Arendasy, M., Bühner, M., Freudenthaler, H., & Neubauer, A. C. (2014). Neural efficiency as a function of task demands. *Intelligence*, *42*(100), 22–30.

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.

Faber Taylor, A., & Kuo, F. E. M. (2011). Could Exposure to Everyday Green Spaces Help Treat ADHD? Evidence from Children's Play Settings. *Applied Psychology. Health and Well-Being*, *3*(3), 281–303.

Fishburn, F. A., Norr, M. E., Medvedev, A. V., & Vaidya, C. J. (2014). Sensitivity of fNIRS to cognitive state and load. *Frontiers in Human Neuroscience*, *8*, 76.

Fleischer, A. (2012). A room with a view—A valuation of the Mediterranean Sea view. *Tourism Management*, *33*(3), 598–602.

Fuegen, K., & Breitenbecher, K. H. (2018). Walking and Being Outdoors in Nature Increase Positive Affect and Energy. *Ecopsychology*, *10*(1), 14–25.

Glasgow, T. E., Le, H. T. K., Scott Geller, E., Fan, Y., & Hankey, S. (2019). How transport modes, the built and natural environments, and activities influence mood: A GPS smartphone app study. *Journal of Environmental Psychology*, *66*, 101345.

Goodman, A., Lamping, D. L., & Ploubidis, G. B. (2010). When to use broader internalising and externalising subscales instead of the hypothesised five subscales on the Strengths and Difficulties Questionnaire (SDQ): data from British parents, teachers and children. *Journal of Abnormal Child Psychology*, *38*(8), 1179–1191.

Grassini, S., Revonsuo, A., Castellotti, S., Petrizzo, I., Benedetti, V., & Koivisto, M. (2019). Processing of natural scenery is associated with lower attentional and cognitive load compared with urban ones. *Journal of Environmental Psychology*, *62*, 1–11.

Harrell, F. E., Jr. (2018). *Package "rms."*

Hartig, T., Böök, A., Garvill, J., Olsson, T., & Gärling, T. (1996). Environmental influences on psychological restoration. *Scandinavian Journal of Psychology*, *37*(4), 378–393.

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

Hartig, T., & Staats, H. (2006). The need for psychological restoration as a determinant of environmental preferences. *Journal of Environmental Psychology*, *26*(3), 215–226.

Herff, C., Heger, D., Fortmann, O., Hennrich, J., Putze, F., & Schultz, T. (2014). Mental workload during n-back task—quantified in the prefrontal cortex using fNIRS. *Frontiers in Human Neuroscience*, *7,* 935.

Hocke, L. M., Oni, I. K., Duszynski, C. C., Corrigan, A. V., Frederick, B. D., & Dunn, J. F. (2018). Automated Processing of fNIRS Data-A Visual Guide to the Pitfalls and Consequences. *Algorithms*, *11*(5).

Hosseini, S. M. H., Bruno, J. L., Baker, J. M., Gundran, A., Harbott, L. K., Gerdes, J. C., & Reiss, A. L. (2017). Neural, physiological, and behavioral correlates of visuomotor cognitive load. *Scientific Reports*, *7*(1), 8866.

Huppert, T. J. (2016). Commentary on the statistical properties of noise and its implication on general linear models in functional near-infrared spectroscopy. *Neurophotonics*, *3*(1), 010401.

Huppert, T. J., Diamond, S. G., Franceschini, M. A., & Boas, D. A. (2009). HomER: a review of time-series analysis methods for near-infrared spectroscopy of the brain. *Applied Optics*, *48*(10), D280–D298.

Huppert, T. J., Hoge, R. D., Diamond, S. G., Franceschini, M. A., & Boas, D. A. (2006). A temporal comparison of BOLD, ASL, and NIRS hemodynamic responses to motor stimuli in adult humans. *NeuroImage*, *29*(2), 368–382.

Ibarra, F. F., Kardan, O., Hunter, M. R., Kotabe, H. P., Meyer, F. A. C., & Berman, M. G. (2017). Image Feature Types and Their Predictions of Aesthetic Preference and Naturalness. *Frontiers in Psychology*, *8*, 632.

Jacques, S. L. (2013). Optical properties of biological tissues: a review. *Physics in Medicine and Biology*, *58*(11), R37–R61.

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.

Joye, Y. (2007). Architectural Lessons from Environmental Psychology: The Case of Biophilic Architecture. *Review of General Psychology: Journal of Division 1, of the American Psychological Association*, *11*(4), 305–328.

Joye, Y., & De Block, A. (2011). “Nature and I are Two”: A Critical Examination of the Biophilia Hypothesis. *Environmental Values*, *20*(2), 189–215.

Joye, Y., Steg, L., Ünal, A. B., & Pals, R. (2016). When complex is easy on the mind: Internal repetition of visual information in complex objects is a source of perceptual fluency. *Journal of Experimental Psychology. Human Perception and Performance*, *42*(1), 103–114.

Joye, Y., & van den Berg, A. (2011). Is love for green in our genes? A critical analysis of evolutionary assumptions in restorative environments research. *Urban Forestry & Urban Greening*, *10*(4), 261–268.

Joye, Y., & van den Berg, A. E. (2018). Restorative environments. *Environmental psychology: An introduction*, 65-75.

Jung, M., Jonides, J., Northouse, L., Berman, M. G., Koelling, T. M., & Pressler, S. J. (2017). Randomized Crossover Study of the Natural Restorative Environment Intervention to Improve Attention and Mood in Heart Failure. *The Journal of Cardiovascular Nursing*, *32*(5), 464–479.

Kahn, P. H., Jr. (1997). Developmental psychology and the biophilia hypothesis: Children's affiliation with nature. *Developmental Review: DR*, *17*(1), 1–61.

Kaplan, R., & Herbert, E. J. (1987). Cultural and sub-cultural comparisons in preferences for natural settings. *Landscape and Urban Planning*, *14*, 281–293.

Kaplan, R., & Kaplan, S. (1989). *The Experience of Nature: A Psychological Perspective*. CUP Archive.

Kaplan, R., & Kaplan, S. (2002). Adolescents and the natural environment: A time out. *Children and Nature: Psychological, Sociocultural, and Evolutionary Investigations*, 227–257.

Kaplan, R., Kaplan, S., & Brown, T. (1989). Environmental Preference: A Comparison of Four Domains of Predictors. *Environment and Behavior*, *21*(5), 509–530.

Kaplan, R., & Yang, B. (1990). The perception of landscape style: a cross-cultural comparison. *Landscape and Urban Planning*, *19*, 252–261.

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

Kaplan, S., & Berman, M. G. (2010). Directed Attention as a Common Resource for Executive Functioning and Self-Regulation. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *5*(1), 43–57.

Kaplan, S., Kaplan, R., & Wendt, J. S. (1972). Rated preference and complexity for natural and urban visual material. *Perception & Psychophysics*, *12*(4), 354–356.

Kardan, O., Adam, K. C. S., Mance, I., Churchill, N. W., Vogel, E. K., & Berman, M. G. (2020). Distinguishing cognitive effort and working memory load using scale-invariance and alpha suppression in EEG. *NeuroImage*, *211*, 116622.

Kardan, O., Demiralp, E., Hout, M. C., Hunter, M. R., Karimi, H., Hanayik, T., Yourganov, G., Jonides, J., & Berman, M. G. (2015). Is the preference of natural versus man-made scenes driven by bottom–up processing of the visual features of nature? *Frontiers in Psychology*, *6*(471), 1–13.

Kardan, O., Gozdyra, P., Misic, B., Moola, F., Palmer, L. J., Paus, T., & Berman, M. G. (2015). Neighborhood greenspace and health in a large urban center. *Scientific Reports*, *5*, 11610.

Kardan, O., Shneidman, L., Krogh-Jespersen, S., Gaskins, S., Berman, M. G., & Woodward, A. (2017). Cultural and Developmental Influences on Overt Visual Attention to Videos. *Scientific Reports*, *7*(1), 11264.

Karmanov, D., & Hamel, R. (2008). Assessing the restorative potential of contemporary urban environment(s): Beyond the nature versus urban dichotomy. *Landscape and Urban Planning*, *86*(2), 115–125.

Kellert, S. R. (1993). The biological basis for human values of nature. *The Biophilia Hypothesis*, 42–69.

Kellert, S. R. (2012). *Building for Life: Designing and Understanding the Human-Nature Connection*. Island Press.

Kellert, S. R., & Wilson, E. O. (1995). *The Biophilia Hypothesis*. Island Press.

Kleinschmidt, A., Obrig, H., Requardt, M., Merboldt, K. D., Dirnagl, U., Villringer, A., & Frahm, J. (1996). Simultaneous recording of cerebral blood oxygenation changes during human brain activation by magnetic resonance imaging and near-infrared spectroscopy. *Journal of Cerebral Blood Flow and Metabolism: Official Journal of the International Society of Cerebral Blood Flow and Metabolism*, *16*(5), 817–826.

Kondo, M. C., Jacoby, S. F., & South, E. C. (2018). Does spending time outdoors reduce stress? A review of real-time stress response to outdoor environments. *Health & Place*, *51*, 136–150.

Kotabe, H. P., Kardan, O., & Berman, M. G. (2017). The nature-disorder paradox: A perceptual study on how nature is disorderly yet aesthetically preferred. *Journal of Experimental Psychology. General*, *146*(8), 1126–1142.

Krishnan, A., Williams, L. J., McIntosh, A. R., & Abdi, H. (2011). Partial Least Squares (PLS) methods for neuroimaging: a tutorial and review. *NeuroImage*, *56*(2), 455–475.

Kuo, F. E., & Sullivan, W. C. (2001a). Environment and Crime in the Inner City: Does

Vegetation Reduce Crime? *Environment and Behavior*, *33*(3), 343–367.

Kuo, F. E., & Sullivan, W. C. (2001b). Aggression and Violence in the Inner City: Effects of Environment via Mental Fatigue. *Environment and Behavior*, *33*(4), 543–571.

Kuo, F. E., Sullivan, W. C., Coley, R. L., & Brunson, L. (1998). Fertile ground for community: Inner-city neighborhood common spaces. *American Journal of Community Psychology, 26*(6), 823-851.

Kuo, M., Barnes, M., & Jordan, C. (2019). Do Experiences With Nature Promote Learning? Converging Evidence of a Cause-and-Effect Relationship. *Frontiers in Psychology*, *10*, 305.

Kuruvilla, M. S., Green, J. R., Ayaz, H., & Murman, D. L. (2013). Neural correlates of cognitive decline in ALS: an fNIRS study of the prefrontal cortex. *Cognitive Neuroscience*, *4*(2), 115–121.

Laumann, K., Gärling, T., & Stormark, K. M. (2003). Selective attention and heart rate responses to natural and urban environments. *Journal of Environmental Psychology*, *23*(2), 125–134.

Lawrence, M. A., & Lawrence, M. M. A. (2016). Package “ez.” *R Package Version*, 4–4.

Lee, J., Park, B.-J., Tsunetsugu, Y., Kagawa, T., & Miyazaki, Y. (2009). Restorative effects of viewing real forest landscapes, based on a comparison with urban landscapes. *Scandinavian Journal of Forest Research / Issued Bimonthly by the Nordic Forest Research Cooperation Committee* , *24*(3), 227–234.

Lesaffre, E., & Spiessens, B. (2001). On the effect of the number of quadrature points in a logistic random effects model: an example. *Applied Statistics*, *50*(3), 325–335.

