



ORIGINAL ARTICLE

Open Access



A pilot study examining the suitability of the mental arithmetic task and single-item measures of affective states to assess affective, physiological, and attention restoration at a wooden desk

Dean Lipovac^{1,2*} , Jure Žitnik^{1,3} and Michael D. Burnard^{1,2}

Abstract

People seem to function and feel better in indoor natural environments, including spaces furnished with wood. When restorative effects of indoor spaces are not detected, suboptimal methodological approaches may be responsible, including stress-inducing activities and measures of affective states and cognitive performance. Our primary objectives were to test (1) whether the Mental Arithmetic Task (MAT) can reliably induce stress and measure cognitive performance, and (2) whether two single-item measures of pleasure and arousal can detect changes in affective states in restoration research. Our secondary objective was to examine whether stress recovery and cognitive performance differ between indoor settings furnished with or without wood. Twenty-two participants, allocated to a space furnished with either a wooden or a white desktop, completed MAT twice, while their electrodermal and cardiovascular activity and affective states were monitored. Participants on average responded to MAT with increased subjective arousal but unchanged subjective pleasure, and with increased physiological arousal on some but not all parameters, suggesting that MAT was effortful but not necessarily stressful. Scores on MAT improved at the 2nd administration, suggesting that MAT did not induce cognitive fatigue at the 1st administration and that its role as a cognitive task in restoration research may be limited. The items assessing affective states performed well. The measured outcomes did not differ between the wooden and non-wooden setting, suggesting that substantial restorative effects of a wooden desktop are unlikely, and that higher wood coverage is needed to increase the chances of observing restorative effects.

Keywords: Restorative environments, Indoor nature, Human stress, Cognitive performance

Introduction

A large body of evidence suggests that people feel and perform better after spending time in natural, restorative environments [1–4]. These observations are usually explained with the stress reduction theory (SRT) [5], attention restoration theory (ART) [6, 7], or both.

SRT claims that the positive outcomes following contact with nature result from the connections humans have developed with nature during the evolution of the species. According to SRT, pleasant, non-threatening natural environments elicit pleasant feelings, hold interest of people and reduce stressful thoughts, and decrease physiological arousal if the initial level is high [5]. ART focuses on the ability of nature to hold human interest: it claims that nature possess qualities that attract effortless attention of people, allowing their directed (effortful) attention to rest and replenish, as fatigued directed

*Correspondence: dean.lipovac@innorenew.eu

¹ InnoRenew CoE, Livade 6, 6310 Izola, Slovenia

Full list of author information is available at the end of the article

attention supposedly leads to stress and decreased cognitive performance [6].

In line with predictions of SRT and ART, several recent reviews observed that participants who spend time in a natural environment generally report improved affective states, exhibit lower physiological arousal [1–4], and perform better on cognitive tasks [8, 9]. However, not all studies observe positive effects following exposure to nature. In those cases, it is challenging to discern whether the tested environment does not lead to restoration or whether the restorative effects do exist but are not observed due to the particular study design and outcomes.

This is especially problematic in studies that test for presumably smaller effects of exposure to nature in indoor spaces, where nature is present only indirectly or in smaller quantities, such as in spaces furnished with natural materials, like wood [10–12]. Indeed, while some studies observed promising effects of wooden indoor environments on occupants [13, 14], others detected no positive effects [15], or reported inconclusive results [16]. Future research should clarify whether (and in what contexts) wood impacts people positively, as bringing nature indoors can be a valuable intervention [17] because most people spend most of their time indoors [18].

Future studies examining effects of indoor nature exposure would benefit from clearer guidelines that would minimize the possibility to miss differences in restorative effects between tested environments due to the study protocol and outcomes, and thus maximize the chances to distinguish restorative from non-restorative environments. A typical study in the field measures some combination of affective states, physiological arousal, and cognitive performance before and after exposure to environments [17]. Researchers must select specific affective, physiological, and cognitive measures from numerous options and then decide when in the study protocol to administer those measures. If the selected measures and the timings of their implementation are inappropriate, results can be misleading. Researchers may incorrectly conclude that tested environments do not differ in terms of restorativeness, when, in fact, the particular study protocol and outcomes are responsible for the lack of observed differences.

One issue can arise from the selection of tools that capture affective states in restoration research. Some assessment tools seem to show higher effect sizes than others, in part presumably because natural environments likely elicit specific affective states that different tools capture to different extents [1]. Currently, however, we do not know which affective states are most reliably influenced by the natural environment [1]. Even tools such as Positive and Negative Affect Schedule (PANAS) [19], which

show relatively high effect sizes [1], could be far from optimal in restoration research, as they capture specific affective states (such as “guilty” and “proud”) that may not be reliably influenced by exposure to nature. PANAS and other commonly used measures also tend to be long (e.g., PANAS contains 20 items), which makes them less suitable for frequent administration and thus more likely to miss changes of affective states in longer exposures to (restorative) environments. Assessment tools based on dimensional approach—an approach describing all affective states on a set of selected dimensions (e.g., pleasure and arousal) [20, 21]—have been underused in restoration research [1]. These assessment tools are recommended in conditions that often characterize restoration studies: (1) when based on the current theory it cannot be anticipated how manipulations will impact affect [22], and (2) when subjects are required to report affective states on several occasions of a study [21].

Another opportunity for misleading results occurs when viewing lower physiological arousal as a positive outcome without additional information [12]. According to SRT, pleasant natural environments can either not influence, decrease, or increase arousal, depending on the initial arousal level [5]. In addition, physiological arousal can reflect states other than stress, including digestion, effort, and attention [23]; and, importantly, both pleasant and unpleasant states can be reflected in either higher or lower physiological arousal [24, 25]; for example, high arousal can indicate vigor [26] and low arousal can signal fatigue [27]. These observations suggest that measures of physiological arousal should be corroborated by measures of affective states, and that a stress-inducing activity should be included so the higher physiological arousal can be more easily attributed to an unpleasant state, such as fear, instead of a pleasant state, such as excitement [28]. Despite the importance of assessing affective states and inducing stress, a recent review of 43 studies reported that only about two-thirds of the studies in the field used a self-report measure and only one in ten studies experimentally induced stress in participants [2].

Including a stressful activity is important, but not all stressors are equally effective. Some approaches, such as exposing people to noise or inducing specific emotions with videos, do not lead to reliable increases in stress (as reflected in cortisol—a commonly used biomarker of stress), while the greatest increases in stress occur with the combination of a cognitive task and public speaking [29]. This combination is present in the commonly used Trier Social Stress Test (TSST) [30, 31], which induces stress relatively reliably even with the variations in the TSST protocol [32]. The downside of TSST is its requirement of three individuals acting as an evaluative audience (i.e., “judges”) and its duration: not counting the

acclimatation and recovery periods, TSST typically lasts 20 min [30, 31]. Despite the advantages of TSST, at least some restoration studies could benefit from a shorter yet reliable stress-induction method that is simpler to implement.

On top of challenges related to assessing affective states and inducing stress, studies can encounter issues when assessing attention restoration—how people perform cognitively after spending time in nature [8, 9, 12, 33, 34]. Results can depend on the specific cognitive function that is measured and on how fatigued participants are: restoration may be more likely to occur in domains of cognitive flexibility and working memory [8], and in participants who are fatigued, either because of a cognitively fatiguing task within an experiment or because of fatiguing day-to-day occurrences, such as attending lectures [8]. Experimentally inducing fatigue with cognitive tasks can be problematic as it can be lengthy—up to 40 min in studies identified by Stevenson et al. [8], while uncontrolled fatiguing day-to-day occurrences are less likely to lead to uniform levels of fatigue among study participants (e.g., during lectures, some students may exert more mental effort and get more fatigued than others). An approach that could sidestep these limitations is increasing cognitive fatigue by inducing stress—according to ART, attentional resources can decline due to stress and not only task demand [6]. This approach is currently used in few studies [8], although it might provide a briefer standardized method to increase participants' need for attention restoration. This opens an interesting possibility of using the mental arithmetic task (MAT)—a part of the TSST stress-inducing protocol—as a stressor as well as a fatiguing cognitive task and test of cognitive performance. MAT involves subtracting the number 13 or 17 from a 4-digit number and reporting answers aloud [30]. As a stressor, MAT can be effective because it involves a social-evaluative threat—task performance could be negatively judged by others [29]; and as a cognitive task, MAT can be suitable because it taps the working memory domain [35], which can be influenced by natural environments [8].

In summary, current research shows that people benefit from spending time in natural environments, but the effects are less clear when people are exposed to some indoor elements of nature, such as wood, possibly due to suboptimal methodological approaches. It is unclear (1) whether the changes of affective states in restorative environments are detected by the tools based on dimensional approach (e.g., pleasure and arousal dimensions), and (2) whether a cognitive task that acts as a stressor (i.e., MAT) is a viable inducer of stress and cognitive fatigue and a viable measure of cognitive performance in restorative environments.

Objectives

Our study primarily aimed to test the suitability of a selected task and outcomes for restoration research, specifically in the context of people's exposure to indoor wood. We aimed to test whether MAT reliably induces stress, as reflected in cardiovascular and electrodermal activity, and affective states, as captured by two items assessing pleasure and arousal (based on the circumplex model of affect) [20]. We were additionally interested in whether MAT can be a viable cognitive task in restoration research. The secondary aim of our study was to examine whether the inspected physiological, affective, and cognitive outcomes differ between wooden and non-wooden indoor settings.

Methods

Participants

A convenience sample of 22 subjects (18 females) participated in the study, with 19 subjects between the ages of 18 and 34, and three subjects between the ages of 35 and 54. Subjects were eligible to participate in the study if they had no health issues or characteristics that would have interfered with the study tasks (e.g., very poor computer skills). Before the experiment, subjects signed an informed consent form that informed them about the study purpose and procedure, rights of participants, and data management practices. The study protocol was approved by the National Medical Ethics Committee of Slovenia (No. 0120–298/2020/3) and the research was carried out in compliance with the Oviedo convention. As a compensation for participating in the study, subjects received a report of their results (in reference to aggregated results of other participants).

Test setting

The experiment was conducted in spaces of University of Primorska in Koper, Slovenia. The test setting included a preparation desk with a smaller top surface (100 cm × 70 cm) and a test desk with a larger top surface (200 cm × 90 cm). The two desks were placed at the opposite sides of the space. The top surface of the smaller desk was covered with beige melamine, while the top surface of the larger desk was made of oak (*Quercus robur*) veneer—light colored wood with darker streaks and with a clear lacquer finish applied by the vendor. The oak veneer was exposed in the experimental condition and covered with a white tablecloth in the control condition (Fig. 1). Windows in the test setting were covered with white drapes to prevent participants from viewing the outdoor environment. The



Fig. 1 The test desk in the control (left) and experimental condition (right)

experiment took place in August of 2020, and the testing was conducted throughout the entire day.

