Validity and scope sensitivity of the Work for Environmental Protection Task

Florian Lange & Siegfried Dewitte

Behavioral Economics and Engineering Group, KU Leuven, Leuven, Belgium

Correspondence:

Florian Lange

BEE – Behavioral Economics and Engineering Group

KU Leuven

Naamsestraat 69

3000 Leuven

Belgium

florian.lange@kuleuven.be

This version of this manuscript has been accepted for publication in the *Journal of Environmental Psychology* on 21 January 2023 (DOI: 10.1016/j.jenvp.2023.101967).

Acknowledgements

This project was supported by a FWO postdoctoral fellowship (No 12U1221N) awarded to Florian Lange.

Declarations of interest: none.

Abstract

Pro-environmental behaviors are defined by their consequences, but often studied using inconsequential procedures. The validity problems associated with self-report proxies and hypothetical scenarios can be addressed by studying consequential pro-environmental behavior in behavioral paradigms. Here, we examine the validity of a recently developed paradigm that involves repeated trade-offs between individual and environmental consequences. On the Work for Environmental Protection Task (WEPT) participants can make real voluntary efforts to produce actual donations to an environmental organization. Responding to the call for cross-culturally robust methods and insights, we administered the WEPT to residents of the UK, the US, and South Africa (total $N = 1175$). Preregistered analyses revealed WEPT efforts to be internally consistent, effectively deterred by behavioral costs, sensitive to the scope of environmental impact, and correlated to people's general pro-environmental propensity. These results suggest the WEPT to be suited for the experimental analysis of pro-environmental behavior in the investigated populations.

Keywords: pro-environmental behavior; measurement; performance-based assessment; validity; scope sensitivity; Work for Environmental Protection Task

1. Introduction

Pro-environmental behavior is defined by its consequences for the natural environment. To be considered pro-environmental, a behavior either needs to produce relative environmental benefits (impact-oriented definition) or to be performed because of its environmental benefits (intent-oriented definition; Stern, 2000). However, these benefits for the environment are not the only consequence linked to pro-environmental behavior. Environmental benefits are typically produced at a behavioral cost (e.g., in terms of money, time, or effort) that may deter people from engaging in pro-environmental behaviors (Gifford, 2011; Kaiser, 2021). Understanding how people navigate the trade-off between environmental benefits and behavioral costs is critical for a science of pro-environmental behavior and behaviorally informed strategies for mitigating environmental issues.

In spite of their definitional importance, the consequences of pro-environmental behavior often go by the board when pro-environmental behavior is studied. Rather than studying situations of actual environmental relevance and behaviors with actual costs, large parts of pro-environmental behavior research focus on the analysis of inconsequential verbal responses to self-report items and hypothetical scenarios (Gifford, 2014; Lange et al., 2018; Steg & Vlek, 2009). Such verbal responses can be of scientific interest per se (e.g., Wille & Lange, 2022) and they may serve as useful indicators of people's commitment towards environmental protection (Kaiser & Wilson, 2004; Kaiser & Lange, 2021; Kaiser et al., 2018). Yet they cannot be expected to produce results that would generalize to situations of actual environmental relevance (Klein & Hilbig, 2019; Lewandowski & Strohmetz, 2009). External validity (i.e., the generalizability of results obtained in an experimental situation to naturally occurring situations) depends on whether the consequences implemented in the experimental situation reflect the relevant consequences in the naturally occurring situation of interest (Lange & Dewitte, 2019; Schmuckler, 2001). As making verbal statements about engagement

in pro-environmental behaviors shares few of the consequences of actual behavioral engagement, the external validity of verbal response experiments seems questionable.

An alternative to the study of inconsequential verbal responses is the use of behavioral paradigms or experimental models of pro-environmental behavior. Behavioral paradigms are systematically arranged model situations that involve the same relationships between behaviors and their consequences as the situations that are supposed to be modelled (Lange, 2022). A recent review (Lange, 2022) has revealed a diverse range of pro-environmental behavior paradigms used in environmental psychology, environmental economics, and adjacent fields of research. For example, researchers have confronted their participants with consequential product choices (e.g., between organic and conventional products; Taube & Vetter, 2019), observed how much energy and water participants used when testing the qualities of a towel (e.g., Zhang et al., 2021), or provided them with an opportunity to recycle their study materials (e.g., Linder et al., 2021). A core limitation of most of these behavioral paradigms is that they have been created ad hoc to directly address a substantive research question and did not undergo an independent validation process (Lange, 2022). As a corollary, it often remains unknown whether people are sensitive to the consequences implemented in a behavioral paradigm, how consistently they respond to the arranged model situations, and how their performance relates to people's propensity to behave pro-environmentally in general. Reliance on unvalidated methods will likely jeopardize the robustness of behavioral research (Flake & Fried, 2020) and the proliferation of ad hoc procedures complicates the systematic integration of research findings across studies (Lange, 2019).