Lilienthal, L., Tamez, E., Shelton, J. T., Myerson, J., & Hale, S. (2013). Dual n-back training increases the capacity of the focus of attention. *Psychonomic Bulletin & Review*, *20*(1), 135–141.

MacKerron, G., & Mourato, S. (2013). Happiness is greater in natural environments. *Global Environmental Change: Human and Policy Dimensions, 23,* 992-1000.

Macmillan, N. A., & Creelman, D. C. (2004). *Detection Theory: A User’s Guide: Vol. xix* (2nd ed.). Psychology Press.

Mandrick, K., Derosiere, G., Dray, G., Coulon, D., Micallef, J.-P., & Perrey, S. (2013). Prefrontal cortex activity during motor tasks with additional mental load requiring attentional demand: a near-infrared spectroscopy study. *Neuroscience Research*, *76*(3), 156–162.

Mandrick, K., Peysakhovich, V., Rémy, F., Lepron, E., & Causse, M. (2016). Neural and psychophysiological correlates of human performance under stress and high mental workload. *Biological Psychology*, *121*(Pt A), 62–73.

Marteau, T. M., & Bekker, H. (1992). The development of a six-item short-form of the state scale of the Spielberger State-Trait Anxiety Inventory (STAI). *The British Journal of Clinical Psychology / the British Psychological Society*, *31 (Pt 3)*, 301–306.

Mårtensson, F., Boldemann, C., Söderström, M., Blennow, M., Englund, J.-E., & Grahn, P. (2009). Outdoor environmental assessment of attention promoting settings for preschool children. *Health & Place*, *15*(4), 1149–1157.

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.

McCullagh, P. (1980). Regression Models for Ordinal Data. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, *42*(2), 109–142.

McIntosh, A. R., & Lobaugh, N. J. (2004). Partial least squares analysis of neuroimaging data: applications and advances. *NeuroImage*, *23 Suppl 1*, S250–S263.

McMahan, E. A., & Estes, D. (2015). The effect of contact with natural environments on positive and negative affect: A meta-analysis. *The Journal of Positive Psychology*, *10*(6), 507–519.

Mencarelli, L., Neri, F., Momi, D., Menardi, A., Rossi, S., Rossi, A., & Santarnecchi, E. (2019). Stimuli, presentation modality, and load-specific brain activity patterns during n-back task. *Human Brain Mapping*, *40*(13), 3810–3831.

Moore, R. C. (1986). The power of nature: Orientations of girls and boys toward biotic and abiotic play settings on a reconstructed schoolyard. *Children's Environments Quarterly*, *3*(3), 52–69.

Morais, G. A. Z., Balardin, J. B., & Sato, J. R. (2018). fNIRS Optodes' Location Decider (fOLD): a toolbox for probe arrangement guided by brain regions-of-interest.

Scientific Reports, *8*(1), 3341.

Murata, A., Park, J., Kovelman, I., Hu, X., & Kitayama, S. (2015). Culturally non-preferred cognitive tasks require compensatory attention: a functional near infrared spectroscopy (fNIRS) investigation. *Culture and Brain*, *3*(1), 53–67.

Neubauer, A. C., & Fink, A. (2009). Intelligence and neural efficiency. *Neuroscience and Biobehavioral Reviews*, *33*(7), 1004–1023.

Nielsen, T. S., & Hansen, K. B. (2007). Do green areas affect health? Results from a Danish survey on the use of green areas and health indicators. *Health & Place*, *13*(4), 839–850.

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, Critical Reviews*, *19*(7), 305–343.

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping*, *25*(1), 46–59.

Pauling, L., & Coryell, C. D. (1936). The Magnetic Properties and Structure of Hemoglobin, Oxyhemoglobin and Carbonmonoxyhemoglobin. *Proceedings of the National Academy of Sciences of the United States of America*, *22*(4), 210–216.

Piñeyro Salvidegoitia, M., Jacobsen, N., Bauer, A.-K. R., Griffiths, B., Hanslmayr, S., & Debener, S. (2019). Out and about: Subsequent memory effect captured in a natural outdoor environment with smartphone EEG. *Psychophysiology*, *56*(5), e13331.

Pinti, P., Scholkmann, F., Hamilton, A., Burgess, P., & Tachtsidis, I. (2018). Current Status and Issues Regarding Pre-processing of fNIRS Neuroimaging Data: An Investigation of Diverse Signal Filtering Methods Within a General Linear Model Framework. *Frontiers in Human Neuroscience*, *12*, 505.

Pinti, P., Tachtsidis, I., Hamilton, A., Hirsch, J., Aichelburg, C., Gilbert, S., & Burgess, P. W. (2018). The present and future use of functional near-infrared spectroscopy (fNIRS) for cognitive neuroscience. *Annals of the New York Academy of Sciences, 40*, 1-25.

Ravizza, S. M., Delgado, M. R., Chein, J. M., Becker, J. T., & Fiez, J. A. (2004). Functional dissociations within the inferior parietal cortex in verbal working memory.

NeuroImage, *22*(2), 562–573.

Ries, A. J., Touryan, J., Vettel, J., McDowell, K., & Hairston, W. D. (2014). A Comparison of Electroencephalography Signals Acquired from Conventional and Mobile Systems. *Journal of Neuroscience and Neuroengineering*, *3*(1), 10–20.

Rolls, E. T., Joliot, M., & Tzourio-Mazoyer, N. (2015). Implementation of a new parcellation of the orbitofrontal cortex in the automated anatomical labeling atlas. *NeuroImage*, *122*, 1–5.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*.

Ryan, R. M., Weinstein, N., Bernstein, J., Brown, K. W., Mistretta, L., & Gagné, M. (2010). Vitalizing effects of being outdoors and in nature. *Journal of Environmental Psychology*, *30*(2), 159–168.

Salingaros, N. A. (1998). A scientific basis for creating architectural forms. *Journal of Architectural and Planning Research*, *15*(4), 283–294.

Santosa, H., Fishburn, F., Zhai, X., & Huppert, T. J. (2019). Investigation of the sensitivity-specificity of canonical- and deconvolution-based linear models in evoked functional near-infrared spectroscopy. *Neurophotonics*, *6*(2), 025009.

Santosa, H., Zhai, X., Fishburn, F., & Huppert, T. (2018). The NIRS Brain AnalyzIR Toolbox. *Algorithms*, *11*(5), 73.

Sato, H., Yahata, N., Funane, T., Takizawa, R., Katura, T., Atsumori, H., Nishimura, Y., Kinoshita, A., Kiguchi, M., Koizumi, H., Fukuda, M., & Kasai, K. (2013). A NIRS–fMRI investigation of prefrontal cortex activity during a working memory task. *NeuroImage*, *83*, 158–173.

Sayala, S., Sala, J. B., & Courtney, S. M. (2006). Increased neural efficiency with repeated performance of a working memory task is information-type dependent. *Cerebral Cortex* , *16*(5), 609–617.

Schertz, K. E., & Berman, M. G. (2019). Understanding Nature and Its Cognitive Benefits. *Current Directions in Psychological Science*, 0963721419854100.

Schertz, K. E., Saxon, J., Cardenas-Iniguez, C., Bettencourt, L., & Berman, M. (2019). *Neighborhood street activity and greenspace usage uniquely contribute to predicting crime*. https://doi.org/10.31234/osf.io/qzf7a

Scholkmann, F., Kleiser, S., Metz, A. J., Zimmermann, R., Mata Pavia, J., Wolf, U., &

Wolf, M. (2014). A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology. *NeuroImage*, *85 Pt 1*, 6–27.

Schroeter, M. L., Kupka, T., Mildner, T., Uludağ, K., & von Cramon, D. Y. (2006). Investigating the post-stimulus undershoot of the BOLD signal—a simultaneous fMRI and fNIRS study. *NeuroImage*, *30*(2), 349–358.

Shimizu, T., Hirose, S., Obara, H., Yanagisawa, K., Tsunashima, H., Marumo, Y., Haji, T., & Taira, M. (2009). Measurement of frontal cortex brain activity attributable to the driving workload and increased attention. *SAE International Journal of Passenger Cars-Mechanical Systems*, *2*(2009-01-0545), 736–744.

Shin, W. S., Shin, C. S., Yeoun, P. S., & Kim, J. J. (2011). influence of interaction with forest on cognitive function. *Scandinavian Journal of Forest Research / Issued Bimonthly by the Nordic Forest Research Cooperation Committee* , *26*. https://doi.org/10.1080/02827581.2011.585996

Staats, H., Kieviet, A., & Hartig, T. (2003). Where to recover from attentional fatigue: An expectancy-value analysis of environmental preference. *Journal of Environmental Psychology*, *23*(2), 147–157.

Stenfors, C. U. D., Van Hedger, S. C., Schertz, K. E., Meyer, F. A. C., Smith, K. E. L., Norman, G. J., Bourrier, S. C., Enns, J. T., Kardan, O., Jonides, J., & Berman, M. G. (2019). Positive Effects of Nature on Cognitive Performance Across Multiple Experiments: Test Order but Not Affect Modulates the Cognitive Effects. *Frontiers in Psychology*, *10*, 1413.

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, Critical Reviews*, *21*(4), 227–268.

Strangman, G., Culver, J. P., Thompson, J. H., & Boas, D. A. (2002). A quantitative comparison of simultaneous BOLD fMRI and NIRS recordings during functional brain activation. *NeuroImage*, *17*(2), 719–731.

Tang, I.-C., Tsai, Y.-P., Lin, Y.-J., Chen, J.-H., Hsieh, C.-H., Hung, S.-H., Sullivan, W. C., Tang, H.-F., & Chang, C.-Y. (2017). Using functional Magnetic Resonance Imaging (fMRI) to analyze brain region activity when viewing landscapes. *Landscape and Urban Planning*, *162*, 137–144.

Tennessen, C. M., & Cimprich, B. (1995). Views to nature: Effects on attention.

Journal of Environmental Psychology, *15*(1), 77–85.

Thompson, T. W., Waskom, M. L., & Gabrieli, J. D. E. (2016). Intensive Working Memory Training Produces Functional Changes in Large-scale Frontoparietal Networks. *Journal of Cognitive Neuroscience*, *28*(4), 575–588.

Tilt, J. H., Unfried, T. M., & Roca, B. (2007). Using objective and subjective measures of neighborhood greenness and accessible destinations for understanding walking trips and BMI in Seattle, Washington. *American Journal of Health Promotion: AJHP*, *21*(4 Suppl), 371–379.

Tong, Y., Lindsey, K. P., & deB Frederick, B. (2011). Partitioning of physiological noise signals in the brain with concurrent near-infrared spectroscopy and fMRI. *Journal of Cerebral Blood Flow and Metabolism: Official Journal of the International Society of Cerebral Blood Flow and Metabolism*, *31*(12), 2352–2362.

Tost, H., Reichert, M., Braun, U., Reinhard, I., Peters, R., Lautenbach, S., Hoell, A., Schwarz, E., Ebner-Priemer, U., Zipf, A., & Meyer-Lindenberg, A. (2019). Neural correlates of individual differences in affective benefit of real-life urban green space exposure. *Nature Neuroscience, 22,* 1389–1393.

Tsunetsugu, Y., Lee, J., Park, B.-J., Tyrväinen, L., Kagawa, T., & Miyazaki, Y. (2013). Physiological and psychological effects of viewing urban forest landscapes assessed by multiple measurements. *Landscape and Urban Planning*, *113*, 90–93.