Measures

Affective states

Affective states were assessed with two single-item measures capturing the states of pleasure and arousal [21]. The scales, based on the circumplex model of affect [20], capture the broad state of core affect—simplest consciously accessible feelings, rather than specific emotions or longer lasting moods [36]. Despite their brevity, the scales have proven to be reliable and valid [21, 37]. The two administered items asked participants: “How pleasant/activated do you feel at this moment?”. Participants responded on a 9-point rating scale (1 = especially unpleasant/inactivated, 5 = neutral, 9 = especially pleasant/activated).

Cognitive performance

Cognitive performance was assessed with MAT—a part of the TSST [30]. Participants were instructed to (mentally) sequentially subtract the number 13 from a 4-digit number (1022 and 1059 in the 1st and 2nd task administration, respectively) as fast and as accurately as possible and report their results verbally, while the researcher was monitoring the correctness of their responses. If participants made a subtraction error, they were instructed to start subtracting again from the initial 4-digit number. The subtraction period lasted for 5 min at each task administration.

Physiological arousal

Physiological arousal was examined with measures of cardiovascular activity, which reflects the activity of the heart and blood vessels, and electrodermal activity, which reflects the activity of the sweat glands in the skin.

Different measures correspond to different branches of the autonomic nervous system. Electrodermal activity predominantly reflects the sympathetic branch [38], heart rate corresponds to both sympathetic and parasympathetic branches, and heart rate variability largely relates to the parasympathetic branch [39]. As indicators of stress, measures of cardiovascular and electrodermal activity have been frequently used in psychophysiological research in general [38–40] and restoration research in particular [17].

Participants were equipped with wireless sensors that captured cardiovascular and electrodermal activity. Cardiovascular activity was monitored with a chest strap (Equivital Life Monitor EQ02; Hidalgo, Cambridge, UK), and electrodermal activity was assessed with a galvanic skin response sensor (EQ-ACC-34; Hidalgo, Cambridge, UK), which was attached to pre-gelled Ag/AgCl electrodes placed on two fingers (index and middle finger) of participants’ left hand. Cardiovascular activity was parametrized as heart rate (beats per minute) and heart rate variability. The root mean square of successive beat-to-beat interval differences (RMSSD) was used as a representative measure of heart rate variability [41]. Electrodermal activity was parametrized as skin conductance level (SCL; i.e., tonic level of electrical conductivity of skin), percentage of skin conductance responses (SCR; i.e., brief increase in conductance following physiologically arousing external or internal stimuli), and SCR amplitude (i.e., the extent of the increase in conductance at the SCR [38]). The physiological arousal data were captured and processed in LabChart 8.1 [42]. Electrodermal activity data were additionally processed with the Python package NeuroKit2 [43]. Electrodermal activity data from one subject were excluded from the analysis, due to unusually high SCL and odd SCL patterns, suggesting a systematic error in the measurement process.

Experimental procedure

Participants were instructed ahead of the study to abstain from caffeinated beverages on the day of the testing, as these might interfere with measurements of physiological arousal [44]. Before the experiment, participants received an overview of the upcoming study protocol and instructions on completing the study tasks. They were then equipped with physiological activity sensors and provided with the opportunity to ask questions related to the study.

Participants were guided through the experimental protocol by a web platform (developed with the R package psychTestR [45]), which delivered instructions, captured self-reported data, and provided timers. The experimental protocol (Fig. 2) started with the baseline period (Baseline), to ensure participants acclimatized to the test testing and that the baseline values of physiological activity were captured. Participants then responded to a measure of affective states, completed MAT (Task (1)), and responded to the measure of affective states again immediately after. Afterwards, they started with the recovery period (Recovery): they relocated to a larger desk at the opposite side of the room, which had its wooden surface either exposed (experimental condition) or covered with a white tablecloth (control condition), where they rested for 10 min (half of the participants were randomly assigned to one of the two conditions). Note that the duration of the resting period should suffice for cardiovascular [46] and electrodermal activity [47, 48] to return to baseline levels after stress induction. Finally, participants responded to a measure of affective states for the third (and final) time and completed MAT for the 2nd (and final) time (Task (2)).

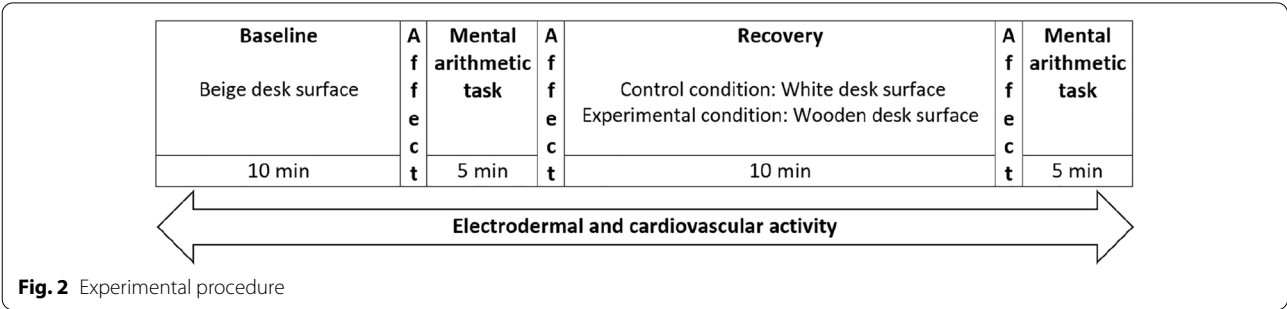
Statistical analysis

The data were processed and analyzed with R 4.0.2 [49] and Python 3.9.2 [50] using RStudio 1.4.1106 [51] with the packages janitor [52], NeuroKit2 [43], broom.mixed [53], rstatix [54], reticulate [55], lme4 [56], lmerTest [57], emmeans [58], DHARMA [59], flextable [60], and the collection of packages tidyverse [61]. Summary statistics were reported as means (M) and confidence

intervals (CI), and visualized as boxplots. In the boxplots, the box represents the interquartile range, which spans from the first (lower) quartile at the bottom hinge to the third (upper) quartile at the top hinge. The thicker line passing through the box represents the median (second quartile). The whiskers extend from the hinges to the largest (for the upper whisker) or smallest value (for the lower whisker) that is no further from the hinge than $1.5 \times$ interquartile range—distance between the first and third quartiles. The overlaid dots represent raw data points.

Our data would commonly be analyzed with a mixed analysis of variance (ANOVA), where the desktop condition (i.e., wooden or white desk) would be treated as a between-subject factor and the study phase (i.e., baseline, task, recovery) as a within-subject factor. Instead of using the mixed ANOVA, we based our analysis on (generalized) linear mixed models, which are becoming increasingly more widespread and recommended approach to analyze within-subject data due to their flexibility and robustness [62].

We typically fitted a linear mixed model, where the residual error term is expected to follow a normal distribution. In one instance, we fitted a binomial mixed model, which can handle dependent variables whose residual error does not follow a normal distribution (i.e., a binary dependent variable whose error distribution is binomial) [63]. The (generalized) linear mixed models were fitted with the R packages lme4 [56] and lmerTest [57]. In all models, subjects were treated as random effects and desktop conditions, study phase, and/or task administration were treated as fixed effects. All models tested for interactions between fixed effects. Variables representing electrodermal and cardiovascular activity were included in the model as dependent variables after the mean values were calculated for each participant at each study phase. At Baseline and Recovery, only the 5 min of the lowest physiological activity (according to skin conductance values) for each period were taken for further analysis, to minimize the presence of physiological arousal resulting from the period before the experiment and from the anticipation of the upcoming task



during the experiment. The variables representing affective states and cognitive performance were included as dependent variables in their raw form.

If statistically significant main effects or interaction effects were detected, post hoc comparisons were conducted with the R package *emmeans* [58], where *p* values were adjusted with the Tukey method and estimated marginal means (*EMM*) were reported. In one instance of the linear mixed model, the dependent variable (i.e., SCR peaks) underwent square root transformation to improve model fit; however, reported *EMMs* were back-transformed and presented in the original unit of the dependent variable, while the corresponding contrasts (differences between the values of two dependent variables) generally cannot be back-transformed and were reported as differences between two square roots. Model diagnostics were conducted with the R package *DHARMA* [59], which uses a simulation-based approach to analyze residuals of (generalized) linear mixed models. None of the reported models exhibited issues with fit to the data.

In some cases, we examined the data in more detail after uncovering atypical response patterns in some participants (i.e., atypical responses on the affective state of pleasure). Here, we split the participants into two groups: if the participant's score on the affective state of pleasure increased or stayed the same from Baseline to Task (1) and decreased or stayed the same from Task (1) to Recovery, the participant was classified as an atypical responder; otherwise, the participant was classified as a typical responder. The results (i.e., physiological activity and cognitive task scores) of these two groups were compared with Wilcoxon tests (the Wilcoxon signed-rank

test was used as a paired difference test and the Wilcoxon rank-sum test was used as an unpaired two-sample test). By splitting our sample, we created two smaller groups of participants (with unequal sizes), which lowers the statistical power of significance tests [64]. For this reason, the *p* values were not adjusted for multiple comparisons, to decrease the possibility of the Type II error (i.e., false negative).

Results

Affective states

Participants on average reported values around the middle of the scale for the affective states of arousal ($M=4.53$, 95% CI [4.09, 4.97]) and pleasure ($M=5.15$, 95% CI [4.70, 5.60]). The results of the linear mixed model showed that the arousal scores significantly changed throughout the study phases, while the pleasure scores did not (Table 1, Fig. 3). The arousal and pleasure scores did not differ between desktop conditions, and there were no interaction effects between desktop conditions and study phases. Post hoc comparisons showed that arousal scores were higher at Task (1) ($EMM=5.59$, 95% CI [4.87, 6.31]) than at Baseline ($EMM=3.73$, 95% CI [3.01, 4.45]; Task (1)—Baseline = 1.86, 95% CI [1.08, 2.64], $p<0.001$) and Recovery ($EMM=4.27$, 95% CI [3.55, 4.99]; Task (1)—Recovery = 1.32, 95% CI [0.54, 2.10], $p<0.001$), while the scores did not significantly differ between Baseline and Recovery (Baseline—Recovery = -0.55 , 95% CI [-1.33 , 0.24], $p=0.217$).