Validated paradigms for the study of consequential pro-environmental behavior can provide an obvious solution to these problems and have begun to receive attention in pro-environmental behavior research (Berger & Wyss, 2021a; Klein et al., 2017; Lange et al.,

2018). Most recently, Lange and Dewitte (2022) have developed the Work for Environmental Protection Task (WEPT) as a multi-trial procedure involving repeated opportunities to make a behavioral effort to the benefit of the natural environment. On every WEPT trial, participants have the option to voluntarily work on a tedious number screening task and for every page of numbers they complete, a monetary donation is made to an environmental organization. An initial validation study indicated that people take these consequences into account when making choices in the WEPT (Lange & Dewitte, 2022). Participants' choices were found to be sensitive to manipulations of both the behavioral costs (i.e., the amount of effort required to complete a page of number) and the environmental benefits (i.e., the amount of money being donated) of WEPT behavior. The paradigm thus seems to involve an actual trade-off between environmental benefits and behavioral costs as is characteristic for many naturally occurring situations of environmental relevance. In addition, WEPT performance was found to be highly reliable (r_{SB} = .92) and to be correlated to other observations of pro-environmental behavior and to self-report measures related to the general propensity to behave pro-environmentally (e.g., environmental attitude, environmental concern, biospheric values; Lange & Dewitte, 2021, 2022).

Importantly, this support for the validity of the WEPT was generated in a sample drawn from a student-dominated subject pool of a Belgian university. Given that few researchers will be able to sample participants from that same population, replications in other populations (or cross-cultural validation studies; Tam & Milfont, 2020) are necessary to gauge the generality of these results. The psychometric properties of test scores are necessarily population-dependent (e.g., Caruso, 2000; Kopp et al., 2021) and people's sensitivity to the consequences implemented in the WEPT may vary across populations as well. For example, people from other (non-Belgian, non-student) populations may not mind the small time and effort costs associated with completing WEPT pages (e.g., when being less

pressed for time or having more autonomy about their schedules) or they may not care about the associated donations (e.g., because they do not trust or value the work of environmental organizations or they judge the monetary amounts to be negligible). In both these cases, the WEPT would not involve a trade-off between valued environmental benefits and behavioral costs and WEPT-based results could not be expected to be externally valid in the respective population.

1.1 Scope sensitivity of pro-environmental behavior

People's responsiveness to the environmental consequences of completing WEPT pages is also of particular relevance for the broader literature on scope (in)sensitivity in the valuation of public environmental goods (e.g., Baron & Greene, 1996; Desvouges et al., 1992; Hsee & Rottenstreich, 2004; Kahneman & Knetsch, 1992; Ojea & Loureiro, 2011). In hypothetical scenarios, participants often report being willing to make very similar contributions to the protection of an environmental good, irrespective of how large that good is (i.e., their behavior appears to be scope insensitive). For example, average willingness to pay was $80, $78, and $88 for the protection of 2,000, 20,000, and 200,000 water birds, respectively, in the study by Desvouges and colleagues (1992). However, the existing support for the scope insensitivity of pro-environmental behavior has been limited by the predominant use of hypothetical methods (Karlan & List, 2007). In addition, even when participants were required to incur actual costs to benefit the environment (Karlan & List, 2007), these costs have been exclusively financial in nature. The WEPT allows testing the generality of scope insensitivity by investigating consequential choices that involve behavioral costs in time and effort. As reviewed above, initial WEPT validation studies have already found that people are more likely to exert real effort for the environment when this effort produces larger environmental benefits (i.e., a €0.30 vs. €0.20 vs. €0.10 donation to an environmental organization; Lange & Dewitte, 2022). Critically however, this evidence for scope sensitivity

was found in a within-subject design, whereas the studies by Desvouges and colleagues (1992) or Karlan and List (2007) have relied on between-subjects designs. People's sensitivity to variations in environmental impact will likely be larger when variations are made salient through within-subject manipulation (Hsee & Rottenstreich, 2004; Berger & Wyss, 2021a). Without a reference for outcome comparison, people may be equally likely to exert effort for small and large environmental outcomes. This possibility illustrates the need for a comparative analysis of scope sensitivity in the WEPT in within-subject and between-subjects designs.

1.2 The present research

The present study aimed to examine the validity of the WEPT across countries, involving samples from the UK, the US, and South Africa. These countries were selected because they accounted for the three largest subpopulations on the online recruitment platform Prolific. Information about the validity of the WEPT in these subpopulations might thus be of particular interest for online research on pro-environmental behavior.

In a first step, we tested whether participants across countries take into account the behavioral costs and environmental impact of their behavior in the WEPT, that is, if our implementation of conflicting consequences was successful in the investigated populations. To this end, we manipulated the costs (i.e., the size of the number pages to be screened) and the impact (i.e., the amount of money being donated) of completing WEPT pages in a within-subject design. In line with Lange and Dewitte (2022), we hypothesized that the likelihood to complete a WEPT page would decrease with increasing costs (Hypothesis 1) and increase with increasing environmental impact (Hypothesis 2).

To shed further light on the scope sensitivity of pro-environmental behavior, we also analyzed the effect of between-subjects variations of environmental impact. Separate groups of participants completed the WEPT for either small or large pro-environmental donations

and we reasoned that, if people's efforts in the WEPT are directed at producing environmental impact, they should be intensified in the large-impact condition (Hypothesis 3a). In a second between-subjects analysis, we focused on participant's behavior on the very first WEPT trial (at which point it could not have been affected by the comparison with any other WEPT trials) and hypothesized that the completion of this trial would become more likely when linked to larger donation amounts (Hypothesis 3b).

Next, we analyzed whether WEPT performance can also serve as an indicator of people's general propensity to behave pro-environmentally. To this end, we correlated the number of completed WEPT pages to established indicators and measures of that propensity (an observation of pro-environmental donation behavior, a behavior-based measure of environmental attitude, and self-report measures of environmental concern, biospheric value orientation, and belief in climate change). Positive relationships between WEPT performance and these potential correlates (Hypothesis 4) would be consistent with the WEPT tapping into a general pro-environmental propensity.