Ulrich, R. S. (1979). Visual landscapes and psychological well-being. *Landscape Research, 4*(1), 17-23.

Ulrich, R. S. (1983). Aesthetic and Affective Response to Natural Environment. In I. Altman & J. F. Wohlwill (Eds.), *Behavior and the Natural Environment* (pp. 85–125). Springer US.

Ulrich, R. S. (1993). Biophilia, biophobia, and natural landscapes. *The Biophilia Hypothesis*, *7*, 73–137.

Ulrich, R. S., Simons, R. F., Losito, B. D., Fiorito, E., Miles, M. A., & Zelson, M. (1991). Stress recovery during exposure to natural and urban environments. *Journal of Environmental Psychology*, *11*(3), 201–230.

Valtchanov, D., Barton, K. R., & Ellard, C. (2010). Restorative effects of virtual nature settings. *Cyberpsychology, Behavior and Social Networking*, *13*(5), 503–512.

Valtchanov, D., & Ellard, C. G. (2015). Cognitive and affective responses to natural scenes: Effects of low level visual properties on preference, cognitive load and eye-movements. *Journal of Environmental Psychology*, *43*, 184–195.

van Buuren, S., & Groothuis-Oudshoorn, K. (2010). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 1–68.

van den Berg, A. E., Hartig, T., & Staats, H. (2007). Preference for Nature in Urbanized Societies: Stress, Restoration, and the Pursuit of Sustainability. *The Journal of Social Issues*, *63*(1), 79–96.

van den Berg, A. E., Joye, Y., & Koole, S. L. (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.

Van Hedger, S. C., Nusbaum, H. C., Clohisy, L., Jaeggi, S. M., Buschkuehl, M., & Berman, M. G. (2018). Of cricket chirps and car horns: The effect of nature sounds on cognitive performance. *Psychonomic Bulletin & Review, 26,* 522–530.

Van Hedger, S. C., Nusbaum, H. C., Heald, S. L. M., Huang, A., Kotabe, H. P., & Berman, M. G. (2019). The Aesthetic Preference for Nature Sounds Depends on Sound Object Recognition. *Cognitive Science*, *43*(5), e12734.

Verheyen, C., & Göritz, A. S. (2009). Plain Texts as an Online Mood-Induction Procedure. *Social Psychology*, *40*(1), 6–15.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070.

Wells, N. M. (2000). At home with nature: Effects of “greenness” on children’s cognitive functioning. *Environment and Behavior*, *32*(6), 775–795.

Wells, N. M., & Evans, G. W. (2003). Nearby Nature: A Buffer of Life Stress among Rural Children. *Environment and Behavior*, *35*(3), 311–330.

White, M. P., Pahl, S., Wheeler, B. W., Depledge, M. H., & Fleming, L. E. (2017). Natural environments and subjective wellbeing: Different types of exposure are associated with different aspects of wellbeing. *Health & Place*, *45*, 77–84.

White, M., Smith, A., Humphryes, K., Pahl, S., Snelling, D., & Depledge, M. (2010). Blue space: The importance of water for preference, affect, and restorativeness ratings of natural and built scenes. *Journal of Environmental Psychology*, *30*(4), 482–493.

Xiao, J., Hays, J., Ehinger, K. A., Oliva, A., & Torralba, A. (2010). SUN database: Large-scale scene recognition from abbey to zoo. *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 3485–3492.

Yücel, M. A., Selb, J. J., Huppert, T. J., Franceschini, M. A., & Boas, D. A. (2017). Functional Near Infrared Spectroscopy: Enabling Routine Functional Brain Imaging. *Current Opinion in Biomedical Engineering, 4,* 78–86.

Zink, R., Hunyadi, B., Van Huffel, S., & Vos, M. D. (2016). Mobile EEG on the bike: disentangling attentional and physical contributions to auditory attention tasks. *Journal of Neural Engineering*, *13*(4), 046017.

Appendix A: Supplementary Stimulus Materials for Chapter 1

Additional Stimuli Information: Full sized images of our stimuli can be found here on OSF https://osf.io/axn9q/?view_only=fa88e665f5a74885bd857d79f5a7ce4a. Smaller versions of the picture sets are below, each with a graph showing where the images fall on naturalness and aesthetic preference dimensions (based on ratings of a separate adult sample):

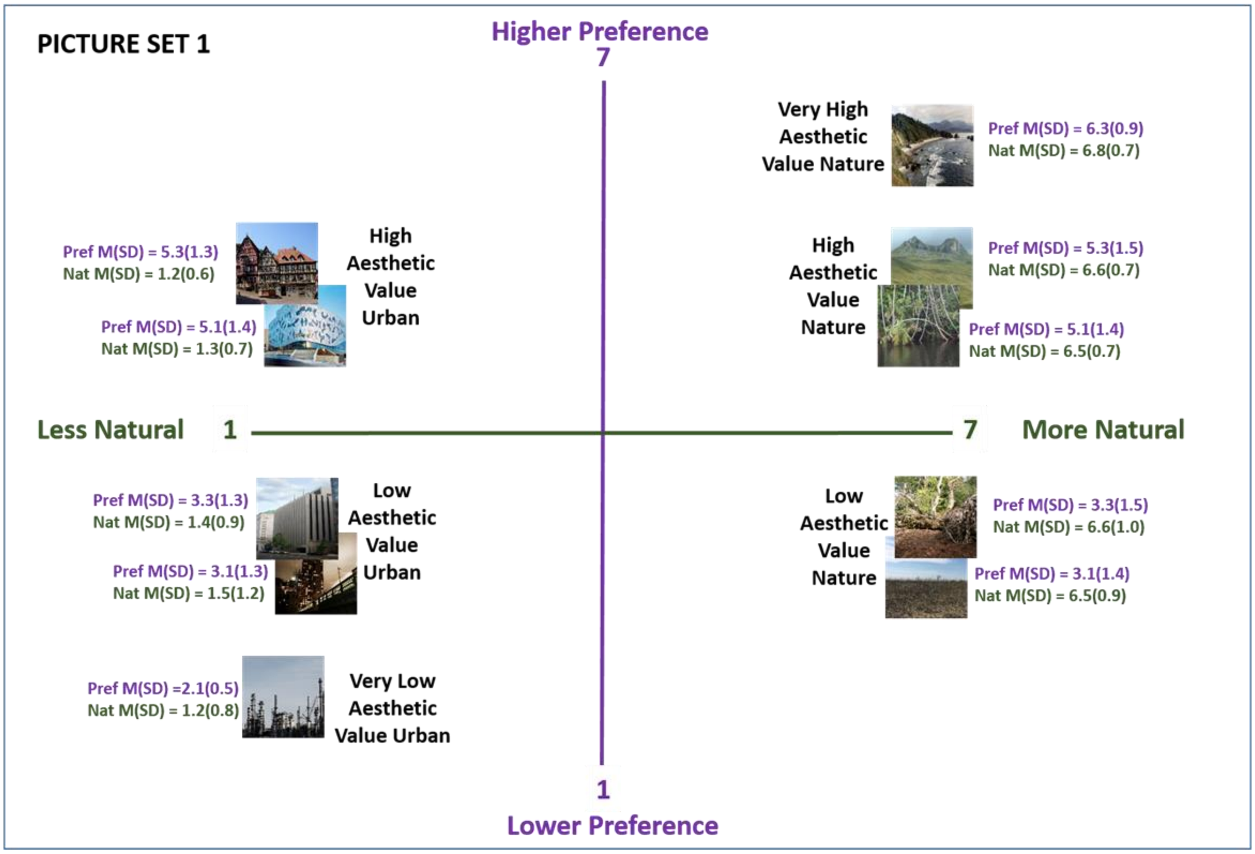


Figure S1.1. Picture Set 1 images along naturalness and preference dimensions.


Very High Aesthetic Nature


High Aesthetic Nature 1


High Aesthetic Nature 2


Low Aesthetic Nature 1


Low Aesthetic Nature 2


High Aesthetic Urban 1


High Aesthetic Urban 2


Low Aesthetic Urban 1


Low Aesthetic Urban 2


Very Low Aesthetic Urban

Figure S1.2. Picture Set 1 Images

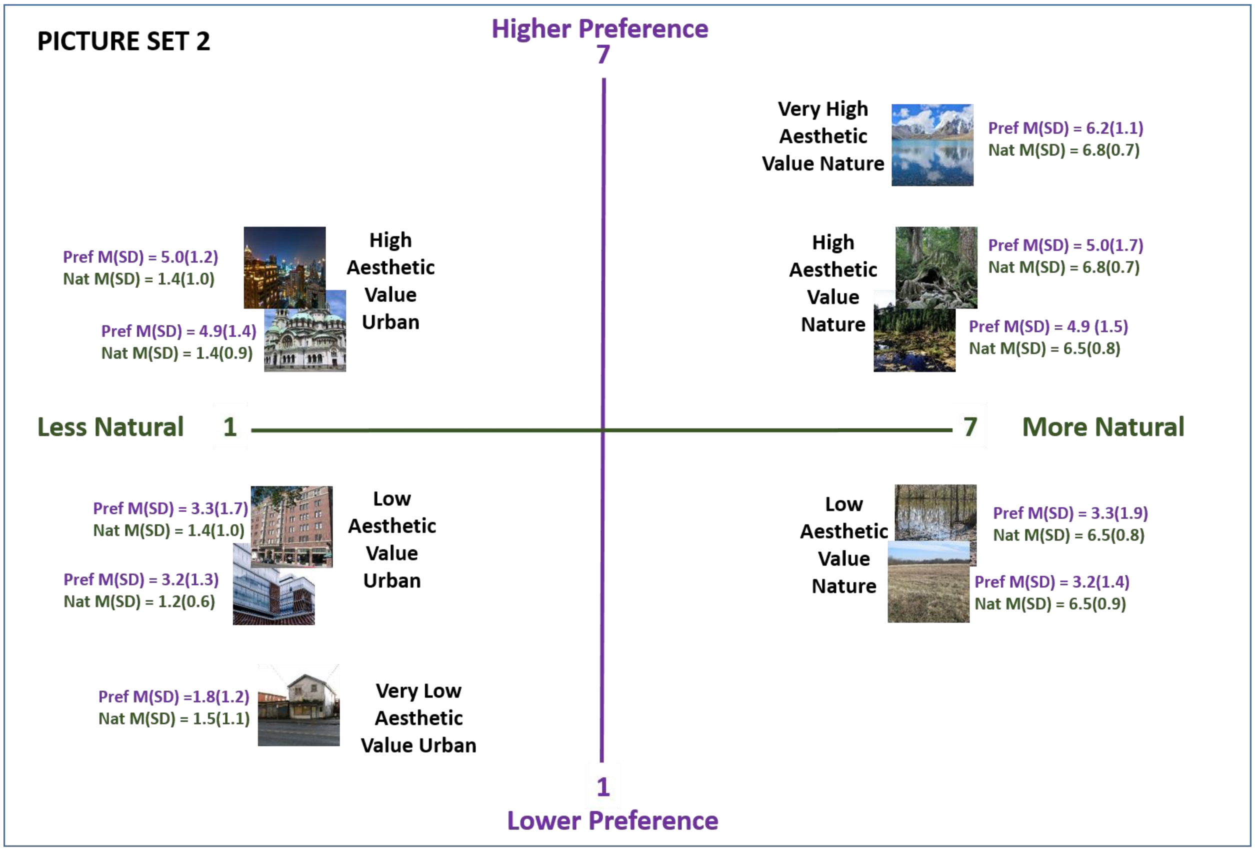


Figure S1.3. Picture Set 2 images along naturalness and preference dimensions.