It should be noted that the pleasure scores, even though they have not (on average) significantly changed in any one direction between study phases, still varied within participants (Fig. 4)—few participants reported

Table 1 Results of the linear mixed models with affective states as dependent variables

Outcome	Predictor	Estimate	SE	95% CI	df	<i>t</i>	<i>p</i>
Arousal	Intercept	3.82	0.50	2.83 to 4.80	35.17	7.59	<0.001
	Condition: Wood	−0.18	0.71	−1.58 to 1.21	35.17	−0.26	0.800
	Phase: Task (1)	2.00	0.45	1.11 to 2.89	40.00	4.41	<0.001
	Phase: Recovery	0.73	0.45	−0.16 to 1.62	40.00	1.60	0.117
	Wood * Task (1)	−0.27	0.64	−1.53 to 0.98	40.00	−0.43	0.673
	Wood * Recovery	−0.36	0.64	−1.62 to 0.89	40.00	−0.57	0.574
Pleasure	Intercept	5.64	0.56	4.55 to 6.73	44.06	10.12	<0.001
	Condition: Wood	0.18	0.79	−1.36 to 1.72	44.06	0.23	0.818
	Phase: Task (1)	−0.91	0.60	−2.08 to 0.26	40.00	−1.52	0.136
	Phase: Recovery	−0.91	0.60	−2.08 to 0.26	40.00	−1.52	0.136
	Wood * Task (1)	0.09	0.84	−1.56 to 1.75	40.00	0.11	0.915
	Wood * Recovery	0.09	0.84	−1.56 to 1.75	40.00	0.11	0.915

Significant predictors ($p<0.05$) are shown in bold

SE: standard error; CI: confidence interval; df: degrees of freedom; *t*: test statistic *t*; *p*: *p*-value

*Interaction between predictors

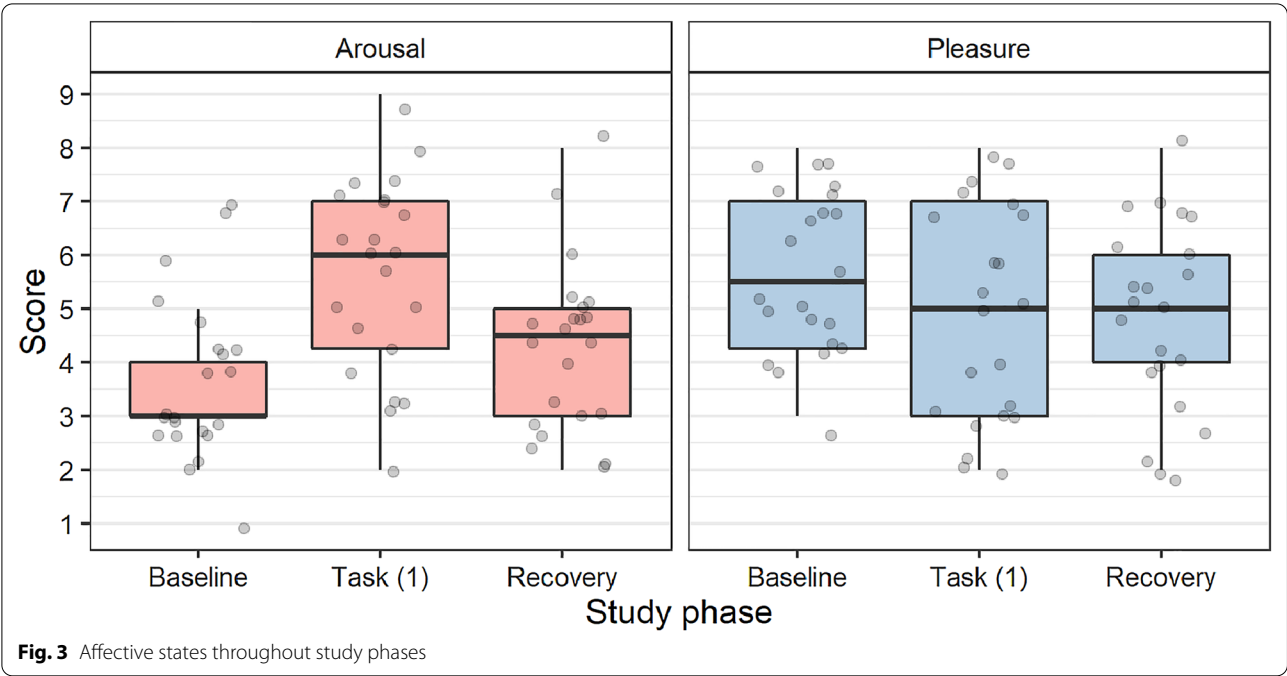


Fig. 3 Affective states throughout study phases

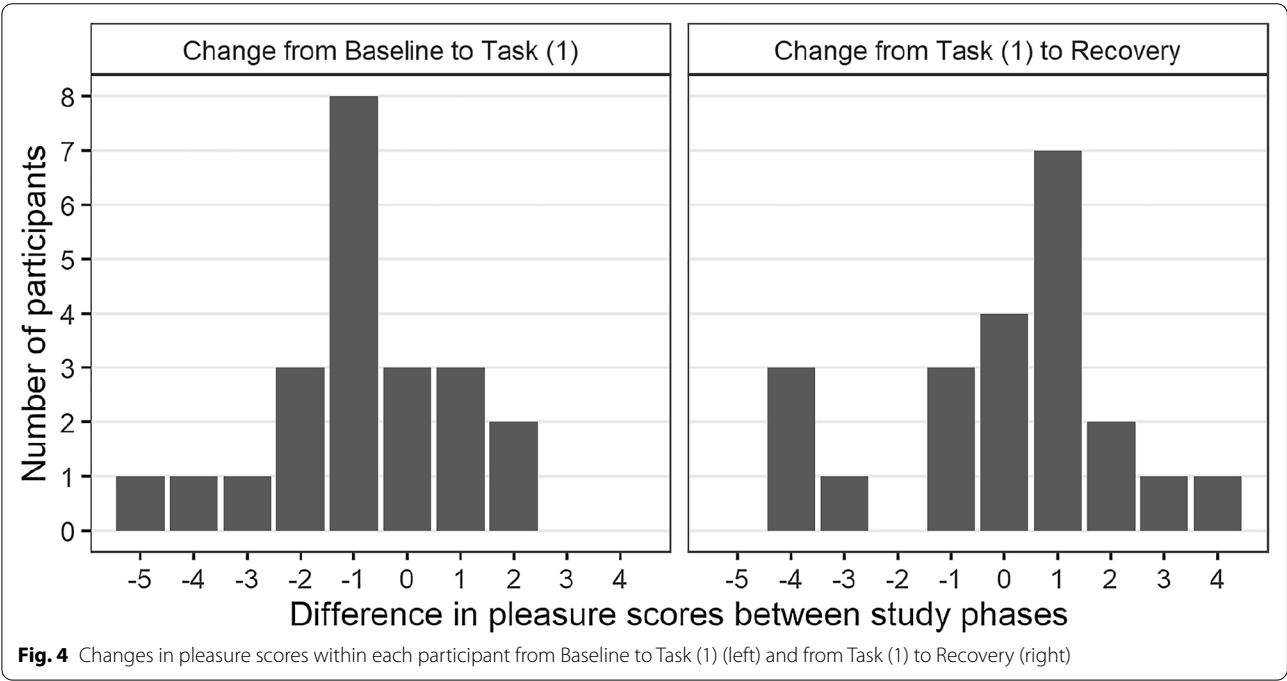


Fig. 4 Changes in pleasure scores within each participant from Baseline to Task (1) (left) and from Task (1) to Recovery (right)

no change in their pleasure scores between study phases, while many reported either decreases or increases in pleasure both from Baseline to Task (1) and from Task (1) to Recovery.

Further examination identified six participants with atypical responses, for whom pleasure seems to have increased or stayed the same from Baseline to Task

(1) and decreased or stayed the same from Task (1) to Recovery, in contrast with 16 participants with typical responses, for whom pleasure appears to have decreased from Baseline to Task (1) and increased from Task (1) to Recovery (Fig. 5, Additional file 1: Table S1).

Atypical and typical responders had similar scores on subjective arousal at Task (1) and Recovery (Additional

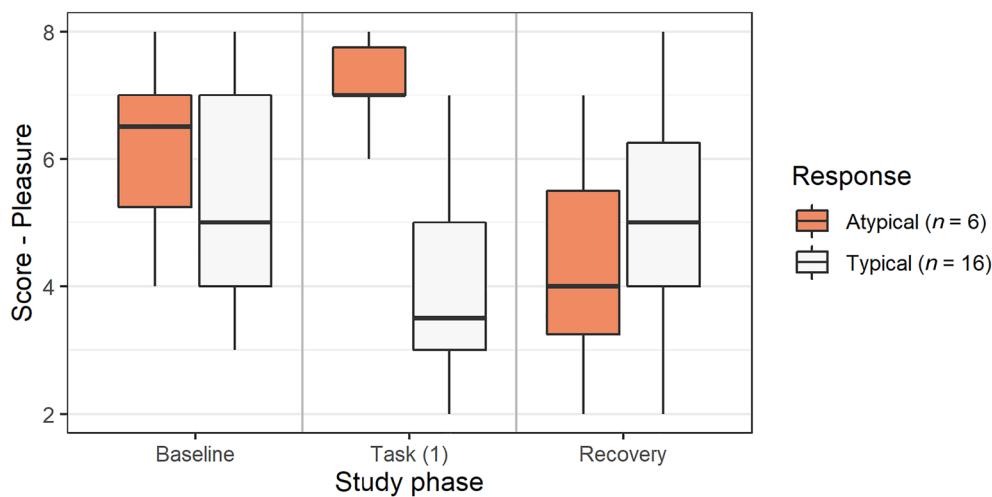


Fig. 5 Pleasure scores across study phases for participants with atypical and typical responses

file 1: Figure S1, Tables S2, 3), but different scores at Baseline, where atypical responders had somewhat lower scores compared to typical responders (difference: -1.00 , 95% CI $[-3.00, 0.00]$, $p = 0.046$).

Physiological arousal

Electrodermal activity

Throughout all study phases, the mean of (means of) exhibited values was $6.35 \mu\text{S}$ (95% CI $[5.46, 7.24]$) for SCL, 0.51% (95% CI $[0.31, 0.71]$) for SCR, and $0.23 \mu\text{S}$ (95% CI $[0.17, 0.29]$) for SCR amplitude. The linear mixed models showed that the SCL and SCRs (but not SCR amplitude) changed throughout study phases, while there were no main effects of desktop condition or interactions between desktop condition and study phases (Table 2, Fig. 6).