Finally, we examined the relationship between participants' sensitivity to the environmental impact of their behavior and their general pro-environmental propensity. People whose behavior responds more strongly to environmental impact variations (e.g., to an increase of WEPT-related donations from 10 to 20 cents) place higher value on that impact and can thus be expected to care more about environmental protection in general. In line with this reasoning, Berger and Wyss (2021b) and Wyss et al. (2022) have found positive relationships between impact sensitivity in a different pro-environmental behavior paradigm and general pro-environmental propensity measures. Here, we sought to replicate these findings using environmental impact variations in the WEPT.

These hypothesis tests were complemented by an analysis of the WEPT's split-half reliability and a number of exploratory analyses. The study was approved by the local ethics

committee (G-2021-3748-R2(AMD)) and preregistered (https://osf.io/t3ehx/?view_only=db4f0417c31e4844ba3abca13e722fd2). We confirm that we have reported all measures, conditions, data exclusions, and how we determined our sample sizes. All materials, data, and analysis scripts can be found at https://osf.io/fhg3t/?view_only=63aee08ef99b4cc38b0aef8042228e50.

2. Methods

2.1 Participants

Our sample size rationale was based on the target sample size of the initial WEPT validation study (N = 184, Lange & Dewitte, 2022). This sample size allows detecting Spearman correlations (as examined for the test of Hypothesis 4) of r = .30 with 95% statistical power (given a corrected significance level of α = .01, one-tailed test). In contrast to the initial study, the present study included two versions of the WEPT (small impact, large impact) and three separate study populations (UK, US, South Africa). To ensure that power was kept high in each of the resulting 3 × 2 = 6 cells, we multiplied our sample size by six, yielding a target sample size of N = 1104 (or n = 368 from each study population). When data were pooled across cells, this sample size allowed detecting small effects of r = .10 or d = 0.20 with 95% statistical power (α = .05, one-tailed tests).

Participants were recruited from the online data-collection platform Prolific. We created three versions of the study, advertised to Prolific users who named the UK, the US, or South Africa as their current country of residence, respectively. In all three cases, the study was advertised to participants who indicated to be fluent in English and who had not participated in any prior WEPT studies from our group. Effective populations were about 40,000 (UK), 47,000 (US), and 11,000 (South Africa). Sampling on Prolific continued until the target sample size was reached. The study was advertised for a payment of £2.00. This amount was retrospectively adjusted to €3.64 in the South African sample due to the average

study completion time being much higher than originally expected. We considered this payment increase to be fair (as our original time estimate was based on our experiences in UK samples) and to be unlikely to influence our results (as there was no way for participants to anticipate the increase).

Participants who took longer than 58 minutes to complete the study were timed out and did not count towards the sample size in Prolific (i.e., their spot was opened again and another participant could complete the study). As preregistered, we kept the data from timed out participants, which resulted in the target sample size being exceeded by two (UK), six (US), and 53 (South Africa) participants, respectively. From the total sample of $N = 1175$, all participants completed the WEPT. Three participants had missing data for climate change beliefs and five participants did not make a donation decision (due to an error, the respective questions had not been made mandatory). As preregistered, data from those participants were excluded from all analyses involving the affected variables, but included in all other analyses. Sample characteristics are displayed in Table 1.

Table 1

Sociodemographic characteristics and key study variables as a function of subsample

	total sample (N = 1175)	UK (n = 370)	US (n = 374)	SA (n = 421)
WEPT	M = 6.42, SD = 4.86	M = 5.75, SD = 4.30	M = 4.34, SD = 4.13	M = 8.87, SD = 4.86
GEB	M = -0.29, SD = 0.73, Rel = .75	M = 0.02, SD = 0.73, Rel = .78	M = -0.44, SD = 0.74, Rel = .78	M = -0.43, SD = 0.62, Rel = .65
NEP	M = 3.70, SD = 0.59, α = .82	M = 3.80, SD = 0.57, α = .84	M = 3.83, SD = 0.67, α = .88	M = 3.50, SD = 0.48, α = .68
egoistic values	M = 3.67, SD = 1.65, α = .83	M = 2.99, SD = 1.33, α = .75	M = 2.98, SD = 1.52, α = .79	M = 4.87, SD = 1.28, α = .75
altruistic values	M = 5.61, SD = 1.25, α = .79	M = 5.40, SD = 1.06, α = .70	M = 5.26, SD = 1.40, α = .78	M = 6.12, SD = 1.09, α = .81
biospheric values	M = 5.30, SD = 1.46, α = .91	M = 5.14, SD = 1.28, α = .88	M = 4.85, SD = 1.64, α = .92	M = 5.86, SD = 1.24, α = .89
climate change beliefs	M = 4.47, SD = 0.71, α = .88	M = 4.47, SD = 0.67, α = .89	M = 4.47, SD = 0.79, α = .91	M = 4.48, SD = 0.66, α = .82
SRA	M = 2.89, SD = 0.75, α = .92	M = 2.82, SD = 0.63, α = .89	M = 2.55, SD = 0.67, α = .91	M = 3.26, SD = 0.75, α = .92
SCS	M = 3.23, SD = 0.71, α = .86	M = 3.13, SD = 0.68, α = .86	M = 3.18, SD = 0.77, α = .89	M = 3.36, SD = 0.67, α = .82
age (years)	M = 32.65, SD = 12.38	M = 36.62, SD = 13.39	M = 34.11, SD = 13.56	M = 27.86, SD = 8.06
gender	68% female, 30% male, 1% pts	77% female, 21% male, 1% pts	67% female, 30% male, 2% pts	62% female, 37% male, 1% pts
employment status	22% student, 41% full-time, 16% part-time, 11% unemployed, 3% retired, 6% pts	17% student, 45% full-time, 19% part-time, 7% unemployed, 5% retired, 7% pts	19% student, 41% full-time, 17% part-time, 12% unemployed, 5% retired, 5% pts	29% student, 36% full-time, 13% part-time, 13% unemployed, 1% retired, 6% pts
English native speakers	68%	87%	85%	35%
bonus donation	61%	77%	57%	51%