Very High Aesthetic Nature


High Aesthetic Nature 1


High Aesthetic Nature 2


Low Aesthetic Nature 1


Low Aesthetic Nature 2


High Aesthetic Urban 1


High Aesthetic Urban 2


Low Aesthetic Urban 1


Low Aesthetic Urban 2


Very Low Aesthetic Urban

Figure S1.4. Picture Set 2 Images

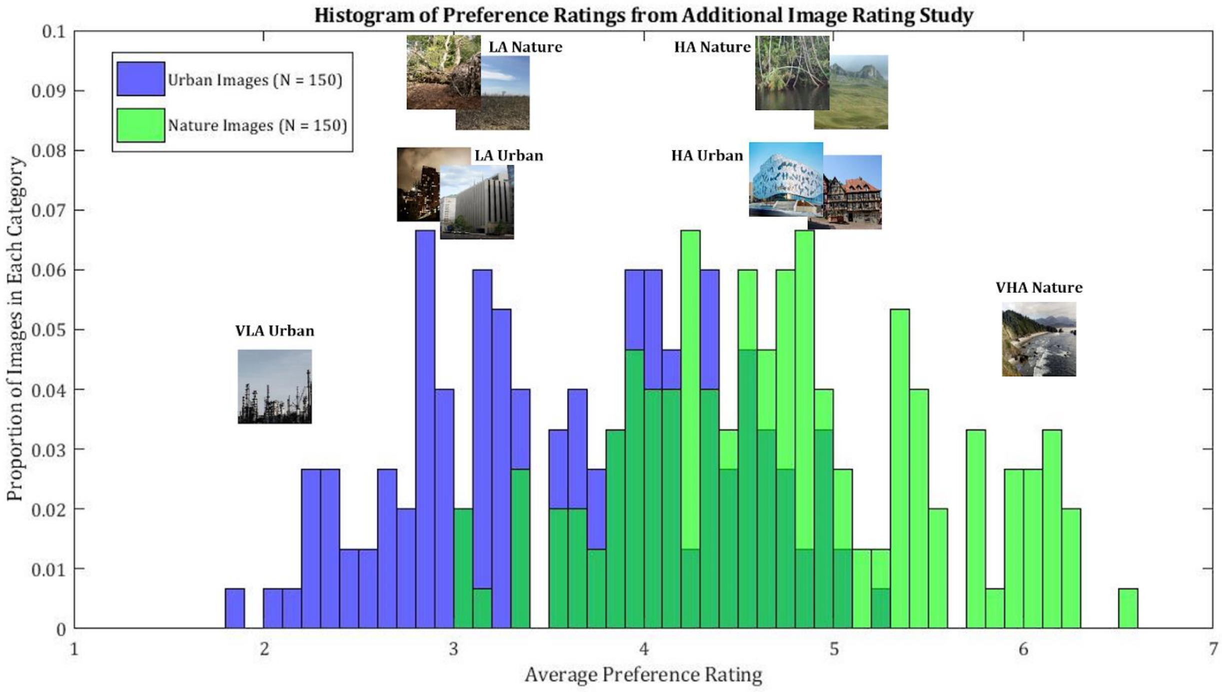


Figure S1.5. Histogram of preference ratings across all images in separate image rating study

These images were chosen based on the data from a separate adult sample. This separate adult sample rated 300 images. The goal of this rating was to find aesthetic preference-matched datasets (with many images in each set) to be used in subsequent studies. Each image was rated by a minimum of 50 adults on both naturalness and preference. This figure plots the distribution of preference ratings, along with the 10 stimuli from Picture Set 1 to show where the images lie on the distribution of aesthetic preference ratings

Appendix B. Supplementary Analyses for Chapter 1

Coded by format

Results from Reported analyses

Results without Excluded Adults (N =15)

Results using Multiple Imputation Procedure

Results using Non-correlated Sample (excluding siblings)

3. Results

3.1 Environmental Preferences differ between Children and Adults

A significant interaction of categorical age and environment type was found **[Reported results: $B = 1.44$, $Z = 11.45$, $OR = 0.27$, 95% CI [0.18, 0.30], $p < 0.001$]** ***[All adults results: $B = 1.38$, $Z = 11.2$, $OR = 0.25$, 95% CI [0.19, 0.32], $p < 0.001$]*** ***[Non-Correlated sample results: $B = 1.37$, $Z = 10.1$, $OR = 0.25$, 95% CI [0.19, 0.33], $p < 0.001$],*** where children showed a greater preference for urban images compared to adults.

Adults and children also differed in their preference for photos based on aesthetic value **[Reported results: $B = 0.16$, $Z = 2.90$, $OR = 0.70$, 95% CI [0.54, 0.89], $p = 0.003$]** ***[All adults results: $B = 0.12$, $z = 1.0$, $OR = 0.88$, 95% CI [0.70, 1.12], $p = 0.3$]*** ***[Non-Correlated sample results: $B = 0.39$, $Z = 2.95$, $OR = 0.67$, 95% CI [0.62, 0.88], $p = 0.003$].*** Specifically, adults preferred the high aesthetic value images (compared to low aesthetic value ones) to a greater extent than did children.

3.2 Children's Environmental Preferences

3.2.1 Group Effects

Results of this analysis showed a main effect of environment, where children generally preferred the urban environments compared to natural ones **[Reported results: $B = -0.83$, $Z = -7.10$, $OR = 0.44$, 95% CI [0.35, 0.55], $p < 0.001$]** ***[Non-Correlated sample results: $B = -0.76$, $Z = -5.7$, $OR = 0.46$, 95% CI [0.36, 0.61], $p < 0.001$].***

There was also a main effect of aesthetic value where children exhibited greater preferences for high aesthetic value images **[Reported results: $B = 1.73$, $Z = 14.5$, $OR = 5.65$, 95% CI [4.47, 7.14], $p < 0.001$]** ***[Non-Correlated sample results: $B = 1.67$, $Z = 12.0$, $OR = 5.33$, 95% CI [4.1, 7.0], $p < 0.001$].***

A significant interaction of environment and aesthetic value was also found **[Reported results: $B = 0.70$, $Z = 4.31$, $OR = 0.50$, 95% CI [0.36, 0.68], $p < 0.001$]** ***[Non-Correlated sample results: $B = 0.66$, $Z = 3.5$, $OR = 0.52$, 95% CI [0.36, 0.75], $p < 0.001$]*** where children exhibited a larger ranking difference between high and low aesthetic value nature scenes compared to urban scenes.

3.2.2. Age-Dependent Preferences

With increasing age, children showed a lessened preference for urban environments over natural ones **[Reported results: $B = 0.32$, $Z = 4.30$, $OR = 1.39$, 95% CI [1.18, 1.63], $p < 0.001$]** ***[Non-Correlated sample results: $B = 0.31$, $Z = 3.2$, $OR =1.36$, 95% CI [1.13, 1.63], $p = 0.001$].***

Additionally, children preferred images of high over low aesthetic value to a greater degree with increasing age **[Reported results: *B* = 0.27, *Z* = 3.31, *OR* = 1.31, 95% CI [1.11, 1.53], *p* < 0.001]** ***[Non-Correlated sample results: B = 0.24, Z = 2.5, OR = 1.27, 95% CI [1.05, 1.53], p = 0.01].***

A significant effect of age was found on rankings of very high aesthetic value nature images **[Reported results: *B* = -0.43, *Z* = -3.62, *OR* = 0.65, 95% CI [0.51, 0.82], *p* < 0.001]** ***[Non-Correlated sample results: B = -0.37, Z = -2.6, OR = 0.69, 95% CI [0.52, 0.91], p = 0.009],*** where higher preference rankings were associated with increasing child age.

There was also a significant effect of age on ranking of very low aesthetic value urban images **[Reported results: *B* = 0.68, *Z* = 5.59, *OR* = 1.98, 95% CI [1.56, 2.52], *p* < 0.001]** ***[Non-Correlated sample results: B = 0.74, Z = 5.17, OR = 2.1, 95% CI [1.59, 2.79], p < 0.001].*** For this category, as age increased, children's preferences for these images decreased.

3.3 Relations between nature exposure, nature preference, and children's behavior

3.3.1 Nearby Nature Predicts Children's Attention

As predicted, greater parent-reported nearby natural features were correlated with lower scores on the SDQ hyperactivity/ inattentiveness subscale **[Reported results: r = -0.21, 95% CI [-0.36, -0.05], p = 0.009] [With Imputation**

results: r = -0.21, 95% CI [-0.36, -0.05], p = 0.008]. However, income was also negatively correlated with hyperactivity/ inattentiveness **[Reported results: r = -0.21, 95% CI [-0.36, -0.04], p = 0.015] [With Imputation results: r = -0.24, 95% CI [-0.4, -0.06], p = 0.009].** Thus, separate analyses were run to see whether natural features were still predictive when income was included in the model.

Results from a linear regression indicated that both greater parent-reported natural features and higher income were significant independent predictors of children's inattentiveness/ hyperactivity **[Reported results: Natural Features B = -0.31, p = 0.015; Income B = -0.31, p = 0.023] [With Imputation results: Natural Features *B* = -0.26, *Z* = -2.32, *p* = 0.022; Income: *B* = -0.29, *Z* = -2.25, *p* = 0.026].** ***When this analysis was conducted on the non-correlated subset of the data which reduced our sample significantly for this analysis (N=106), this effect became non-significant (p = 0.1).*** Surprisingly, play in more natural environments was not significantly related to this SDQ measure (*ps* = 0.2).

3.3.2 Other Nature-Behavior Relations

Results of the full correlation matrix can be found in **Tables S1.1 and S1.2**. As expected, there were strong correlations between NLCD natural features and parent-reported nearby natural features **[Reported results: r = 0.44, 95% CI [0.32, 0.55], p < 0.001] [With Imputation results: r = 0.47, 95% CI [0.32, 0.55], p < 0.001]** as well as NLCD natural features and play in natural environments

[Reported results: r = 0.32, 95% CI [0.17, 0.45], p < 0.001] [With Imputation results: r = 0.32, 95% CI [0.18, 0.44], p < 0.001].

More parent-reported nearby natural features were modestly correlated with lower scores on the SDQ conduct problems subscale **[Reported results:r = -0.15, 95% CI [-0.3, 0.01], p = 0.077] [With Imputation results: r = -0.15, 95% CI [-0.3, -0.01], p = 0.047].** but was not related to any other SDQ subscales. Play in natural environments and nearby natural features calculated from zip-code based national land cover data (NLCD) were not related to any SDQ measures (all *p* > 0.13).

Table S1.1. Correlation matrix of nearby nature, SDQ, demographics, and child environmental preference

Correlation coefficients are listed for all correlations between parent-reported and objective nature, SDQ subscales, SES measures, and children's environmental preferences. Coefficients with ** indicate a *p*-value of less than 0.01, and those with * indicate *p* < 0.05.