Post hoc comparisons showed that SCL scores significantly increased from Baseline ($EMM = 4.36 \mu\text{S}$, 95% CI $[2.88, 5.84]$) to Task (1) ($EMM = 8.05 \mu\text{S}$, 95% CI $[6.57, 9.53]$; Task (1)–Baseline = $3.69 \mu\text{S}$, 95% CI $[3.00, 4.38]$, $p < 0.001$) and then decreased from Task (1) to Recovery ($EMM = 6.57 \mu\text{S}$, 95% CI $[5.09, 8.05]$; Task (1)–Recovery = $1.48 \mu\text{S}$, 95% CI $[0.79, 2.18]$, $p < 0.001$), but remained higher at Recovery than they were at Baseline (Recovery–Baseline = $2.21 \mu\text{S}$, 95% CI $[1.52, 2.90]$, $p < 0.001$). A somewhat similar trend was seen in SCRs, which increased from Baseline ($EMM = 0.13\%$, 95% CI $[0.04, 0.28]$) to Task (1) ($EMM = 0.76\%$, 95% CI $[0.49, 1.10]$; Task (1)–Baseline (difference of square roots) = 0.51 , 95% CI $[0.35, 0.68]$, $p < 0.001$), and decreased from Task (1) to Recovery ($EMM = 0.24\%$, 95% CI $[0.10, 0.43]$; Task (1)–Recovery (difference of square roots) = 0.39 , 95% CI $[0.22, 0.55]$, $p < 0.001$), while they did not significantly

differ between Baseline and Recovery (Baseline–Recovery (difference of square roots) = -0.13 , 95% CI $[-0.29, 0.04]$, $p = 0.155$).

Cardiovascular activity

Throughout all study phases, the mean of (means of) exhibited values were 79.95 beats per minute for heart rate (95% CI $[75.47, 84.43]$) and 36.41 ms for heart rate variability—RMSSD (95% CI $[31.59, 41.23]$). The linear mixed models showed that heart rate (but not heart rate variability) changed throughout study phases (Table 3, Fig. 7). There were no main effects of desktop conditions or interaction effects between desktop conditions and study phases. Post hoc comparisons for heart rate showed a similar trend as electrodermal activity results: heart rate values (beats per minute) were comparatively low at Baseline ($EMM = 74.24$, 95% CI $[67.13, 81.35]$), increased from Baseline to Task (1) ($EMM = 91.92$, 95% CI $[84.80, 99.03]$; Task (1)–Baseline = 17.68 , 95% CI $[10.55, 24.81]$, $p < 0.001$), and decreased from Task (1) to Recovery ($EMM = 73.69$, 95% CI $[66.58, 80.80]$; Task (1)–Recovery = 18.23 , 95% CI $[11.10, 25.36]$, $p < 0.001$), with no significant differences between Baseline and Recovery (Baseline–Recovery = 0.55 , 95% CI $[-6.58, 7.68]$, $p = 0.852$).

Further analysis suggested that participants with atypical pleasure scores (see “Affective states” section) had similar patterns of electrodermal activity but different patterns of cardiovascular activity compared to participants with typical pleasure scores (Fig. 8, Additional file 1: Tables S4, 5). The atypical responders had a relatively stable heart rate across the study phases, while the heart rate of the typical responders increased

Table 2 Results of the linear mixed models with electrodermal activity parameters as dependent variables

Outcome	Predictor	Estimate	SE	95% CI	df	t	p
SCL [μ S]	Intercept	4.66	0.99	2.72 to 6.59	22.2	4.72	< 0.001
	Condition: Wood	− 0.60	1.43	− 3.40 to 2.20	22.2	− 0.42	0.680
	Phase: Recovery	2.60	0.47	1.67 to 3.52	38.0	5.51	< 0.001
	Phase: Task (1)	3.85	0.47	2.93 to 4.78	38.0	8.17	< 0.001
	Wood * Recovery	− 0.78	0.68	− 2.12 to 0.56	38.0	− 1.14	0.262
	Wood * Task (1)	− 0.32	0.68	− 1.66 to 1.02	38.0	− 0.47	0.638
SCR [%] (square root)	Intercept	0.33	0.12	0.10 to 0.55	29.55	2.82	0.008
	Condition: Wood	0.06	0.17	− 0.26 to 0.39	29.55	0.38	0.704
	Phase: Recovery	0.21	0.09	0.03 to 0.40	38.00	2.32	0.026
	Phase: Task (1)	0.59	0.09	0.41 to 0.77	38.00	6.37	< 0.001
	Wood * Recovery	− 0.18	0.13	− 0.44 to 0.09	38.00	− 1.31	0.197
	Wood * Task (1)	− 0.15	0.13	− 0.41 to 0.11	38.00	− 1.12	0.270
SCR amplitude [μ S]	Intercept	0.19	0.07	0.06 to 0.33	39.97	2.79	0.008
	Condition: Wood	0.09	0.10	− 0.11 to 0.28	39.97	0.87	0.388
	Phase: Recovery	0.00	0.07	− 0.14 to 0.14	38.00	0.03	0.978
	Phase: Task (1)	0.07	0.07	− 0.07 to 0.21	38.00	0.98	0.332
	Wood * Recovery	− 0.09	0.10	− 0.3 to 0.11	38.00	− 0.90	0.374
	Wood * Task (1)	− 0.06	0.10	− 0.27 to 0.14	38.00	− 0.62	0.540

Significant predictors ($p < 0.05$) are shown in bold

SE: standard error; CI: confidence interval; df: degrees of freedom; t: test statistic t; p: p-value

*Interaction between predictors

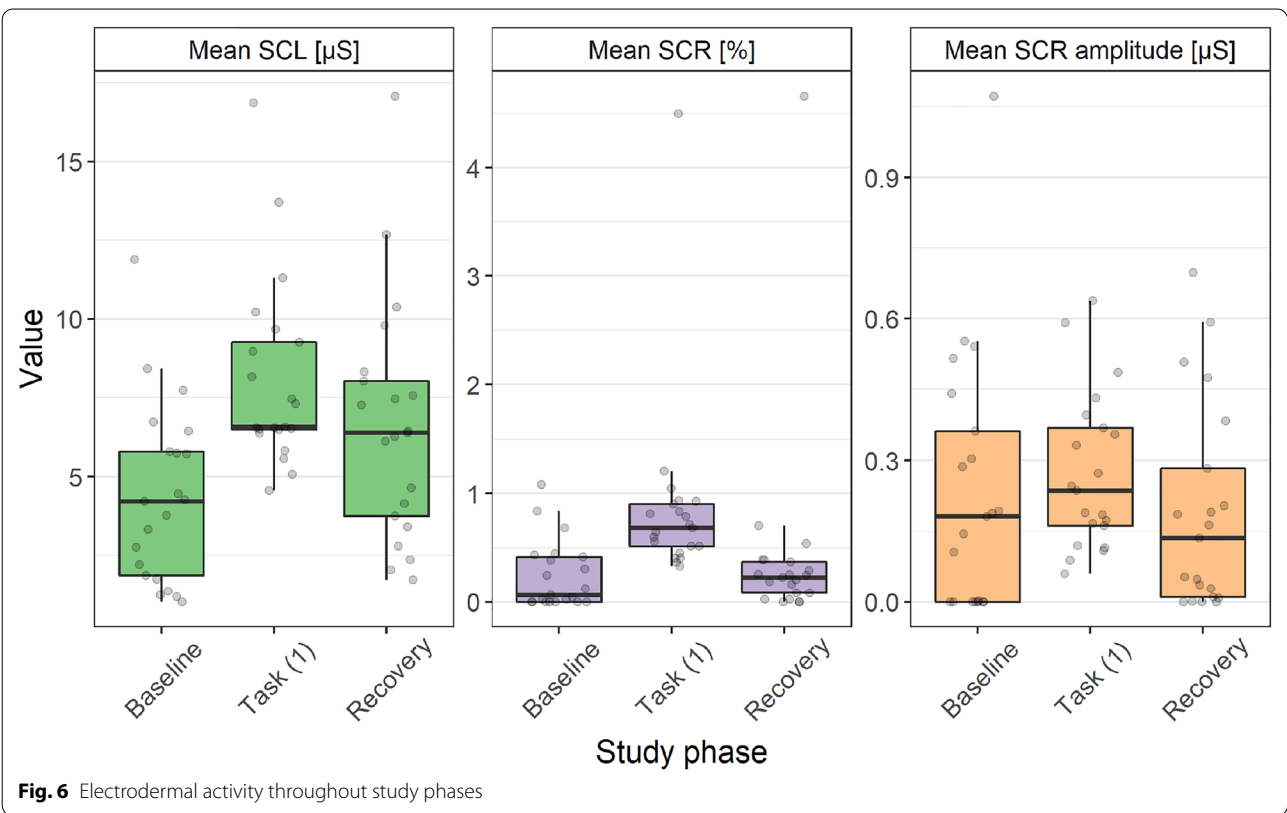


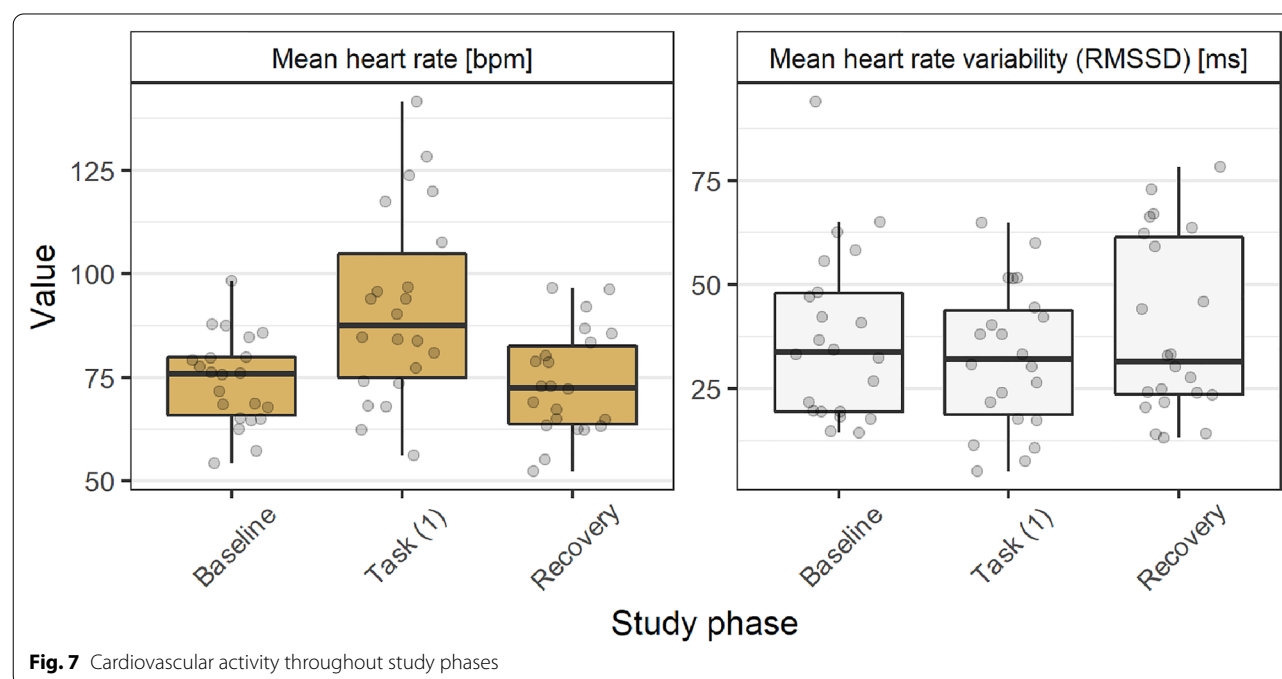
Table 3 Results of the linear mixed models with heart rate and heart rate variability (RMSSD) as dependent variables