Note. WEPT = Work for Environmental Protection Task, GEB = General Ecological Behavior scale, NEP = New Environmental Paradigm, SRA = Self-reported Altruism scale, SCS = Self-control Scale. pts = prefer to self-describe (gender: non-binary: n = 8, agender: n = 1, gender-fluid: n =1, sissy: n = 1, two spirit: n = 1; employment status: self-employed/business owner/freelancer: n = 40, homemaker/stay-at-home parent: n = 7, disabled: n = 8, multiple: n = 9)

2.2 Procedure

The study was launched at 5:00 p.m. local time (CST was chosen for the US sample) on a Tuesday in February 2022 and all participants completed the study on that same evening. Having provided informed consent, participants first completed the WEPT. To introduce them to the task contingencies, we presented them with 18 two-digit numbers and the instruction to identify all those numbers that consist of an even first digit and an odd second digit. Once all numbers had been correctly identified (with automatic corrective feedback, if required), participants received "the option to continue with the number-identification task for a little longer." They were informed (truthfully) that we would make a donation to "an environmental non-profit organization working on the preservation of tropical forests" for each page of numbers that they complete correctly. In the original WEPT validation study (Lange & Dewitte, 2022), the German organization OroVerde was named as a receiving organization. In order to both replicate this study and to examine generalizability across organizations, we randomly allocated participants to either OroVerde or to a more familiar organization (WWF) in the present study. Both organizations were described in the exact same terms. Participants were explicitly informed that there was a maximum of 15 pages, that doing the task was "completely voluntary", and that it was "up to [them] to decide how much time and effort [they] want to invest in the task." They were then presented with the first choice trial, asking them whether they want to check X numbers for a donation of Y pounds for the protection of the rainforest. If they agreed, they were presented with a page containing X numbers to screen without time pressure or corrective feedback. Participants were encouraged to work thoroughly as we would "only count pages that are at least 90% correct". In line with previous WEPT studies (Lange & Dewitte, 2022; Lange & Truyens, 2022), we did not apply such an accuracy-based criterion in our confirmatory analyses, but instead required participants to spend a minimum amount of time working on a WEPT page (see

Section 2.3). An exploratory robustness analysis using accuracy-filtered WEPT performance can be found in Section 3.2.

Across the 15 trials, the number of numbers to be checked (Page Size) and the amount of money to be donated (Relative Impact) was varied. Pages contained 40, 80, 120, 160, or 200 numbers and page completion lead to a donation of a relatively small, medium, or large amount of money. Combination of the factors Page Size and Relative Impact yielded 15 distinct trial types, displayed in random order.

The absolute size of the relatively small, medium, and large amount of money was manipulated between-subjects (Absolute Impact). In the large-impact condition, WEPT page completion generated donations of 10, 20, or 30 pence. In the small-impact condition, WEPT page completion generated donations of 1, 2, or 3 pence. Participants were randomly allocated to either the large-impact or the small-impact condition.

After the last WEPT trial, participants completed a set of self-report scales (see Table 1 for descriptive statistics and reliabilities) that yield established measures related to people's general propensity to protect the environment. These scales (presented in the same fixed order for all participants) included 1) a 50-item version of the General Ecological Behavior scale (GEB, Kaiser & Wilson, 2004) as an environmental attitude measure based on self-reported pro-environmental behaviors, 2) the New Environmental Paradigm Scale (NEP, Dunlap et al., 2000) as a measure of environmental concern, 3) the value instrument reported by de Groot and Steg (2010), which allows to derive, among others, a measure of biospheric value orientation, and 4) a three-item measure of belief in climate change (Heath & Gifford, 2006). The first three measures were included as they have also been included in the original WEPT validation study (Lange & Dewitte, 2022) and the climate change belief scale was chosen in line with Berger and Wyss (2021b) who analyzed it as a correlate of environmental impact sensitivity. In line with Lange and Dewitte (2022), we also obtained another behavioral

indicator of people's pro-environmental propensity (and thus a potential correlate of WEPT performance) by providing participants with a second opportunity to engage in actual pro-environmental behavior. Participants were informed that they received a bonus payment of 20 pence, that they could either keep this money, donate it OroVerde, or donate it to the WWF, and that we would add 50% to every donation that they made in this donation task. As with the WEPT, all this donation-related information was factual and a total of £1095.76 was donated after the study. All monetary amounts were given in pound sterling (as this is currency of payment handled by Prolific), but on the informed consent form, participants were informed that "1 pound (£1.00) corresponds to about 1.34 US dollar or 21 South African rand." Finally, participants completed the Self-reported Altruism scale (SRA, Rushton et al., 1981) and the Brief Self-control Scale (BSCS, Tangney et al., 2004) for exploratory purposes and provided demographic data. Participants were explicitly informed that the study "examines a new procedure on which you can decide how much effort you want to spend on a computer task in order to generate funds for a charitable organization" and that the "duration of the study depends on participants' choices".