	PR Nat Play	NLCD Nat Feat	NLCD Dev	SDQ Emot	SDQ Cond	SDQ Inatt/ Hyper	SDQ Peer	SDQ Prosoc	Income	Mom Ed	Child Env Pref
PR Nat Feat	.37**	.44**	.32**	.00	-.15	-.21**	-.04	.15	.14	.01	-.02
PR Nat Play		.32**	.19**	-.07	-.09	-.11	-.08	.13	.02	.14	.00
NLCD Nat Feat			.56**	-.13	-.09	-.12	-.1	.14	-.06	.01	-.07
NLCD Dev				-.08	-.08	-.1	.02	.14	.05	.06	-.01
SDQ Emot					.31**	.36**	.34**	.26**	-.24**	-.04	.00
SDQ Cond						.49**	.08	-.43**	-.14	-.09	-.13

SDQ Inatt							.16	-.34**	-.21*	-.19*	-.05
SDQ Peer								-.12	-.15	-.02	.05
SDQ Pros									.05	.05	-.03
Income										.45**	-.13
Mom Ed											-.08

[Legend: PR Nat Feat = Parent-reported nearby natural features, PR Nat Play = Parent-reported measure of child's play in nature over built space, NLCD Nat Feat = natural features calculated from zip-code level NCLD data, NLCD Dev = ratio of open/low to medium/high developed space from zip-code level NLCD data, SDQ Emot = Emotional problems subscale, SDQ Conduct = Conduct problems subscale, SDQ Inatt = Inattentive/Hyperactive subscale, SDQ Peer = Peer problems subscale, SDQ Pros = Prosocial behavior subscale, Income = family household income, Mom Ed = Maternal Education, Child Env Pref = Child's preference for natural over urban environments]

Table S1.2. With Imputation. Correlation matrix of nearby nature, SDQ, demographics, and child environmental preference

	PR Nat Play	NLCD Nat Feat	NLCD Dev	SDQ Emot	SDQ Cond	SDQ Inatt/ Hyper	SDQ Peer	SDQ Prosoc	Income	Mom Ed	Child Env Pref
PR Nat Feat	.37**	.47**	.32**	0	-.15*	-.21**	-.06	.13	.15	.03	-.03
PR Nat Play		.32**	.18*	-.08	-.08	-.11	-.1	.11	.05	.14	.02
NLCD Nat Feat			.63**	-.13	-.09	-.1	-.13	.1	-.05	.01	-.03
NLCD Dev				-.08	-.06	-.08	0	.1	.05	.05	-.04
SDQ Emot					.34**	.39**	.36**	.26**	-.3**	-.11	.01
SDQ Cond						.54**	.08	-.44**	-.24*	-.12	-.13

SDQ Inatt	.18*	-.34**	-.24**	-.21*	-.04
SDQ Peer		-.12	-.22	-.08	.07
SDQ Pros			.1	.05	-.03
Income				.46**	-.13
Mom Ed					-.1

ADDITIONAL ANALYSES:

1. Examining the Age x Environment Interaction with an age split (4-6 years vs. 7-11 years) rather than using a continuous measure.

Original Age x Environmental Preference Interaction: B = 0.32, Z = 4.3, $p < 0.001$

Age Split x Environmental Preference Interaction: B = 0.56, Z = 3.4, $p < 0.001$

2. Examining the Age x Aesthetic Preference Interaction with an age split (4-6 years vs. 7-11 years) rather than using a continuous measure.

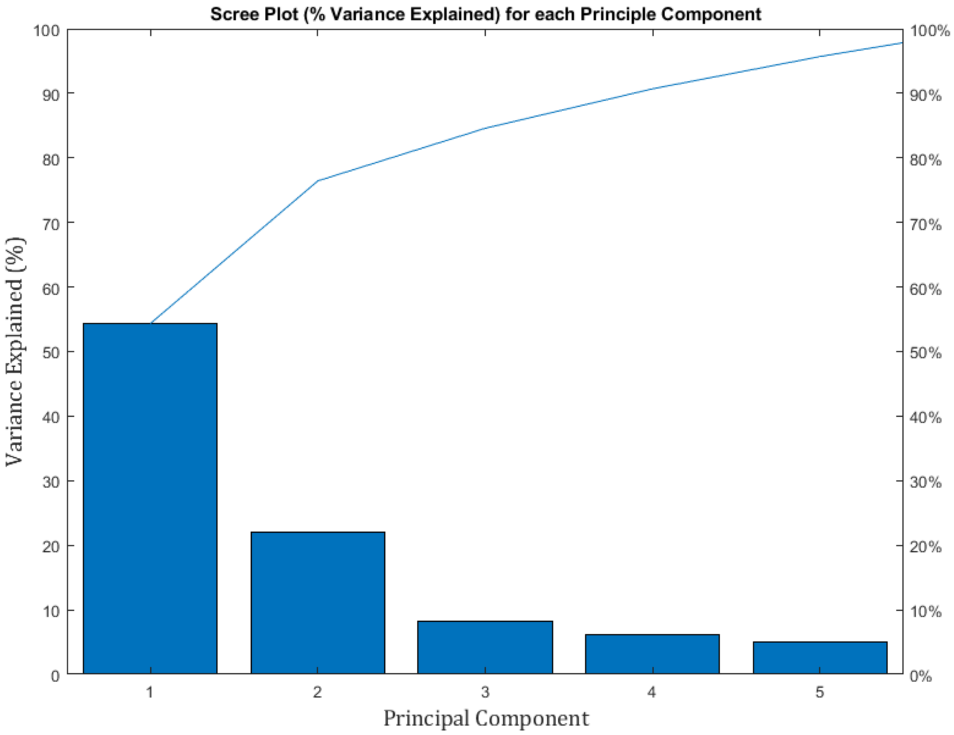
Original Age x Aesthetic Preference Interaction: B = 0.27, Z = 3.31, $p < 0.001$

Age Split x Aesthetic Preference Interaction: B = 0.61, Z = 3.6, $p < 0.001$

Appendix C: Supplementary Analyses for Chapter 2

STUDY 2 Principal Component Analysis

A principal component analysis (PCA) was performed on the baseline (T1) values for the 6 affect measures (VAS for happy, sad, angry, and inspired, and composite scores for STAI Negative Affect items and Positive Affect items), using the function 'pca' in Matlab (Mathworks Inc.). All affect measures were z-scored before the PCA was run. The first two principal components of this analysis were maintained for analysis as they each explained a high proportion of the variance and were easily interpretable. PC 1, which explained 54% of the total variance, mapped onto valence (negative vs. positive affect items). PC 2 explained 22% of the variance and mapped primarily onto high levels of inspiration and generally higher arousal [See **Figure S2.1**]. Next, the six affect measures for post-picture viewing (T2) were z-scored, then multiplied by a vector corresponding to the coefficients for each principal component, and the inner product of this multiplication creates a value for each PC for T2. To calculate a change score for each participant, the value of that participant's PC 1 (valence) at T1 was subtracted from their value of PC 1 at T2. The same calculation was performed for PC 2 (arousal/inspiration).



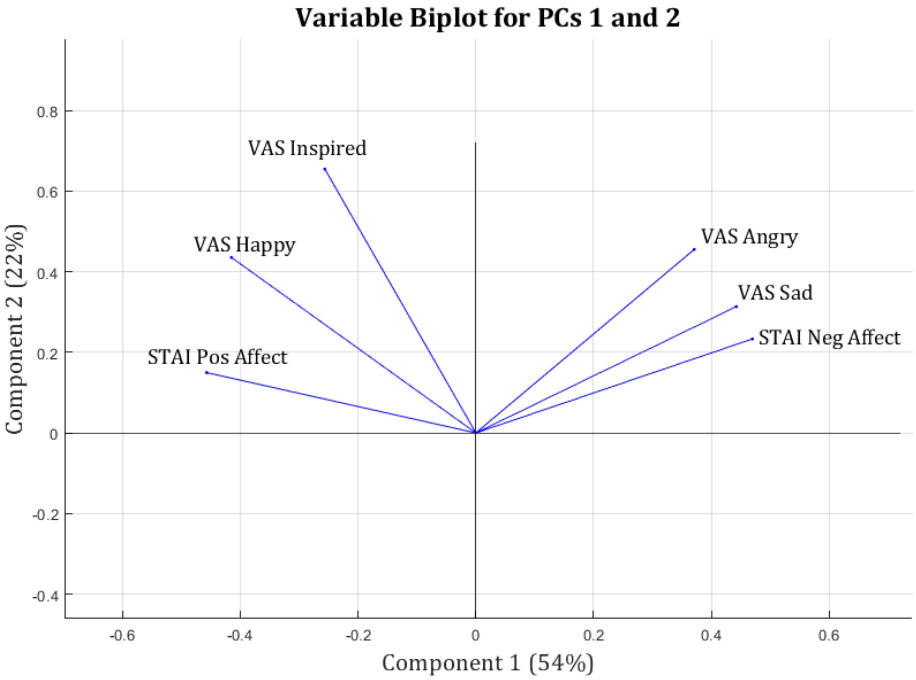


Figure S2.1. PCA for Study 2
TOP: Scree Plot (% variance explained) for each of the 5 principle components
BOTTOM: Variable coefficients plotted for the first two principal components (PCs)

Analysis 2: Affect Change in Completely Crossed Data Subset (Aesthetic Value as Factor)

The first analysis was conducted using only those conditions which were completely crossed (HA-Nat, HA-Urb, LA-Nat, LA-Urb), excluding the VHA-Nat and VLA-Urb conditions. To test the effect of environment and aesthetic value on mood change, 2 (Nat vs. Urb) x 2 (High vs. Low Aesthetic Value) factorial ANOVAs were conducted on change scores (T2 minus T1) for PC 1 (valence) and PC 2 (inspiration/arousal). Results of these analyses did not yield significant effects of environment, aesthetic value condition, or the interaction for either principal component (all $p > 0.15$) [**Figure *S2.2***].

Analysis 3: Affect Change in Completely Crossed Data Subset (Aesthetic Value as Individual Difference Measure)

Analysis 3 was also conducted using only the data with completely crossed conditions, but using participants' own average ratings for the images they viewed, rather than treating aesthetic value as a factor. To do this, two multiple regressions were performed predicting PC 1 and PC 2 by environment type, average image rating, and their interaction.

For PC 1 the overall model was not significant ($R^2 = 0.018$, $F(3,406) = 2.52$, *ps* = 0.058), though participants' average preference ratings were significantly predictive of change in PC 1 ($B = 0.13$, 95% CI [0.02, 0.25], $p = 0.02$). However, neither environment type nor the interaction of environment and individual preference ratings were significantly predictive (all $p > 0.73$). For PC 2, the overall model was also not significant ($R^2 = 0.017$, $F(3,406) = 2.30$, $ps = 0.08$), though again, participants' average preference ratings were significantly predictive of change in PC 2 ($B = 0.09$, 95% CI [0.01, 0.17], $p = 0.02$). Neither environment type nor the interaction of environment and preference ratings were significant predictors (all $p > 0.12$).

Analysis 4: Affect Change between Aesthetic Value Conditions within an Environment

To handle the conditions which were not completely crossed in our design (VHA-Nat and VLA-Urb), separate factorial ANOVAs were conducted on each of the three aesthetic value levels for nature and urban images.

Nature Conditions

A one-way ANOVA with the 3 aesthetic value levels (Very High, High, Low) in the nature condition was performed on each PC. Results of this ANOVA for PC 1 (valence) yielded a significant effect of aesthetic value level $F(2, 306) = 6.39$, $p = 0.002$, $\eta_p^2 = 0.04$, 95% CI [0.006, 0.087]. The partial eta-squared indicates a small-to-medium effect size. Post hoc comparisons were conducted and family-wise error corrected using Tukey's HSD, which showed a significant difference between the VHA-Nat ($M = 0.44$, SD = 0.86) and LA-Nat conditions ($M = 0.0$, $SD = 1.04$, $p = 0.001$, $d = 0.46$, 95% CI [0.18, 0.74]), indicating a greater increase in positive affect/reduction in negative affect for those in the VHA-Nat condition relative to those in the LA-Nat condition. However, there were no differences between VHA-Nat and HA-Nat (M = 0.18, SD = 0.74, *ps* = 0.09) or between HA-Nat and LA-Nat (*ps* = 0.32) [**Figure S2.2**].