Outcome	Predictor	Estimate	SE	95% CI	df	t	p
Heart rate [beats per minute]	Intercept	71.51	4.94	61.83 to 81.20	32.57	14.47	< 0.001
	Condition: Wood	5.45	6.99	− 8.25 to 19.15	32.57	0.78	0.441
	Phase: Recovery	16.40	4.14	8.28 to 24.52	40.00	3.96	< 0.001
	Phase: Task (1)	− 0.75	4.14	− 8.87 to 7.36	40.00	− 0.18	0.856
	Wood * Recovery	2.55	5.86	− 8.93 to 14.03	40.00	0.44	0.666
	Wood * Task (1)	0.41	5.86	− 11.07 to 11.89	40.00	0.07	0.945
RMSSD [ms]	Intercept	43.47	5.82	32.07 to 54.88	33.65	7.47	< 0.001
	Condition: Wood	− 12.23	8.23	− 28.36 to 3.90	33.65	− 1.49	0.146
	Phase: Recovery	− 6.20	5.03	− 16.07 to 3.67	40.00	− 1.23	0.225
	Phase: Task (1)	1.90	5.03	− 7.97 to 11.76	40.00	0.38	0.708
	Wood * Recovery	2.98	7.12	− 10.97 to 16.94	40.00	0.42	0.677
	Wood * Task (1)	− 0.08	7.12	− 14.03 to 13.88	40.00	− 0.01	0.992

Significant predictors ($p < 0.05$) are shown in bold

SE: standard error; CI: confidence interval; df: degrees of freedom; t: test statistic; p: p-value

*Interaction between predictors

**Fig. 7** Cardiovascular activity throughout study phases

markedly at Task (1) (Task (1)—Baseline = 22.61 beats per minute, 95% CI [11.56, 31.70], $p < 0.001$). Similarly, the atypical responders reacted to Task (1) with slightly (but insignificantly) increased heart rate variability (Task (1)—Baseline = 7.62 ms, 95% CI [− 13.48, 28.21], $p = 0.219$), in contrast with the typical responders, for whom the heart rate variability slightly (but insignificantly) decreased at Task (1) (Task (1)—Baseline = − 9.06 ms, 95% CI [− 19.20, 0.18], $p = 0.058$).

Mental arithmetic task

Participants on average generated more than 50 total responses to MAT ($M = 51.41$, 95% CI [45.72, 57.10]) with a very high proportion of correct responses ($M = 0.94$, 95% CI [0.93, 0.95]). The mixed models showed that the proportion of correct responses (Table 4) and the number of responses (Table 5) varied between study phases but not between desktop conditions, and there were no interactions between the desktop

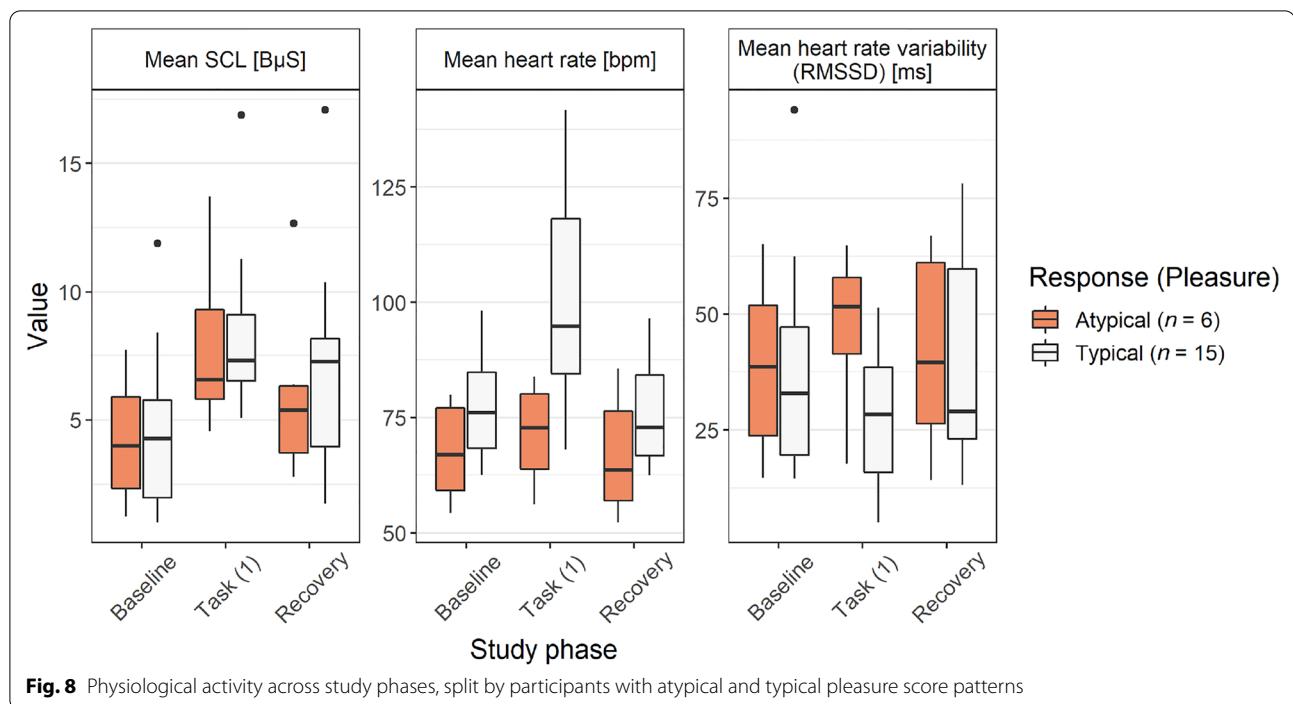


Fig. 8 Physiological activity across study phases, split by participants with atypical and typical pleasure score patterns

Table 4 Results of the binomial mixed model with MAT response correctness as the dependent variable

Outcome	Predictor	Estimate	SE	CI	z	p
Correct responses	Intercept	11.86	3.37	6.79 to 20.71	8.70	< 0.001
	Condition: Wood	1.31	0.55	0.58 to 2.96	0.65	0.513
	Phase: Task (2)	2.21	0.54	1.38 to 3.56	3.28	0.001
	Wood * Task (2)	0.78	0.29	0.37 to 1.63	− 0.66	0.512

The estimates represent odds ratios. Significant predictors ($p < 0.05$) are shown in bold

SE: standard error; CI: confidence interval; z: test statistic z; p: p-value

*Interaction between predictors

Table 5 Results of the linear mixed model with MAT total number of responses as the dependent variable

Outcome	Predictor	Estimate	SE	95% CI	df	t	p
Number of responses	Intercept	53.09	5.27	42.76 to 63.42	22.01	10.07	< 0.001
	Condition: Wood	− 12.91	7.46	− 27.52 to 1.7	22.01	− 1.73	0.097
	Phase: Task (2)	9.55	2.31	5.02 to 14.07	20.00	4.13	< 0.001
	Wood * Task (2)	0.00	3.27	− 6.4 to 6.4	20.00	0.00	1.000

Significant predictors ($p < 0.05$) are shown in bold

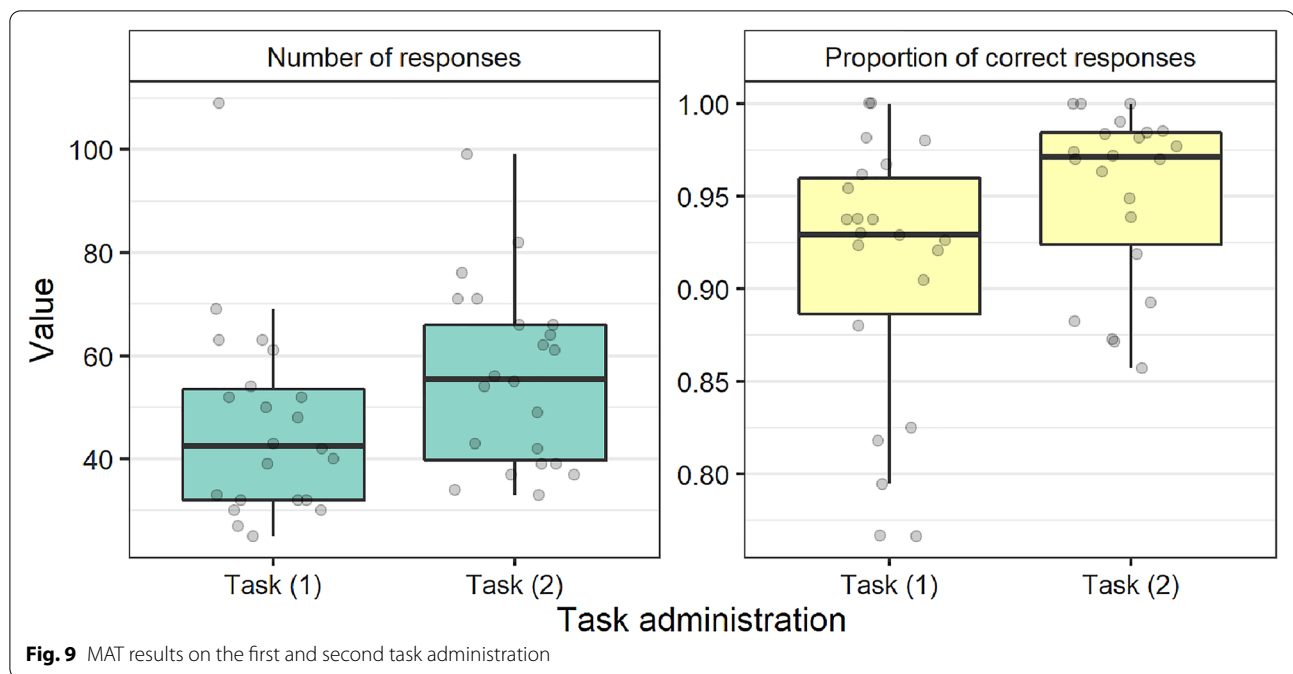
SE: standard error; CI: confidence interval; df: degrees of freedom; t: test statistic t; p: p-value