2.3 Preregistered Analyses

All significance tests were one-sided against a significance level of $\alpha = .05$. For Hypothesis 4 ($\alpha = .05/5 = .01$) and Hypothesis 5 ($\alpha = .05/4 = .0125$), the level of significance was adjusted to account for the number of examined relationships. Logistic regression analyses involving within-subject factors were run using Generalized Estimating Equations with an exchangeable working correlation matrix.

Hypotheses 1 and 2 were tested using a logistic regression analysis with the continuous within-subject factors Page Size (40, 80, 120, 160, 200 numbers) and Relative Impact (small, medium, large) and Page Completion (no = 0, yes = 1) as outcome measure. A WEPT page was counted as completed when the amount of time spent on the page was not

more than two standard deviations below the sample mean for that page. The main effect of the between-subjects factor Absolute Impact (small, large) was added to this regression model in order to examine Hypothesis 3a. To test Hypothesis 3b, we created a new between-subjects factor (Donation Amount) with six levels by combining the factors Relative Impact and Absolute Impact. This factor indicated the donation amount on the very first WEPT trial (1, 2, 3, 10, 20, 30 pence) and its effect on the completion of the first WEPT page was analyzed using a logistic regression analysis.

To test Hypothesis 4, Spearman correlations were calculated between the total number of completed WEPT pages and five potential correlates related to the propensity to behave pro-environmentally: 1) participants' GEB scores in logits (computed according to established procedures, Taube et al., 2018), 2) participants' average score on the NEP, 3) participants' average score on the biospheric value orientation subscale, 4) participants' average score on the measure of belief in climate change, and 5) whether participants donated (1) or not (0) their bonus payment to one of the two environmental organizations. Absolute Impact condition was added as covariate. Positive correlations were considered evidence for Hypothesis 4. To examine Hypothesis 5, four logistic regression analyses were run with the continuous within-subject factor Relative Impact, Page Size and Absolute Impact as covariates, and either 1) GEB scores, 2) NEP scores, 3) biospheric value scores, or 4) climate change belief scores. Significant positive regression coefficients for the interaction between Relative Impact and the self-report measures (indicating that the effect of Relative Impact is more pronounced for people who score high on the self-report measures) were interpreted as evidence for Hypothesis 5.

Finally, we used the RELEX tool developed by Steinke and Kopp (2020) to estimate the split-half reliability of the WEPT. We sampled reliability estimates across 1,000 iterations. Median Spearman-Brown corrected split-half parallel reliability coefficients larger than .90

were interpreted as indication of very good reliability. All preregistered analyses were run on the combined three-country sample and repeated for each of three individual countries separately.

3. Results

Overall, participants completed a total of 7485 WEPT pages or 6.42 pages on average (*SD* = 4.86). Two participants started to complete one additional page each, but those two pages were counted as non-completed because participants spent less than the preregistered minimum of time on the page. The average time participants spent on a completed page increased with page size (40 numbers: 50 s, 80 numbers: 98 s, 120 numbers: 150 s, 160 numbers: 196 s, 200 numbers: 246 s). WEPT pages were completed at an average accuracy (i.e., percentage of identified target numbers) of 93% (*SD* = 11%). The number of completed WEPT pages was highest in the South African subsample and lowest in the US subsample (Table 1).