Results of this analysis for PC 2 (inspiration/arousal) also showed a significant effect of aesthetic value, $F(2,306) = 5.38$, $p = 0.005$, $\eta_p{}^2 = 0.034$, 95% CI [0.003, 0.079]. Post hoc comparisons showed a significant difference between VHA-Nat ($M = 0.25$, $SD = 0.63$) and HA-Nat ($M = 0.0$, $SD = 0.64$, $p = 0.017$, $d = 0.39$, 95% CI = [0.11, 0.66]) as well as between VHA-Nat and LA-Nat ($M = -0.01$, $SD = 0.64$, $p = 0.011$, $d = 0.41$, 95% CI = [0.13, 0.68]). This difference indicates a greater increase in arousal for participants in the VHA-Nat condition relative to the other nature conditions. No difference was found between HA-Nat and LA-Nat (*ps* = 0.98).

Urban Conditions

For the urban images, a one-way ANOVA with the 3 aesthetic values (High, Low, Very Low) was also performed on each PC. Results of the first ANOVA indicated a significant effect of aesthetic value level on PC 1, $F(2, 303) = 14.42$, $p < 0.001$, $\eta_p{}^2 = 0.087$, 95% CI [0.033, 0.148]. Here, the value of the partial eta-squared indicated a medium effect size. Post hoc comparisons were conducted and family-wise error corrected using Tukey's HSD, which showed a significant difference between the VLA-Urb ($M = -0.68$, $SD = 1.34$) and LA-Urb conditions ($M = 0.07$, $SD = 1.09$, $p < 0.001$, $d = 0.62$, 95% CI = [0.34, 0.90]) as well as between VLA-Urb and HA-

Urb (M = -0.01, SD = 0.81, $p < 0.001$*), indicating a lessened increase in positive affect/reduction in negative affect for those in the VLA-Urb condition relative to those in the LA-Urb and HA-Urb conditions [**Figure S2.2**]. No difference was found for PC1 between HA-Urb and LA-Urb (ps = 0.87). The ANOVA on PC 2 did not show a significant effect of aesthetic value (ps = 0.85).

*effect size d not calculated here due to unequal variances preventing pooling

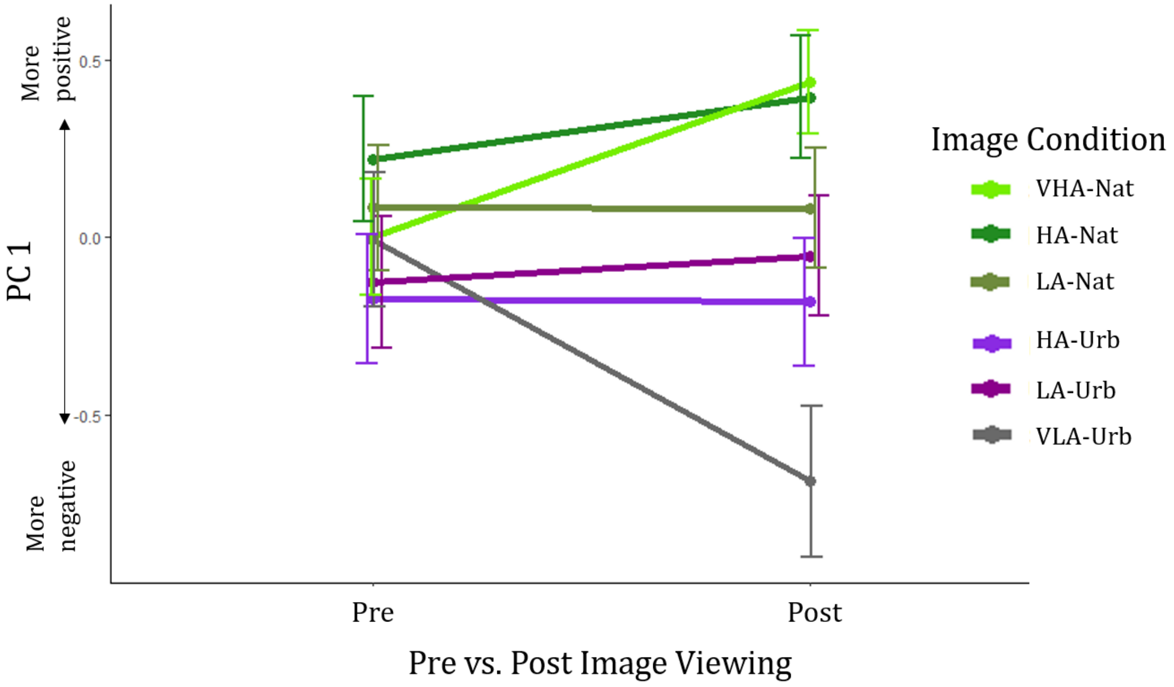


Figure S2.2. Study 2 Change in Affect (PC 1) by Image Condition
Average change in PC1 (affective valence) between baseline (pre) and after image viewing (post) for each of the 6 image conditions. Error bars represent SEM.

Analysis 5: Affect Change as Predicted by Individual Preference Ratings in full Dataset

To examine whether individual differences in participants' preference ratings for the images were related to changes in affect, two linear regressions were conducted to predict change in PCs 1 and 2 by individuals' average preference rating. These analyses were conducted on the full dataset (not only on the completely-crossed conditions). Results of this analysis on PC 1 (valence) indicated that average image preference rating explained 9.2% of the variance in PC 1 change ($R^2 = 0.092$, $F(1,613) = 61.92$, $p < 0.001$). In this case, a higher average preference rating for the images viewed significantly predicted a greater increase in positive affect/decrease in negative affect ($\beta = 0.3$, 95% CI [0.23, 0.38], $p < 0.001$).

For PC 2 (inspiration/arousal), average image rating explained 3.1% of the variance in PC 2 change ($R^2 = 0.031$, $F(1,613) = 19.58$, $p < 0.001$). Here, higher preference ratings also significantly predicted a greater increase in arousal ($\beta = 0.18$, 95% CI [0.10, 0.25], $p < 0.001$).

STUDY 4 Principal Component Analysis

As in Study 2, a principal component analysis (PCA) procedure was performed on the affect measures. Though the pre-registered analyses specified running this PCA on all participants, there were substantial differences in the T1 data based on mood induction condition which affected the PCA decomposition (i.e., MIP significantly affected the covariance between affect measures, which subsequently altered the PC composition, making them difficult to compare). Therefore, the PCA was conducted separately on the participants who underwent the neutral MIP and those that underwent the negative MIP.

The first two principal components of this analysis were maintained for analysis upon examination of the scree plots. For the neutral MIP, principle component 1 explained 43% of the total variance and mapped onto valence (negative vs. positive affect items). PC 2, which explained 29% of the variance, was related to arousal. In particular, this component seemed to map onto positive arousal where Inspired and Happy were the highest variable loadings [**Figure S2.3**]. For the negative MIP, the 1st PC was very similar to that of the neutral MIP as it also explained 43% of the variance and was related primarily to valence. PC 2 for this group was somewhat different than the neutral MIP group, explaining 23% of the total variance and primarily relating to general arousal, loading highest on Angry and Inspired [**Figure S2.4**]. The directionality of PC 1 was inverted in the neutral group, where lower values indicated more positive affect and higher values indicated more negative affect. Thus to compare the two MIP conditions, values for PC 1 in the neutral condition were multiplied by -1 to place them in the same direction as the negative condition. Additionally, because they were so similar, all of the PC 1 data were analyzed together. As PC 2 differed between the MIP conditions, analyses were conducted separately on the negative and neutral MIP groups. All statistical analyses were conducted using the same methods as Study 1.

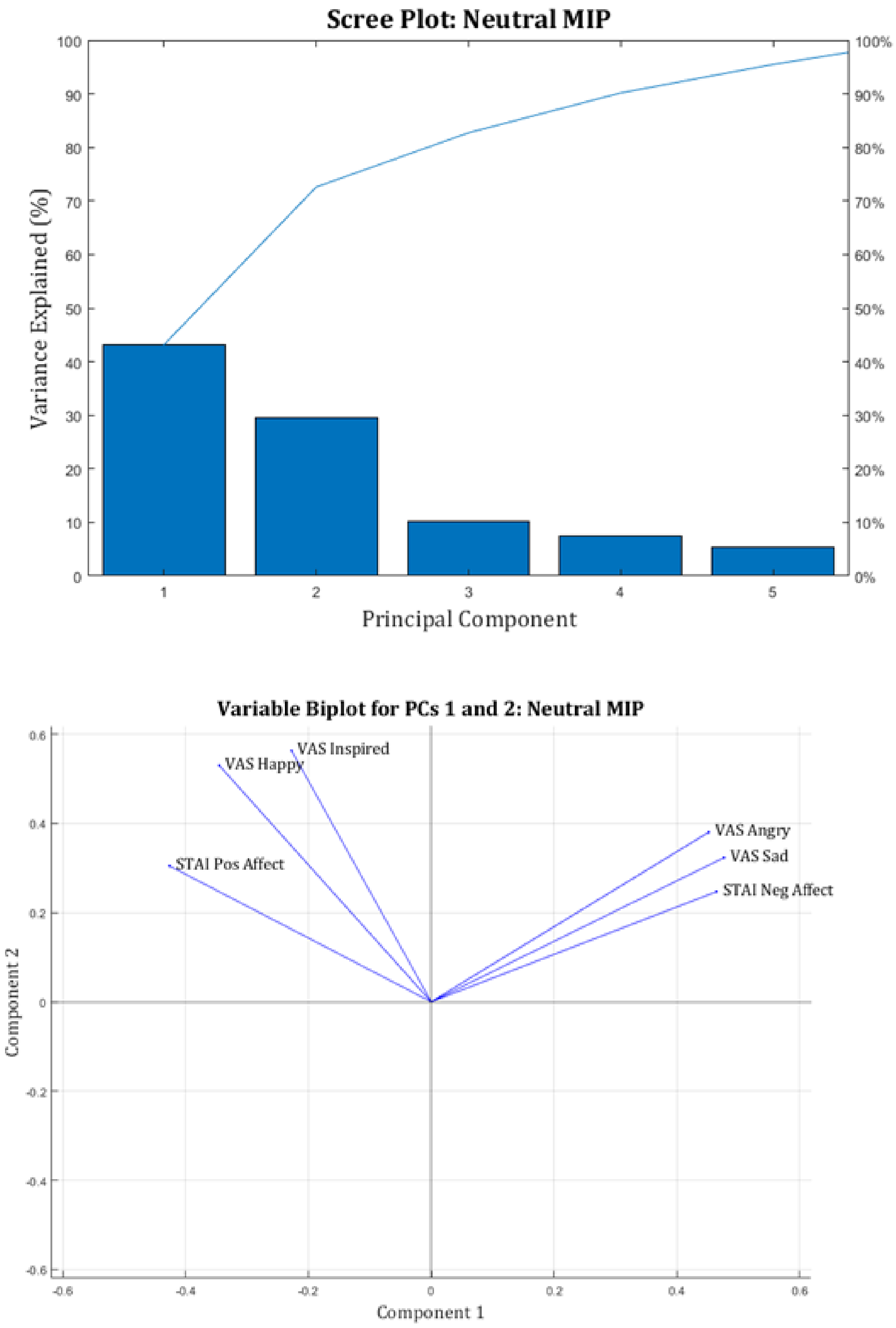


Figure S2.3. Results of PCA in Neutral MIP condition
TOP: Scree Plot (% variance explained) for each of the 5 PCs for the Neutral MIP
BOTTOM: Variable Biplot for the first two PCs for the Neutral MIP.