*Interaction between predictors

condition and task administration. Post hoc comparisons revealed that participants provided fewer responses in Task (1) ($EMM = 46.64$, 95% CI [38.91, 54.37]) than at Task (2) ($EMM = 56.18$, 95% CI [48.45, 63.91]; Task (1)–Task (2) = $− 9.55$, 95% CI [− 13.00, − 6.14], $p < 0.001$), and they were less likely to respond correctly to MAT at Task

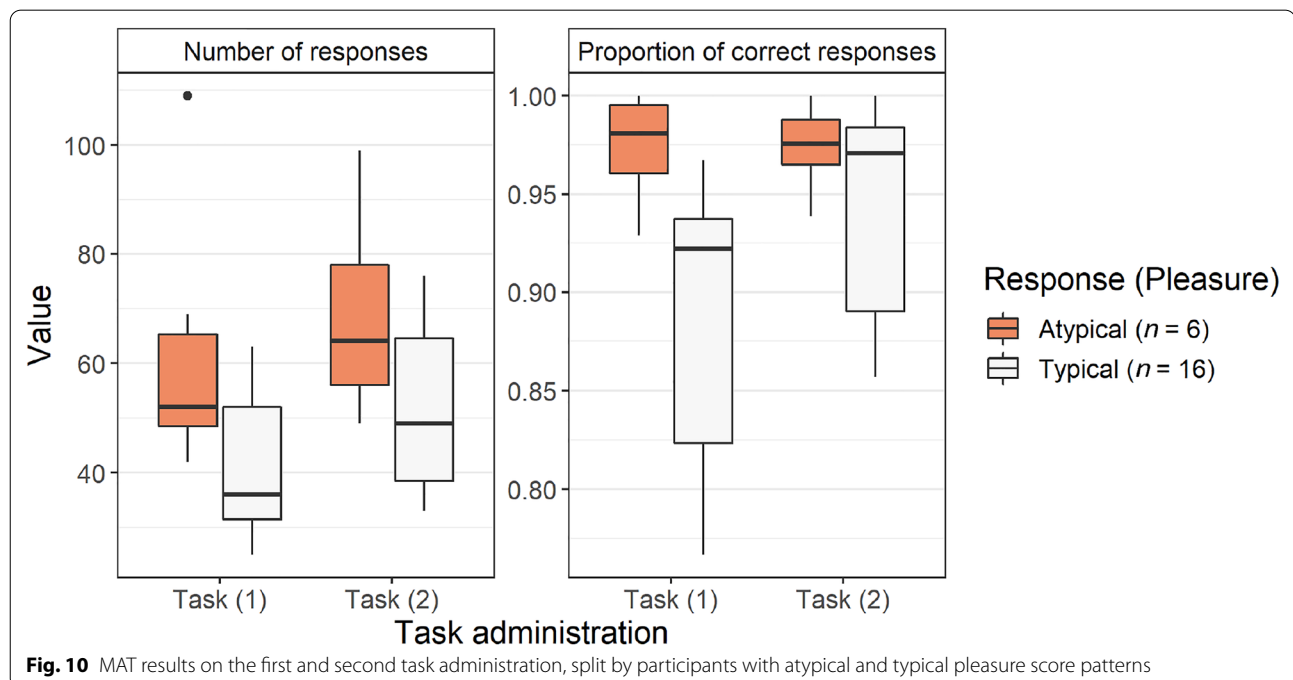
(1) ($EMM = 0.93$, 95% CI [0.90, 0.95]) than at Task (2) ($EMM = 0.96$, 95% CI [0.95, 0.98]; Task (1)/Task (2) (odds ratio) = 0.51, 95% CI [0.35, 0.74], $p < 0.001$) (Fig. 9).

A closer examination of participants who responded atypically on the affective state of pleasure (see “Affective states” section) suggests they provided more responses in



Task (1) than the typical responders (Atypical responders–typical responders=17.00, 95% CI [2.00, 37.00], $p=0.035$). The difference in number of responses between the two groups was similar in Task (2), although the variability of scores was greater and the difference was not statistically significant (Atypical responders–typical responders=16.87, 95% CI [−4.00, 33.00],

$p=0.090$) (Fig. 10, Additional file 1: Table S6). In addition, the atypical responders had a larger proportion of correct responses than the typical responders in Task (1) (Atypical responders–typical responders=0.06, 95% CI [0.03, 0.16], $p=0.004$) but similar proportion of correct responses in Task (2) (Atypical responders–typical responders=0.02, 95% CI [−0.01, 0.09], $p=0.376$).



Discussion

Affective states

Participants generally reacted to MAT with a state of higher arousal accompanied with middle values of pleasure, indicating a highly aroused state close to neutral in terms of valence, such as alertness or tension, but not stress [20]. Had the participants on average experienced a significant amount of stress, the experienced affective states should have been characterized by high arousal and low pleasure, such as anxiety [65]. In contrast, some participants reported *increased* subjective pleasure following MAT, suggesting that MAT sometimes induced more pleasurable states, such as excitement [20]. In the absence of low subjective pleasure, high subjective arousal following MAT likely primarily reflects the effort required to accomplish task demands [66]. This suggests that MAT does not lead to a reliable stress response in at least a subgroup of people, and that different or additional stressors are needed. At Recovery, the subjective arousal that MAT induced returned to levels similar to those observed at Baseline, suggesting that a 10-min recovery period is sufficiently long for affective states to return to initial values. The results also suggest that the deployed single-item measures assessing arousal and pleasure are sensitive enough to detect changes in affective states, as evidenced by the variability in scores, indicating that these scales may prove useful in restoration research.

Physiological activity

Electrodermal activity results generally followed the pattern observed in the self-reports of arousal—an increase after MAT followed by a decrease at Recovery. This pattern, however, differed between electrodermal activity parameters. SCL and SCR both increased from Baseline to Task (1), but SCL was higher at Recovery than at Baseline, while SCR returned to levels similar to those observed at Baseline. This indicates that MAT is capable of inducing increases of electrodermal activity, but that the period of 10 min may not be sufficient for the physiological arousal to return to baseline levels, suggesting that a longer recovery period is warranted. Unlike SCL and SCR, SCR amplitude did not significantly change throughout study phases. High SCL usually co-occurs with a high number of SCRs and large SCR amplitudes [38]; however, different electrodermal activity parameters may represent partially independent sources of information that are uniquely related to different psychophysiological processes. While all three electrodermal activity parameters are associated with strain, SCR amplitude is thought to also reflect preparatory activation, signaling increased perceptual and motor readiness

for an upcoming task [67, 68]. This suggests one possible explanation of the observed results: participants might have anticipated the upcoming task both at Baseline and Recovery, leading to increased values of SCR amplitude at these periods of rest to the point that these values did not significantly differ from those observed at Task (1). Alternatively, SCR amplitude may be less responsive to the specific type of demands placed on participants by MAT.

The patterns of cardiovascular activity resembled those of electrodermal activity for heart rate but not for heart rate variability. Heart rate increased from Baseline to Task (1) and then decreased at Recovery, to the point of being no different than at Baseline. Heart rate variability showed no such variation and remained similar throughout the study phases. When the heart rate increases following a stressor or an effortful cognitive task, the heart rate variability tends to decrease [69], making the cardiovascular responses observed in this study somewhat atypical. However, heart rate and heart rate variability are thought to provide partially independent information when it comes to stress and mental effort. Heart rate variability seems to be somewhat more sensitive to mental strain than heart rate [68], opening the possibility that participants were at least slightly tense at Baseline and Recovery periods, as they might have been anticipating the upcoming task, which could have been reflected in the heart rate variability not being significantly different than at Task (1). An alternative explanation is similar to the above interpretation related to unchanging SCR amplitudes: heart rate variability may be less responsive to the type of demands that participants faced on MAT. It should be noted, though, that the interaction between cardiovascular responses and arousal following stressors or cognitive tasks is complex, and a number of influences could be responsible for the observed results [67, 68].

MAT

The results of MAT showed that participants generally improved from the 1st to 2nd administration on both MAT outcomes: number of provided responses and the proportion of correct responses. This suggests that the potential cognitive fatigue induced by the 1st administration of the task was offset by learning and practice gained from completing the task. Alternatively, participants might have been more distracted at the 1st task administration, before getting acclimatized to the experimental session ahead of the 2nd administration of the task. Higher scores at the 2nd administration of the task could still show positive effects of restorative environments; indeed, many studies exploring attention restoration in natural environments detect higher scores at the 2nd task administration [8, 33]. However, the attention restoration

theory claims that exposure to nature *restores* fatigued cognitive capacities [6]. This suggests that the positive effects of natural environments on cognitive performance will be less likely present if participants are not cognitively fatigued and operate at their peak cognitive capacities, leaving the natural environment no maneuvering space: cognitive capacities cannot be restored if they have not been depleted. It is unclear, though, whether the observed effect of natural environments on cognitive performance is in fact the restoration of a depleted cognitive resource [8, 70, 71]. Still, inducing cognitive fatigue seems more likely to lead to a more reliable restoration effect, at least on some occasions [8], and the 5-min instance of MAT may not be sufficient to induce significant levels of cognitive fatigue.

Atypical versus typical responders on the affective state of pleasure

Some participants reacted to MAT with an increased affective state of pleasure—the opposite of what would be expected if they had experienced stress. In response to MAT, these atypical responders appeared to have similar electrodermal activity but lower cardiovascular activity than the typical responders. Perhaps this discrepancy can be explained by different properties of the two physiological systems: electrodermal activity is a relatively direct measure of sympathetic activity of the autonomic nervous system, while heart rate provides a broader picture of both sympathetic and parasympathetic activity [38]. The atypical responders may have been sufficiently activated for the increased sympathetic activity to be detected on the measure of electrodermal activity but not activated enough for the activation to be evident in heart rate, which also involves parasympathetic activity. Increased parasympathetic activity in the atypical responders could also be indicated by their slight increase in heart rate variability in response to MAT [39].

Interestingly, even though physiological activity was somewhat different between typical and atypical responders in response to MAT, subjective arousal was similar in both groups of participants. This suggests that subjective arousal cannot be fully explained by measures of electrodermal and cardiovascular activity. It is also possible that subjective assessment cannot capture arousal as precisely as physiological measures, due to the subjectivity involved. Based on the identified discrepancies between subjective and physiological arousal, it appears that both types of arousal should be measured to obtain a more complete understanding of arousal in the studied situation.