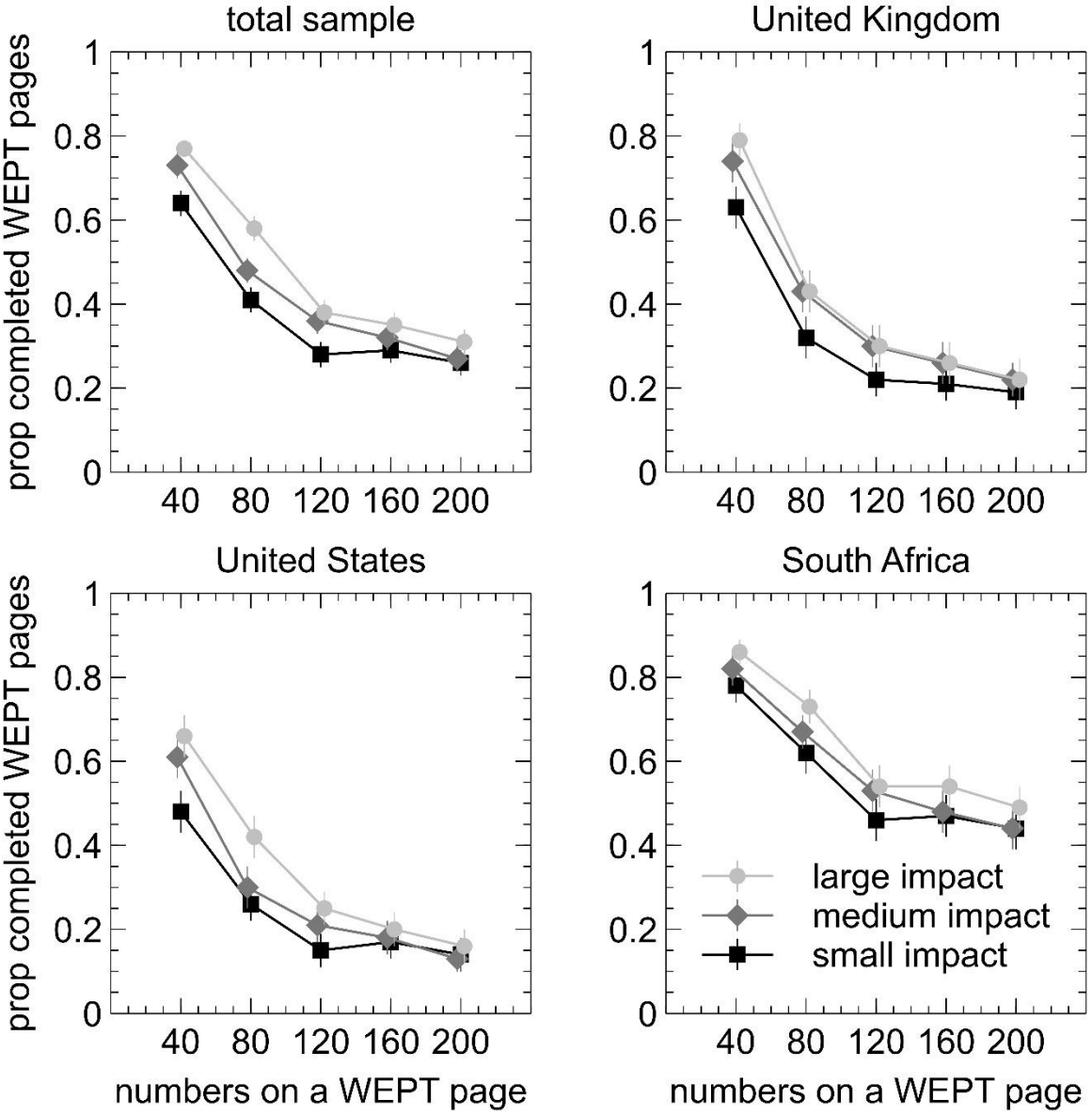
3.1 Confirmatory Analyses

The odds of completing a WEPT page decreased with increasing page size, b = -0.45, 95% CI = [-0.48, -0.42], *Exp(b)*= 0.64, Wald(1) = 835.76, $p < .001$, that is, with every additional 40 numbers to be screened, the odds of completing the page decreased by 36%. In contrast, the odds of completing a WEPT page increased by 26% per level of relative environmental impact, b = 0.23, 95% CI = [0.20, 0.26], *Exp(b)* = 1.26, Wald(1) = 204.43, $p <$ .001 (see Figure 1). Both effects were significant in all three countries, but smaller in the South African sample (Page Size: b = -0.39, 95% CI = [-0.43, -0.35]; Relative Impact: b = 0.18, 95% CI = [0.13, 0.23]), than in the UK sample (Page Size: b = -0.56, 95% CI = [-0.62, -0.49]; Relative Impact: b = 0.32, 95% CI = [0.26, 0.39]) or the US sample (Page Size: b = -0.56, 95% CI = [-0.64, -0.49]; Relative Impact: b = 0.28, 95% CI = [0.21, 0.35]). Hence,

Hypothesis 1 and Hypothesis 2 received support both from the total sample and in the country-specific analyses.

Figure 1

Within-subject analyses of cost and impact sensitivity on the Work for Environmental Protection Task (WEPT)