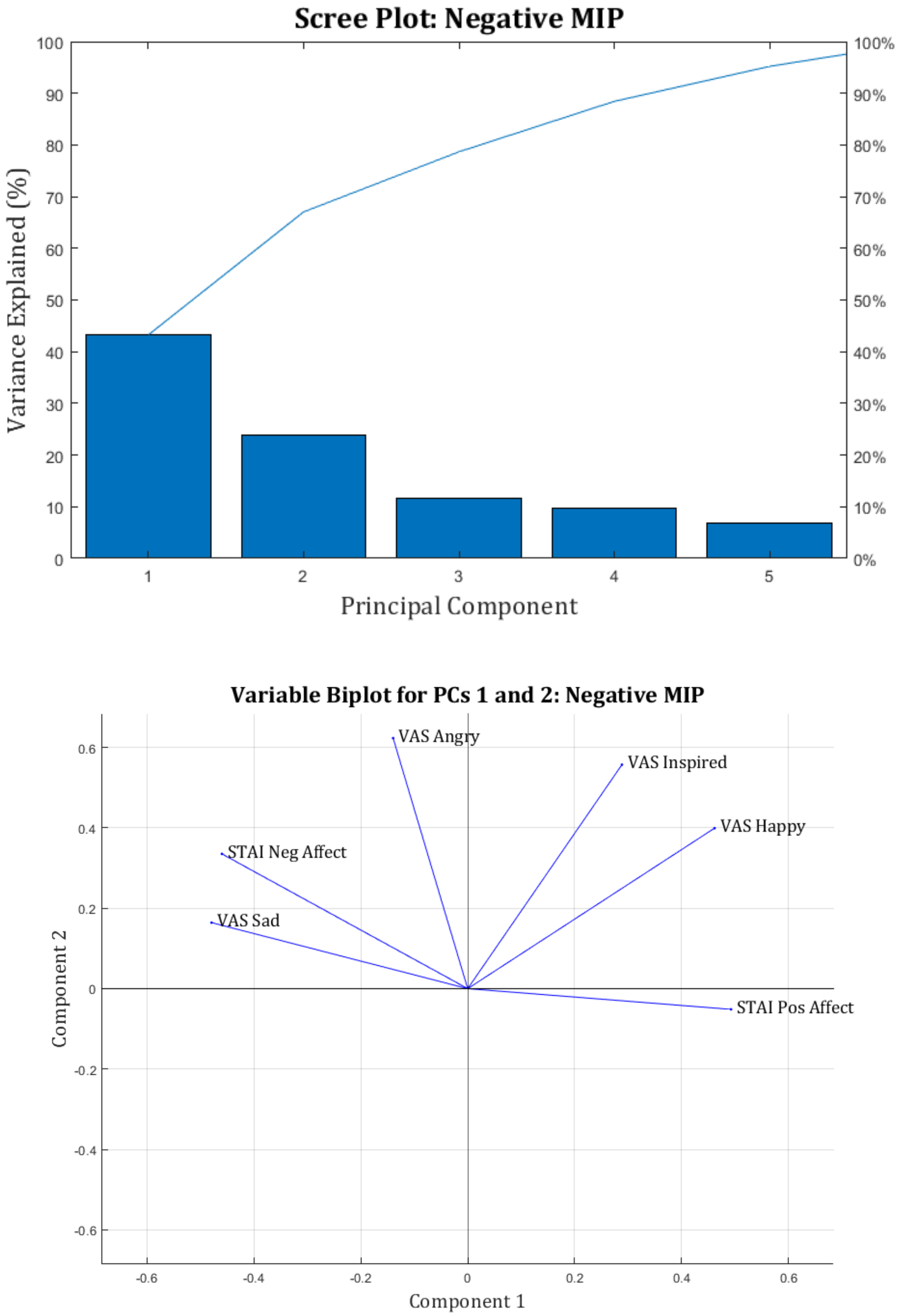


Figure S2.4. Results of PCA in Negative MIP condition
TOP: Scree Plot (% variance explained) for each of the 5 PCs for the Negative MIP
BOTTOM: Variable Biplot for the first two PCs for the Negative MIP.

Analysis 1: Affect Change by Image Condition and MIP

To test the effect of image category and mood induction on affect change, a factorial ANOVA with Image Category (Nature, Urban, Animal) and Mood Induction (Negative, Neutral) as between-subjects variable were conducted on change (T2 minus T1) for PC 1 (valence). For PC 2, two one-way ANOVAs were conducted using PC 2 change as the outcome variable and Image Category as the between-subjects factor. These PC 2 analyses were done separately on the negative and neutral MIP conditions.

PC 1 change (valence) across MIPs

Results of this ANOVA for PC 1 (valence) yielded a trend main effect of image condition, $F(2,598) = 2.95$, $p = 0.053$, $\eta_p^2 = 0.01$, 95% CI [0.00, 0.029]. The partial eta-squared indicates a very small effect of condition. Post hoc comparisons, conducted using Tukey's HSD, showed a significant difference between the Animal ($M = 0.20$, $SD = 1.71$) and Urban categories ($M = -0.20$, $SD = 1.45$, $p = 0.04$, $d = 0.25$, 95% CI [0.05, 0.45]), indicating a greater increase in positive affect/reduction in negative affect for those in the Animal images condition relative to those in the Urban images condition. There were no differences between Animal and Nature conditions ($M = 0.00$, $SD = 1.65$, $ps = 0.43$) or between Nature and Urban ($ps = 0.45$). There was not a significant effect of mood induction procedure ($ps = 0.99$) or an image condition by MIP condition interaction ($ps = 0.88$).

PC 2 (arousal) change in each MIP

Neither the ANOVA examining changes in PC 2 in neutral or negative MIPs yielded a significant effect of image condition on mood change (neutral: *ps* = 0.13, negative = 0.33).

Analysis 2: Affect Change as Predicted by Image Condition, Mood Induction, & Individual Preference Ratings

As the participants in this study rated the images somewhat differently from those in Study 2A, it is possible that the image condition trending effect was due to overall higher ratings for the Animal images versus Urban ones. As such, multiple regressions were conducted to examine the respective contributions of image type and individuals' average preference ratings.

Results of this analysis on PC 1 (valence) is presented in **Table S2.1.** The overall model was significant and explained 7.9% of the variance in PC 1 change ($R^2 = 0.079$, $F(4,597) = 12.79$, $p < 0.001$). In this case, a higher average preference rating for the images viewed significantly predicted a greater increase in positive affect/decrease in negative affect ($B = 0.47$, 95% CI [0.34, 0.61], $p < 0.001$). Importantly, picture condition did not significantly contribute to change in PC 1 [**Figure S2.5**].

Table S2.1. Regression results using PC1 Change as the criterion

Predictor	*b*	*b* 95% CI [LL, UL]	*sr*^{2}	*sr*^{2} 95% CI [LL, UL]	Fit
(Intercept)	-2.63**	[-3.49, -1.76]			
Avg_Pic_Rate	0.47**	[0.34, 0.61]	.07	[.03, .11]	
Pic_Con_Nat	-0.10	[-0.41, 0.20]	.00	[-.00, .00]	
Pic_Con_Urb	-0.05	[-0.37, 0.28]	.00	[-.00, .00]	
MIP_Con	0.06	[-0.19, 0.31]	.00	[-.00, .00]	R^2 = .079**
					95% CI[.04,.12]

A significant b-weight indicates the semi-partial correlation is also significant. b represents unstandardized regression weights. sr2 represents the semi-partial correlation squared. LL and UL indicate the lower and upper limits of a confidence interval, respectively.

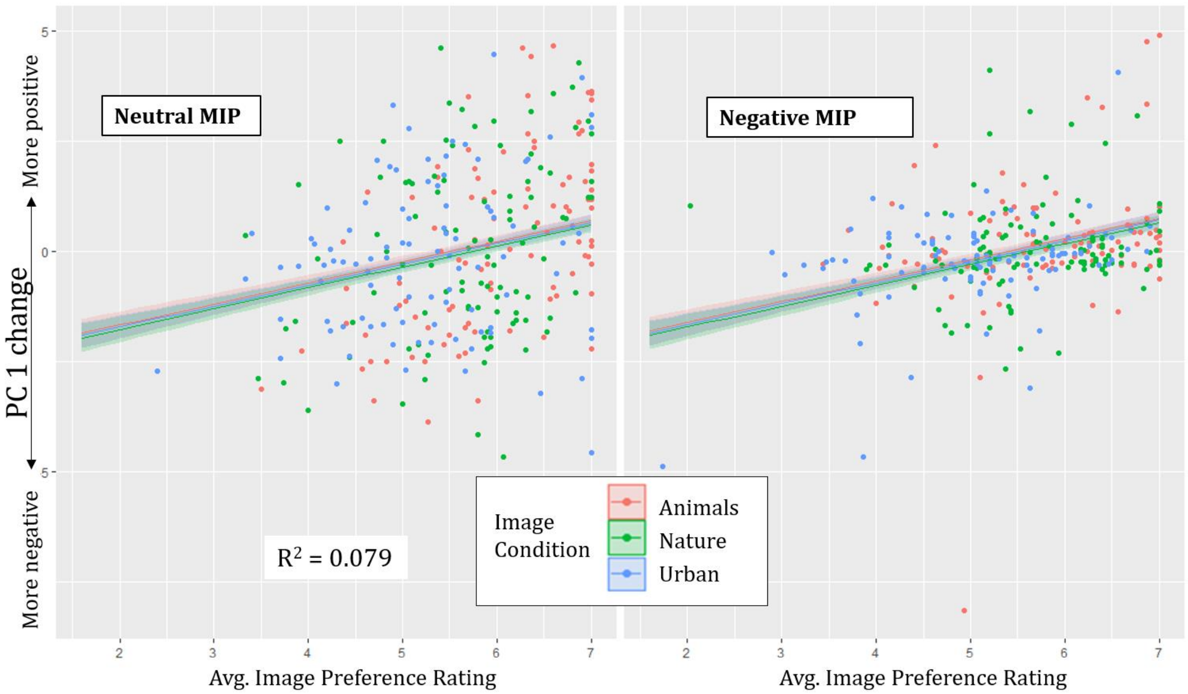


Figure S2.5. Regression Plots for PC 1 change by Mood Induction Condition
Regression plots predicting change in PC 1 (affective valence) by Image Condition + Average Image Rating + MIP Condition, broken up by for neutral (left) and negative (right) MIPs

