Supplemental material.

Restore or get restored: The effect of control on stress reduction and restoration in virtual nature settings

Supplement A – Analysis of active vs. passive control without participants who indicated motion sickness

As it is plausible to assume that motion sickness has an effect on restoration, we repeated the same analyses with those fifty participants that did not report motion sickness. As can be seen below, there are only very small changes. For subjective stress, the interaction is now non-significant; the main effect of time within the no control condition, however, similarly strong.

Table 1. Descriptive statistics of the outcome variables. (*N* = 50)

	Control			
Measure	Yes		No	
	M	*SD*	*M*	*SD*
Positive affect t1	3.16	0.47	3.26	0.65
Positive affect t2	3.51	0.72	3.73	0.68
Negative affect t1	1.30	0.32	1.37	0.46
Negative affect t2	1.14	0.22	1.18	0.26
Stress t1	2.51	0.53	2.6	0.40
Stress t2	2.45	0.58	2.43	0.39
Restoration outcome	3.76	0.80	3.87	0.84
Perceived restorativeness	4.56	0.87	4.56	0.71

Positive and negative affect

We submitted the positive and negative PANAS scores into 2(Control: active vs passive) X 2(Affect: Before vs After) ANOVAs with repeated measures on the second factor. Results revealed that positive affect was significantly higher after the VR experience, $F(1,48) = 30.63$, $p < .001$, $\eta_p^2 = .39$, and negative affect was lower, $F(1,48) = 22.69$, $p < .001$, $\eta_p^2 = .32$, (see Table 1). There was no main effect of control, and no interaction for positive emotions, both $Fs<1$. There was also no main effect and no interaction for negative emotions, both $Fs<1$.

Subjective Stress

We submitted the subjective stress score into a 2(Control: Active vs passive) X2(Stress: Before vs After) ANOVA with repeated measures on the second factor. Results revealed that subjective stress was significantly lower after the VR experience, $F(1, 48) = 14.69$, $p<.001$, $\eta_p^2 = .23$. There was no main effect of control, $F<1$, but the effect of measurement time was qualified by a non-significant interaction, $F(1, 62) = 3.67$, $p=.006$, $\eta_p^2 = .07$, suggesting that among participants in the "no control" condition, stress was reduced significantly from t1 to t2 ($M_{diff\_t1t2stress} = 0.1$, SE = .043, $p < .001$). In the active condition, there was no significant change ($M_{diff\_t1t2stress} = 0.06$, SE = .043, $p = .18$).

Restoration and perceived restorativeness

To test the difference between an active vs. passive VR walk with regard to restoration and perceived restorativeness, we submitted the ROS scale and the perceived

restorativeness scale each to a t-test for independent variables. There were no significant differences – neither for the ROS, t<1, nor for perceived restorativeness, t<1, but mean values were relatively high (i.e., significantly about the scale midpoint "3": $M_{Ros} = 3.82$, $t[49] = 7.11$, $p < .001$; $M_{perceived\_restorativeness} = 4.56$, $t[49] = 14.02$, $p < .001$) and similar compared to a previous study (i.e., Mattila et al., 2020, who report a mean of $M = 5.22$ [SD = 0.97] for ROS on a 7-point Likert scale).

Supplement B.

It is ongoing scientific debate if and under which conditions multiple tests of different psychological constructs require alpha-error correction such as Bonferroni correction [1-4], and we decided to refrain from using them in the main analysis. However, for the repeated measures positive and negative affect, as well as stress, we present a MANOVA with difference scores in the following to account for multiple testing.

We conducted a GLM MANOVA with the difference scores (t1-t2) of positive affect, negative affect, and stress. The overall effect of the MANOVA was non-significant, $F(3,60) = 2.22$, $p = .095$, $\eta^2p = .10$. However, as expected and mirroring the analysis in the main text, there was the effect of the manipulation on stress such that only in the "no control" condition, stress was reduced significantly, $F(1, 62) = 6.63$, $p = .012$, $\eta^2p = .10$. For positive and negative affect, there were no significant differences between the two conditions, both Fs<1. Mean difference scores are depicted in Table 2.

Table 2. *Mean difference scores of the repeated measure variables. Difference sores were computed with t1 – t2. (N=64).*

Variable	Overall		Active control		No control	
	M	*SD*	*M*	*SD*	*M*	*SD*
Positive affect	-.37	.55	-.33	.54	-.41	.56
Negative affect	.15	.26	.14	.16	.16	.34
Stress	.12	.23	.05	.20	.19	.23

References

1. Nakagawa, S. A farewell to Bonferroni: The problems of low statistical power and publication bias. *Behavioral Ecology* **2004**, *15*(6), 1044–1045. https://doi.org/10.1093/beheco/arh107

2. Perneger, T. V. What's wrong with Bonferroni adjustments. *BMJ* **1998**, *316*(7139), 1236–1238. https://doi.org/10.1136/bmj.316.7139.1236

3. Sinclair, J.; Taylor, P. J.; Hobbs, S. J. Alpha level adjustments for multiple dependent variable analyses and their applicability–a review. *Int J Sports Sci Eng* **2013**, *7*(1), 17–20.

4. Forstmeier, W.; Schielzeth, H. Cryptic multiple hypotheses testing in linear models: overestimated effect sizes and the winner's curse. *Behavioral Ecology and Sociobiology* **2011**, *65(1)*, 47-55