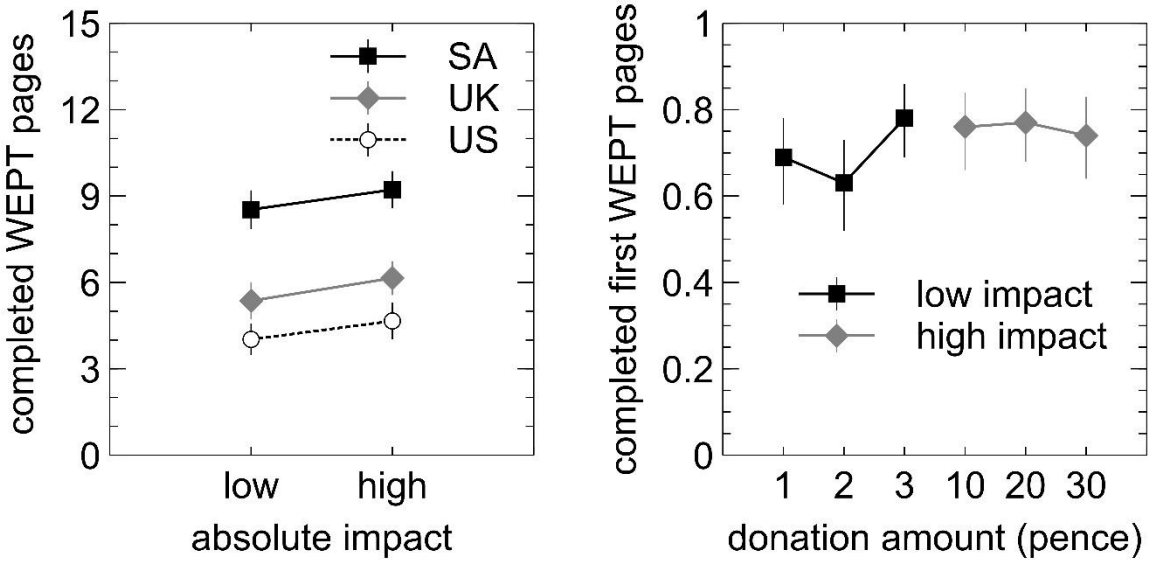


In support of Hypothesis 3a, participants who worked for 10, 20, or 30 pence completed more WEPT pages (M = 6.78, SD = 4.81) than participants who worked for 1, 2, or 3 pence (M = 6.09, SD = 4.88), b = 0.19, 95% CI = [0.04, 0.34], *Exp(b)*= 1.25, Wald(1) = 5.89, p = .008 (one-sided). This effect of Absolute Impact was of similar size in the UK, b = 0.22, the US, b = 0.21, and South Africa, b = 0.19 (see also Figure 2), but it only reached statistical significance in the UK subsample, p = .038 (one-sided). While the effect was significant for participants completing the WEPT to the benefit of the WWF, b = 0.27, 95% CI [0.06, 0.48], but not for participants completing the WEPT to the benefit of OroVerde, b = 0.11, 95% CI [-0.11, 0.33], the large overlap of confidence intervals indicates that the effect size did not differ significantly between organizations.

On the very first WEPT trial, between-subject variation in Donation Amount did not significantly affect pro-environmental effort, b = 0.01, *Exp(b)*= 1.01, 95% CI = [1.00, 1.02], Wald(1) = 2.00, p = .079 (one-sided). The effect was not significant in any of the country-specific samples, all $b < 0.02$, all $p > .093$, or for any of the two environmental organizations, both $b < 0.02$, all $p > .145$. The effect was significant in the low-impact condition (i.e., when Donation Amount ranged from 1 to 3 pence), b = 0.24, 95% CI [0.01, 0.46], but WEPT completion likelihood did not monotonously increase with donation amount (Figure 2). In the high-impact condition, between-subject variation from 10 to 30 pence did not significantly affect WEPT performance, b = 0.00, 95% CI [-0.02, 0.03]. Hence, we did not find conclusive support for Hypothesis 3b.

Figure 2

Between-subjects analyses of impact sensitivity in the WEPT



Note. Left: The number of completed WEPT pages as a function of study subsample and Absolute Impact condition (low: 1, 2, and 3 pence; high: 10, 20, and 30 pence). Right: The proportion of participants completing the first WEPT trial as a function of Donation Amount. Vertical bars are 95% confidence intervals.

Across the entire sample, the number of completed WEPT pages was positively related to GEB person parameters, $r = .06$, $p = .025$, 95% CI [.00, .11], but this partial correlation (controlled for Absolute Impact) was not significant at the adjusted significance level of $\alpha = .01$. It was significant when study subsample (UK, US, SA) was added as an additional control variable, $r = .09$, $p = .001$, 95% CI [.03, .15] (not preregistered). The correlation was significant ($p < .01$) in the US subsample $r = .15$, 95% CI [.05, .25], in people completing the WEPT for OroVerde, $r = .13$, 95% CI [.04, .20], and in the high Absolute Impact version of the procedure, $r = .13$, 95% CI [.05, .21].

Overall, we did not find a positive correlation between WEPT performance and NEP scores, $r = -.07$, $p = .990$, 95% CI [-.13, -.01]. The correlation flipped sign, but did not

become significant when additionally adjusted for study subsample, $r = .05$, $p = .052$, 95% CI [-.01, .11] (not preregistered). The strongest positive correlation appeared in the UK subsample, $r = .10$, 95% CI [.00, .20], but did not reach significance either, $p = .028$ (one-sided).

A significant, medium-sized correlation emerged between WEPT performance and biospheric values, $r = .26$, $p < .001$, 95% CI, [.21, .31] and was attenuated by adding study subsample as an additional control variable, $r = .16$, $p < .001$, 95% CI, [.10, .21] (not preregistered). This correlation was consistently positive and significant across all subsamples.

Climate change beliefs were not significantly related to the number of completed WEPT pages, neither in the total sample, $r = .03$, $p = .194$, 95% CI, [-.03, .08], nor in any of the subsamples. Adjusting for study subsample did not markedly change this result, $r = .04$, $p = .103$, 95% CI, [-.02, .09] (not preregistered).

Finally, we found a significant correlation between WEPT performance and participants' donations to an environmental organization. This correlation was significant in the total sample, $r = .09$, $p = .002$, 95% CI, [.03, .14], stronger when adjusted for study subsample, $r = .14$, $p < .001$, 95% CI, [.08, .19] (not preregistered), and significant in the UK and US subsample.

In sum, these correlation analyses provide some support for convergence between WEPT performance and conceptually related measures (Hypothesis 4). This support was strongest for the US sample (where the strongest correlation with donation behavior and the only significant correlation with GEB parameters emerged), and weakest for the South African sample. Zero-order correlations between WEPT performance, conceptually related measures, and exploratory measures are displayed in Table 2.