The affective and physiological response of the atypical responders—higher pleasure and lower

physiological activity—might be explained in part by their better performance on MAT. Perhaps these participants reacted to MAT atypically due to their higher ability or affinity for cognitive tasks, suggesting that MAT might not lead to stress especially in people who are more capable or motivated to perform on cognitive tasks.

Outcomes in wooden versus non-wooden desktop conditions

Affective states, electrodermal and cardiovascular activity, and cognitive performance did not differ between desktop conditions (i.e., wooden desktop versus desktop covered with a white cloth). This can be due to the low number of participants, making the study underpowered to detect presumably small effects of the exposure to a wooden setting. Another reason for the lack of detected differences can stem from the absence of a clear stress response and cognitive fatigue in participants: if participants did not experience stress or cognitive fatigue, it could have been more difficult for the environment to provide restorative effects [8]. The lack of observed differences between environments could also have resulted from the specific wood furnishings: the wooden desk may not have provided sufficient stimulation to induce restorative effects. The existing studies that observed the most promising effects of wood exposure on people used rooms with larger wood coverage [13, 14, 72], suggesting that even a relatively large desk surface tested in our study might not be sufficiently large to provide restorative effects.

Limitations

The most obvious limitations of the study are related to the nature and size of the study sample. Most participants were at least loosely acquainted with the study's first author, who was leading the experimental sessions. This might have urged participants to behave and respond differently than they would have in a more neutral context. In addition, the study sample was imbalanced in terms of gender, with most of the participants being female, and we did not control for the menstrual cycle phase, which could have impacted the results. The age range of participants was somewhat wide, and some variability in stress reactivity between participants may have been a result of differences in age. A relatively small sample size may have left the study underpowered—not only unable to detect potential differences in outcomes between desktop conditions but also unable to identify some of the potential

subtle changes in outcomes across all participants, such as small differences in pleasure scores between study phases.

Conclusions and recommendations for future studies

On average, MAT may not lead to a reliable stress response. The task generally increased self-reported arousal and most measures of physiological arousal, indicating that it successfully activated participants to an extent. However, MAT did not impact all measures of physiological arousal, and it did not significantly affect the self-reported affective state of pleasure, indicating that the *average* response of participants cannot be straightforwardly interpreted as a stress response, but instead as activation required to successfully meet task demands. Clear stress response in the entire sample may have not appeared mainly due to a subgroup of participants who reacted to MAT positively—with increased affective state of pleasure. The role of MAT as a cognitive task in restoration research seems similarly limited, at least when MAT lasts only 5 min and when the goal is to reliably induce cognitive fatigue. However, MAT might become more useful if it would be longer (to attempt to induce cognitive fatigue) and if the testing condition would be more threatening (to attempt to induce stress), for example, by including a larger evaluative audience. The single-item measures that examined affective states seemed to be sufficiently sensitive to detect changing states of pleasure and arousal for their use to be recommended in restoration research. The comparison of outcomes between desktop conditions revealed that a larger wooden desktop is unlikely to lead to considerable restorative effects, but larger studies might detect potential (smaller) effects of the exposure to wooden desks, especially if the wood coverage increases. Taken together, the results of this study can inform and guide future studies, increasing their chances to recognize restorative environments.

Future studies may benefit from piloting their experimental design and measures before engaging larger subject pools. Methodological investigations are needed to identify how to induce an adequate degree of stress and cognitive fatigue for restoration studies, which would support more robust and comparable research in the field. For example, testing a longer version of MAT may reveal more about its capacity to reliably induce cognitive fatigue and stress. The single-item measures of affective states used in this study were robust, and we encourage other researchers to use them. However, comparing them with more commonly used (and longer) measures (e.g., PANAS) in the context of restoration research would be a useful contribution. The

settings where the studies are deployed should be assessed in detail to examine how people are affected by characteristics such as indoor air quality, amount of natural elements (e.g., plants, wood), light quality, and other properties.

Abbreviations

ANOVA: Analysis of variance; ART: Attention restoration theory; CI: Confidence interval; df: Degrees of freedom; EMM: Estimated marginal mean; M: Mean; MAT: Mental Arithmetic Task; PANAS: Positive and Negative Affect Schedule; RMSSD: Root mean square of successive beat-to-beat interval differences; SCL: Skin conductance level; SCR: Skin conductance response; SE: Standard error; SRT: Stress reduction theory; TSST: Trier Social Stress Test.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s10086-022-02042-5>.

Additional file 1: Table S1. Pleasure scores across study phases for participants with unusual and usual responses. **Table S2.** Arousal scores across study phases for participants with unusual and usual pleasure response patterns. **Table S3.** Comparison of arousal scores across study phases between participants with unusual and usual pleasure response patterns using the Wilcoxon rank-sum test. **Table S4.** Physiological activity across study phases for participants with unusual and usual pleasure response patterns. **Table S5.** Comparison of physiological activity across study phases within participants with unusual and usual pleasure response patterns using the Wilcoxon signed-rank test. **Table S6.** MAT results on the first and second task administration, split by participants with unusual and usual pleasure score patterns. **Figure S1.** Arousal scores across study phases for participants with unusual and usual pleasure response patterns.

Acknowledgements

The authors gratefully acknowledge the European Commission for funding the InnoRenew CoE project (Grant Agreement 739574) under the Horizon2020 Widespread-Teaming program and the Republic of Slovenia (Investment funding of the Republic of Slovenia and the European Union of the European Regional Development Fund).

Author contributions

DL and MB designed the study. DL and JŽ collected the data. DL analyzed the data and prepared the first draft of the manuscript. All authors read and approved the final manuscript.

Funding

The research was funded through the InnoRenew CoE project (Grant Agreement 739574) under the Horizon2020 Widespread-Teaming program and the Republic of Slovenia (Investment funding of the Republic of Slovenia and the European Union of the European Regional Development Fund). The funding body had no influence on the design of the study; collection, analysis, and interpretation of data; and writing the manuscript.

Availability of data and materials

The datasets and the R analysis code used in the current study are available in the Zenodo repository, <https://doi.org/10.5281/zenodo.5616116>.

Declarations

Competing interests

The authors declare that they have no competing interests.

Author details

¹InnoRenew CoE, Livade 6, 6310 Izola, Slovenia. ²Andrej Marušič Institute, University of Primorska, Muzejski trg 2, 6000 Koper, Slovenia. ³Faculty of Health Sciences, University of Primorska, Polje 42, 6310 Izola, Slovenia.