Study 5 Pre-Registered Analysis

As specified in the OSF pre-registration (https://osf.io/u2e6n/?view_only=ddc4107c7af64b3e8783db56d03ec949), beauty ratings were compared in the 180 images corresponding to those in the High and Low Aesthetic value Nature and Urban conditions. A one-way ANOVA was conducted to compare the aesthetics (beauty) ratings from Study 5 (this study) across the two image sets which were equated on preference (affinity) from Study 1. This analysis was conducted using the function ‘aov’ in R. Results of the 1-way ANOVA comparing beauty ratings of the 90 urban and 90 nature preference-matched images did not yield a significant effect of environment on beauty rating, $F(1,178) = 0.01$, $p = 0.91$, $\eta_p^2 = 0.00$, 95% CI [-0.01, 0.02]. Distributions of average beauty ratings for images in each environment type are plotted in **Figure S2.6.**

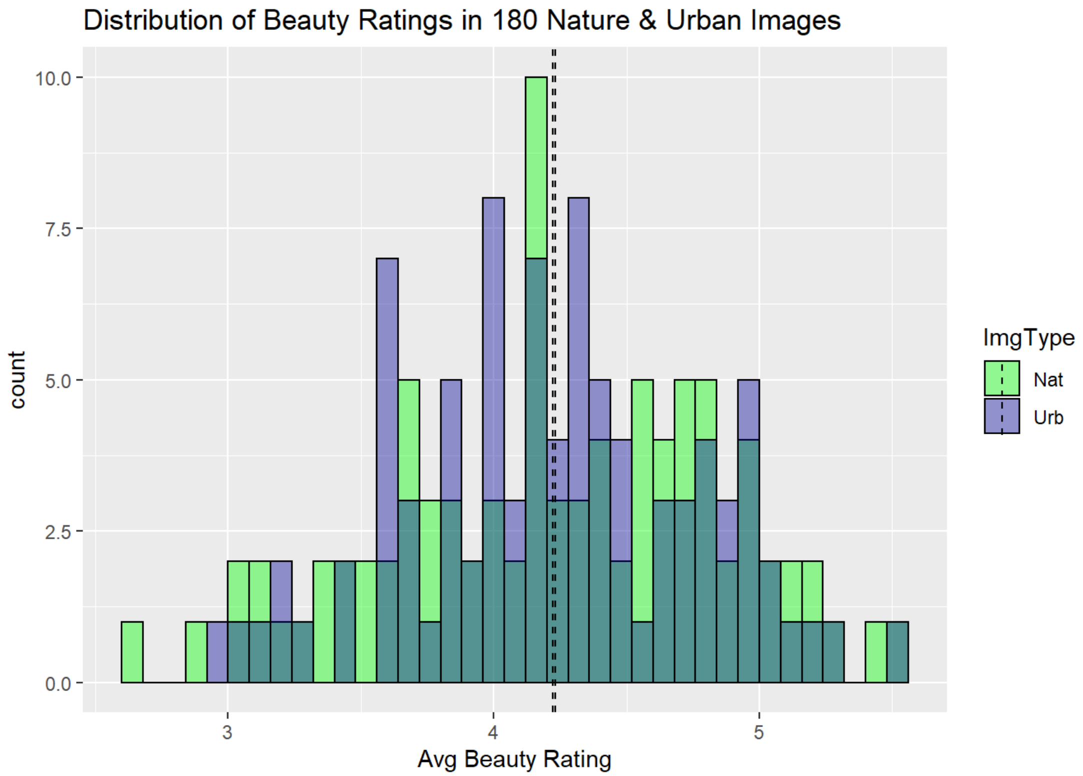


Figure S2.6. Histogram of Beauty Ratings for Preference-Equated Nature & Urban Images
Aesthetics (beauty) ratings from Study 5 for the 90 nature and 90 urban images with overall similar preference (affinity) ratings from Study 1.

Appendix D: Supplementary Analyses for Chapter 4

Behavioral Results: Main N-back performance, Outliers Removed (<60% acc)

Results of the repeated measures ANOVA examining accuracy as a function of n-back level in the main task yielded a significant effect of n-back level on accuracy ($F(2,92) = 57.1$, $p < 0.001$, $\eta_p{}^2 = 0.54$, 95% CI [0.41 0.64]. As expected, accuracy for the 1-back task ($M = 0.94$, $SD = 0.03$) was significantly better than accuracy for the 2-back task ($M = 0.86$, $SD = 0.1$, $t(46) = 5.87$, $p < 0.001$) and for the 3-back task ($M = 0.80$, $SD = 0.1$, $t(46) = 9.7$, $p < 0.001$). Accuracy for the 2-back task was also significantly higher than for the 3-back task ($t(46)$, 5.5, $p < 0.001$). [**Figure S4.1**]

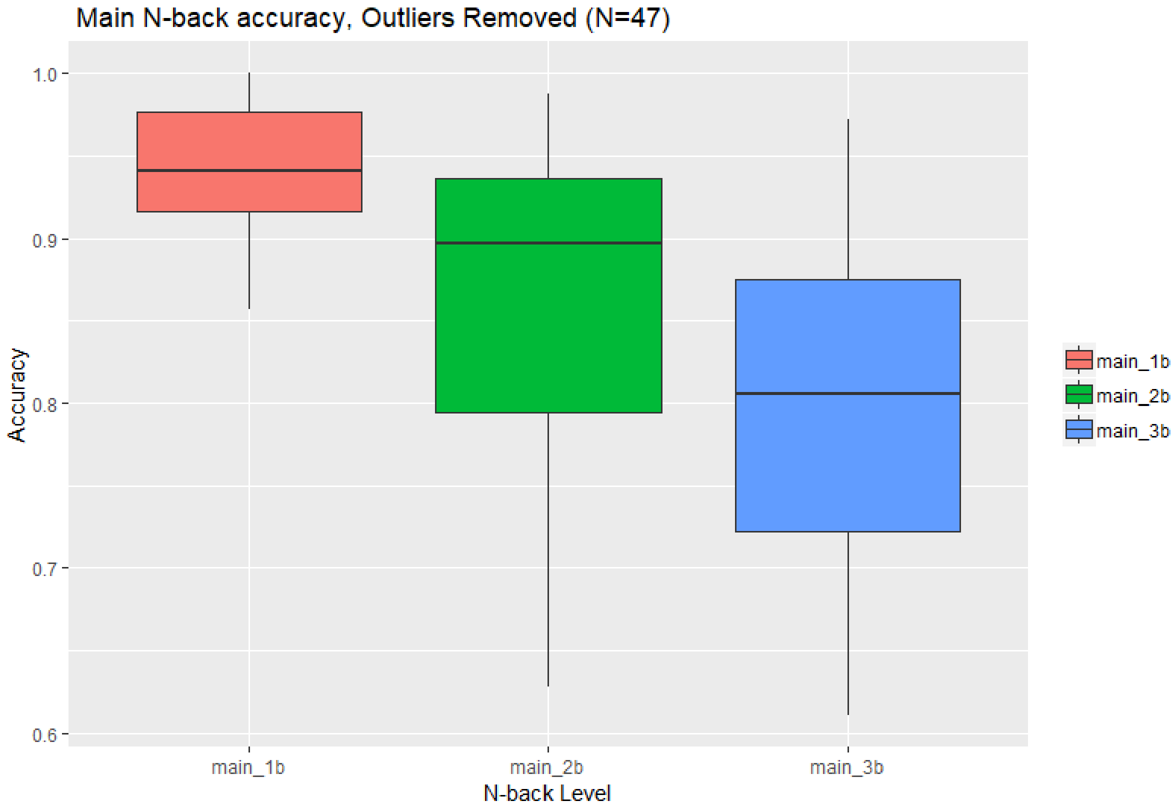


Figure S4.1. Average accuracy in Main N-back Round, outliers removed

Behavioral Results: Change in N-back Performance, Outliers Removed

To test for the effect of environment type on change in 3-back performance, a 2 (time: pre-video, post-video) x 2 (video: nature, urban) mixed-effects ANOVA was conducted with accuracy as the dependent variable. Results of this ANOVA yielded a significant main effect of time, $F(1,45) = 23.2$, $p < 0.001$, $\eta_p^2 = 0.34$, 95% CI [0.13 0.51], where participants performed significantly better in the post-video 3-back blocks ($M = 0.86$, $SD = 0.1$) relative to the pre-video 3-back blocks ($M = 0.80$, $SD = 0.1$, $t(46) = 4.9$, $p < 0.001$). No main effect of video or an interaction of video and time was found (all $p > 0.67$). When examined as a change score (post-video 3-back

— main 3-back, controlling for main 3-back performance), the results were unaffected (*p* = 0.86). [**Figure S4.2**]

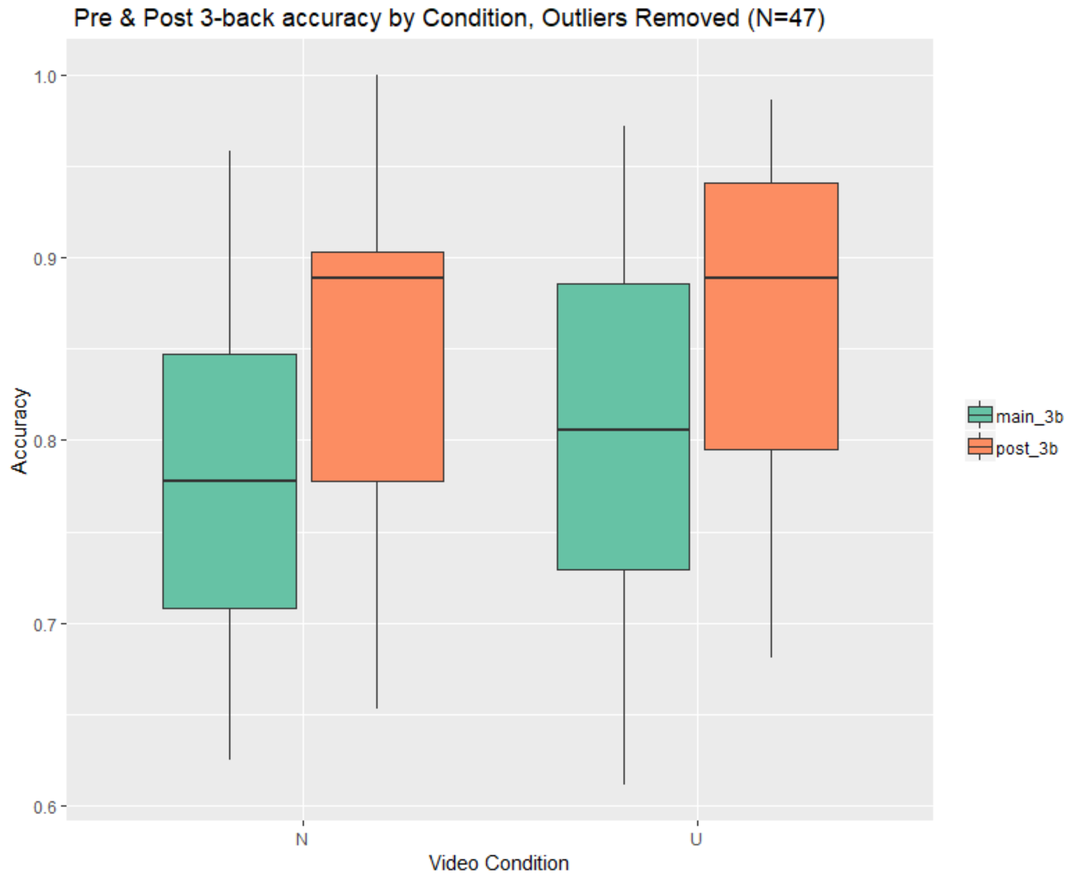


Figure S4.2. A Main 3-back (Pre) and Post-Video 3-back Performance by Video Condition, outliers removed