Received: 15 November 2021 Accepted: 3 June 2022
Published online: 15 June 2022

References

- McMahan EA, Estes D (2015) The effect of contact with natural environments on positive and negative affect: a meta-analysis. *J Posit Psychol* 10:507–519. <https://doi.org/10.1080/17439760.2014.994224>
- Kondo MC, Jacoby SF, South EC (2018) Does spending time outdoors reduce stress? A review of real-time stress response to outdoor environments. *Heal Place* 51:136–150. <https://doi.org/10.1016/j.healthplace.2018.03.001>
- Roberts H, van Lissa C, Hagedoorn P, Kellar I, Helbich M (2019) The effect of short-term exposure to the natural environment on depressive mood: a systematic review and meta-analysis. *Environ Res* 177:1–14. <https://doi.org/10.1016/j.envres.2019.108606>
- Yao W, Zhang X, Gong Q (2021) The effect of exposure to the natural environment on stress reduction: a meta-analysis. *Urban For Urban Green* 57:1–12. <https://doi.org/10.1016/j.ufug.2020.126932>
- Ulrich RS (1983) Aesthetic and affective response to natural environment. *Behav Nat Environ Adv Theory Res* 6:85–125. <https://doi.org/10.1007/978-1-4613-3539-9>
- Kaplan S (1995) The restorative benefits of nature: toward an integrative framework. *J Environ Psychol* 15:169–182. [https://doi.org/10.1016/0272-4944\(95\)90001-2](https://doi.org/10.1016/0272-4944(95)90001-2)
- Kaplan R, Kaplan S (1989) *The experience of nature: a psychological perspective*. Cambridge, UK
- Stevenson MP, Schilhab T, Bentsen P (2018) Attention restoration theory II: a systematic review to clarify attention processes affected by exposure to natural environments. *J Toxicol Environ Heal Part B Crit Rev* 21:227–268. <https://doi.org/10.1080/10937404.2018.1505571>
- Ohly H, White MP, Wheeler BW, Bethel A, Ukoumunne OC, Nikolaou V, Garside R (2016) Attention restoration theory: a systematic review of the attention restoration potential of exposure to natural environments. *J Toxicol Environ Heal - Part B Crit Rev* 19:305–343. <https://doi.org/10.1080/10937404.2016.1196155>
- Burnard MD, Nyrud AQ, Bysheim K, Kutnar A, Vahtikari K, Hughes M (2017) Building material naturalness: perceptions from Finland, Norway and Slovenia. *Indoor Built Environ* 26:92–107. <https://doi.org/10.1177/1420326X15605162>
- Burnard MD, Kutnar A (2015) Wood and human stress in the built indoor environment: a review. *Wood Sci Technol* 49:969–986. <https://doi.org/10.1007/s00226-015-0747-3>
- Lipovac D, Burnard MD (2020) Effects of visual exposure to wood on human affective states, physiological arousal and cognitive performance: a systematic review of randomized trials. *Indoor Built Environ*. <https://doi.org/10.1177/1420326X20927437>
- Fell DR (2010) Restorative properties of wood in the built indoor environment (PhD Thesis). The University of British Columbia, Vancouver
- Burnard MD, Kutnar A (2020) Human stress responses in office-like environments with wood furniture. *Build Res Inf* 48:316–330. <https://doi.org/10.1080/09613218.2019.1660609>
- Lipovac D, Podrekar N, Burnard MD, Šarabon N (2020) Effect of desk materials on affective states and cognitive performance. *J Wood Sci* 66:1–12. <https://doi.org/10.1186/s10086-020-01890-3>
- Tsunetsugu Y, Miyazaki Y, Sato H (2007) Physiological effects in humans induced by the visual stimulation of room interiors with different wood quantities. *J Wood Sci* 53:11–16. <https://doi.org/10.1007/s10086-006-0812-5>
- McSweeney J, Rainham D, Johnson SA, Sherry SB, Singleton J (2015) Indoor nature exposure (INE): a health-promotion framework. *Health Promot Int* 30:126–139. <https://doi.org/10.1093/heapro/dau081>
- Klepeis NE, Nelson WC, Ott WR, Robinson JP, Tsang AM, Switzer P, Behar JV, Herr SC, Engelmann WH (2001) The national human activity pattern survey (NHAPS): a resource for assessing exposure to environmental pollutants. *J Expo Sci Environ Epidemiol* 11:231–252. <https://doi.org/10.1038/sj.jea.7500165>
- Watson D, Clark LA, Tellegen A (1988) Development and validation of brief measures of positive and negative affect: the PANAS scales. *J Pers Soc Psychol* 54:1063–1070. <https://doi.org/10.1037/0022-3514.54.6.1063>
- Russell JA (1980) A circumplex model of affect. *J Pers Soc Psychol* 39:1161–1178. <https://doi.org/10.1037/h0077714>
- Russell JA, Weiss A, Mendelsohn GA (1989) Affect grid: a single-item scale of pleasure and arousal. *J Pers Soc Psychol* 57:493–502. <https://doi.org/10.1037/0022-3514.57.3.493>
- Ekkekakis P (2013) *The measurement of affect, mood, and emotion: A guide for health-behavioral research*. Cambridge University Press, New York
- Berntson GG, Cacioppo JT (2000) From homeostasis to allodynamic regulation. In: Cacioppo JT, Tassinary LG, Berntson GG (eds) *Handbook of psychophysiology*, 2nd edn. Cambridge University Press, New York, pp 459–481
- Kreibitz SD (2010) Autonomic nervous system activity in emotion: a review. *Biol Psychol* 84:14–41. <https://doi.org/10.1016/j.biopsycho.2010.03.010>
- Siegel EH, Sands MK, Van den Noortgate W, Condon P, Chang Y, Dy J, Quigley KS, Barrett LF (2018) Emotion fingerprints or emotion populations? A meta-analytic investigation of autonomic features of emotion categories. *Psychol Bull* 144:343–393. <https://doi.org/10.1037/bul0000128>
- Yoshino K, Matsuoka K (2011) Correlation between mood and heart rate variability indices during daily life. *Health* 3:553–556. <https://doi.org/10.4236/health.2011.39094>
- Matusz A, van der Linden D, Kisander Z, Hernádi I, Kázmér K, Csathó Á (2021) Enhanced cardiac vagal tone in mental fatigue: analysis of heart rate variability in Time-on-Task, recovery, and reactivity. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0238670>
- Feldman PJ, Cohen S, Hamrick N, Lepore SJ (2004) Psychological stress, appraisal, emotion and cardiovascular response in a public speaking task. *Psychol Health* 19:353–368. <https://doi.org/10.1080/0887044042000193497>
- Dickerson SS, Kemeny ME (2004) Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. *Psychol Bull* 130:355–391. <https://doi.org/10.1037/0033-2909.130.3.355>
- Birkett MA (2011) The trier social stress test protocol for inducing psychological stress. *JoVE* 56:1–6. <https://doi.org/10.3791/3238>
- Labuschagne I, Grace C, Rendell P, Terrett G, Heinrichs M (2019) An introductory guide to conducting the Trier Social Stress Test. *Neurosci Biobehav Rev* 107:686–695. <https://doi.org/10.1016/j.neubiorev.2019.09.032>
- Goodman WK, Janson J, Wolf JM (2017) Meta-analytical assessment of the effects of protocol variations on cortisol responses to the Trier Social Stress Test. *Psychoneuroendocrinology* 80:26–35. <https://doi.org/10.1016/j.psyneuen.2017.02.030>
- Hartig T, Jahncke H (2017) Letter to the editor: attention restoration in natural environments: mixed mythical metaphors for meta-analysis. *J Toxicol Environ Heal Part B* 20:305–315. <https://doi.org/10.1080/10937404.2017.1363101>
- Neilson BN, Craig CM, Travis AT, Klein MI (2019) A review of the limitations of Attention Restoration Theory and the importance of its future research for the improvement of well-being in urban living. *Visions Sustain* 11:59–67
- DeStefano D, LeFevre JA (2004) The role of working memory in mental arithmetic. *Eur J Cogn Psychol* 16:353–386. <https://doi.org/10.1080/09541440244000328>
- Ekkekakis P (2008) Affect circumplex redux: the discussion on its utility as a measurement framework in exercise psychology continues. *Int Rev Sport Exerc Psychol* 1:139–159. <https://doi.org/10.1080/17509840802287200>
- Killgore WDS (1998) The affect grid: a moderately valid, nonspecific measure of pleasure and arousal. *Psychol Rep* 83:639–642. <https://doi.org/10.2466/pr0.1998.83.2.639>
- Dawson ME, Schell AM, Filion DL (2007) The electrodermal system. In: University C (ed) *Handbook of psychophysiology*, 3rd. Press, New York, pp 159–181
- Berntson GG, Quigley KS, Norman GJ, Lozano DL (2016) Cardiovascular psychophysiology. In: *Handbook of Psychophysiology*. Cambridge University Press, pp 183–216
- Kim HG, Cheon EJ, Bai D-S, Lee YH, Koo B-H (2018) Stress and heart rate variability: a meta-analysis and review of the literature. *Psychiatry Investig* 15:235–245

41. Laborde S, Mosley E, Thayer JF (2017) Heart rate variability and cardiac vagal tone in psychophysiological research—Recommendations for experiment planning, data analysis, and data reporting. *Front Psychol* 8:1–18. <https://doi.org/10.3389/fpsyg.2017.00213>
42. ADInstruments (2021) LabChart: Life science data acquisition software, ADInstruments, Dunedin, New Zealand
43. Makowski D, Pham T, Lau ZJ, Brammer JC, Lespinasse F, Pham H, Schölzel C, Chen SHA (2021) NeuroKit2: a python toolbox for neurophysiological signal processing. *Behav Res Methods*. <https://doi.org/10.3758/s13428-020-01516-y>
44. Yeragani VK, Krishnan S, Engels HJ, Gretebeck R (2005) Effects of caffeine on linear and nonlinear measures of heart rate variability before and after exercise. *Depress Anxiety* 21:130–134. <https://doi.org/10.1002/da.20061>
45. Harrison P (2020) psychTestR: An R package for designing and conducting behavioural psychological experiments. *J Open Source Softw* 5:1–3. <https://doi.org/10.21105/joss.02088>
46. Allen AP, Kennedy PJ, Cryan JF, Dinan TG, Clarke G (2014) Biological and psychological markers of stress in humans: focus on the trier social stress test. *Neurosci Biobehav Rev* 38:94–124. <https://doi.org/10.1016/j.neubiorev.2013.11.005>
47. Liu Q, Zhang W (2020) Sex differences in stress reactivity to the Trier Social Stress Test in virtual reality. *Psychol Res Behav Manag* 13:859–869. <https://doi.org/10.21203/rs.3.rs-21036/v1>
48. Peck K, Totony de Zepetnek J, Fiocco AJ (2021) Music listening does not inoculate the stress response in young and older adults. *Int J Stress Manag* 28:154–164. <https://doi.org/10.1037/str0000217>
49. R Core Team (2021) R: a language and environment for statistical computing
50. Van Rossum G, Drake Jr FL (1995) Python reference manual. Centrum voor Wiskunde en Informatica, Amsterdam, NL
51. RStudio Team (2021) RStudio: Integrated Development Environment for R. RStudio, Boston
52. Firke S, Denney B, Haid C, Knight R, Grosser M, Zadra J (2020) janitor: Simple tools for examining and cleaning dirty data. R package version 2.1.0
53. Bolker B, Robinson D (2020) broom.mixed: Tidying methods for mixed models. R package version 0.2.6
54. Kassambara A (2020) rstatix: Pipe-friendly framework for basic statistical tests. R package version 0.7.0
55. Ushey K, Allaire JJ, Tang Y (2020) Reticulate: interface to python. R Package Version 1:18
56. Bates D, Maechler M, Bolker B, Walker S (2020) lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1–26
57. Kuznetsova A, Brockhoff PB, Christensen RHB (2017) lmerTest Package: Tests in linear mixed effects models. *J Stat Softw* 82:1–26. <https://doi.org/10.18637/jss.v082.i13>
58. Lenth RV (2021) emmeans: Estimated marginal means, aka least-squares means. R package version 1.6.0
59. Hartig F (2021) DHARMa: Residual diagnostics for hierarchical (multi-level/mixed) regression models. R package version 0.4.1
60. Gohel D (2021) flextable: Functions for tabular reporting. R package version 0.6.8
61. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019) Welcome to the tidyverse. *J Open Source Softw* 4:1–6. <https://doi.org/10.21105/joss.01686>
62. Gueorguieva R, Krystal JH (2004) Move over ANOVA: progress in analyzing repeated-measures data and its reflection in papers published in the Archives of general psychiatry. *Arch Gen Psychiatry* 61:310–317. <https://doi.org/10.1001/archpsyc.61.3.310>
63. Faraway JJ (2016) Extending the linear model with R generalized linear, mixed effects and nonparametric regression models, 2nd edn. CRC Press: Taylor & Francis Group, New York
64. Rusticus SA, Lovato CY (2014) Impact of sample size and variability on the power and type I error rates of equivalence tests: a simulation study. *Pract Assessment Res Eval* 19:1–10
65. Feldman PJ, Cohen S, Lepore SJ, Matthews KA, Kamarck TW, Marsland AL (1999) Negative emotions and acute physiological responses to stress. *Ann Behav Med* 21:216–222. <https://doi.org/10.1007/BF02884836>
66. Kahneman D (1973) Attention and effort. Prentice-Hall Inc., Englewood Cliffs, New Jersey
67. Boucsein W (2012) Electrodermal Activity, 2nd edn. Springer, New York
68. Boucsein W, Backs RW (2008) The psychophysiology of emotion, arousal, and personality: Methods and models. In: Duffy VG (ed) Handbook of Digital Human Modeling. CRC Press: Taylor & Francis Group, New York, pp 35–1–35–18
69. Hilgarter K, Schmid-Zalaudek K, Csanády-Leitner R, Mörtl M, Rössler A, Lackner HK (2021) Phasic heart rate variability and the association with cognitive performance: a cross-sectional study in a healthy population setting. *PLoS ONE* 16:1–19. <https://doi.org/10.1371/journal.pone.0246968>
70. Parsons R, Tassinary LG (2002) Environmental psychophysiology. Handbook of environmental psychology. John Wiley & Sons, Hoboken, pp 172–190
71. Parsons R (2007) Environmental Psychophysiology. In: Cacioppo J, Tassinary LG, Berntson GG (eds) The handbook of psychophysiology, 3rd edn. Cambridge University Press, New York, pp 752–786
72. Zhang X, Lian Z, Ding Q (2016) Investigation variance in human psychological responses to wooden indoor environments. *Build Environ* 109:58–67. <https://doi.org/10.1016/j.buildenv.2016.09.014>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)