Table 2

Zero-order Spearman correlations between main study variables and exploratory measures

			1	2	3	4	5	6	7	8	9	10	11	12	13
1	WEPT		**\ all**	**.06**	**-.07**	**.30**	**.26**	**.26**	**.02**	**.09**	**.26**	**.04**	**.06**	**-.05**	**.06**
2	GEB	UK	.07												
		US	.16	**\ all**	**.28**	**-.15**	**.13**	**.29**	**.26**	**.18**	**.29**	**.11**	**.10**	**.11**	**.00**
		SA	.07												
3	NEP	UK	.10	.33											
		US	.05	.30	**\ all**	**-.33**	**.13**	**.20**	**.53**	**.23**	**-.10**	**-.01**	**.16**	**.12**	**-.08**
		SA	-.02	.15											
4	egoistic values	UK	.11	-.15	-.22										
		US	.13	-.08	-.28	**\ all**	**.38**	**.35**	**-.07**	**-.11**	**.29**	**.09**	**-.10**	**-.30**	**.15**
		SA	.14	-.04	-.15										
5	altruistic values	UK	.21	.30	.22	.14									
		US	.13	.27	.38	.07	**\ all**	**.72**	**.25**	**.03**	**.31**	**.17**	**.09**	**-.12**	**.08**
		SA	.12	.04	.15	.46									
6	biospheric values	UK	.19	.44	.33	.10	.65								
		US	.17	.41	.39	.10	.71	**\ all**	**.28**	**.10**	**.29**	**.18**	**.07**	**-.04**	**.06**
		SA	.12	.21	.21	.44	.69								
7	climate change beliefs	UK	.06	.34	.49	-.09	.31	.26							
		US	.07	.33	.68	-.14	.36	.39	**\ all**	**.11**	**.02**	**-.01**	**.08**	**-.13**	**.07**
		SA	.01	.15	.40	.03	.20	.26							
8	donation behavior	UK	.14	.09	.24	.04	.16	.21	.16						
		US	.26	.14	.16	.04	.17	.24	.13	**\ all**	**.02**	**-.03**	**.13**	**.13**	**-.06**
		SA	.04	.10	.21	-.10	-.01	.06	.06						
9	SRA	UK	.11	.31	.07	.06	.23	.19	.04	.10					
		US	.11	.36	.08	.09	.25	.25	.02	.10	**\ all**	**.16**	**-.06**	**.12**	**-.05**
		SA	.18	.34	-.09	.19	.18	.20	.05	-.02					
10	SCS	UK	.02	.15	.05	.06	.08	.12	-.06	.05	.09				
		US	-.09	.16	.03	-.11	.07	.08	-.02	-.07	.09	**\ all**	**.03**	**.17**	**-.09**
		SA	.06	.10	.04	.16	.21	.21	.07	.02	.23				
11	gender (female = 1, male = 0)	UK	.08	.05	.09	-.05	.13	.12	.16	.13	-.05	.06			
		US	.05	.08	.18	-.03	.21	.14	.09	.18	.08	.04	**\ all**	**.00**	**-.02**
		SA	.16	.06	.12	-.05	.11	.07	.03	.03	-.11	.05			
12	age	UK	.09	.07	.09	-.18	.01	.12	-.18	.05	.33	.20	-.05		
		US	-.13	.06	.05	-.26	-.03	.07	-.13	-.03	.22	.31	-.08	**\ all**	**-.58**
		SA	.18	.02	-.01	-.01	-.05	-.04	-.13	.23	.16	.16	-.02		
13	student	UK	-.02	.07	-.09	.17	.06	.02	.11	.04	-.24	-.12	-.09	-.57	
		US	.20	.03	-.08	.16	.06	.03	.00	.03	-.01	-.09	.05	-.56	\
		SA	-.10	.00	.01	-.02	.02	.03	.09	-.13	-.09	-.11	.01	-.56	

Note. Correlations within the country-specific subsamples are presented below the diagonal. Correlations within the total sample are presented in bold above the diagonal. Light grey highlights: significant at $\alpha = .05$ (two-sided); dark grey highlights: significant at $\alpha = .01$ (two-sided). WEPT = Work for Environmental Protection Task, GEB = General Ecological Behavior scale, NEP = New Environmental Paradigm, SRA = Self-reported Altruism scale, SCS = Self-control Scale

In support of Hypothesis 5, individual differences in NEP scores moderated the effect of Relative Impact on the likelihood to complete a WEPT page, $b = 0.10$, 95% CI = [0.05, 0.15], $Exp(b) = 1.05$, Wald(1) = 14.72, $p < .001$. The effect of Relative Impact on WEPT page completion increased across NEP quartiles, 1st quartile: $b = 0.17$, 95% CI [0.11, 0.23], 2nd quartile: $b = 0.17$, 95% CI [0.11, 0.23], 3rd quartile: $b = 0.27$, 95% CI [0.20, 0.34], 4th quartile: $b = 0.34$, 95% CI [0.27, 0.41]. This interaction was not affected by controlling for study subsample, $b = 0.10$, 95% CI = [0.05, 0.16] (not preregistered), and it did not reach significance in any of the country-specific subsamples. No significant interactions were observed with GEB parameters, biospheric values, and climate change beliefs.

Spearman-Brown corrected split-half parallel reliability across the 15 WEPT trials was $\rho_{SP} = .92$, with 95% of the sampled reliability coefficients lying between $\rho_{SP} = .87$ and $\rho_{SP} = .94$. In addition, reliability estimates were at least .90 in all of the three study subsamples (see Table 3).

Table 3

Overview of confirmatory analyses (regression/correlation coefficients [95% confidence intervals])

		Country			Organization		Absolute Impact	
	total	UK	US	SA	WWF	Oro Verde	high	low
sample size	1165	370	374	421	583	582	575	590
H1: within-subject effect Page Size	-0.45 [-0.48, -0.42]	-0.56 [-0.62, -0.49	-0.56 [-0.64, -0.49]	-0.39 [-0.43, -0.35]	-0.47 [-0.51, -0.42]	-0.43 [-0.48, -0.39]	-0.46 [-0.51, -0.42]	-0.44 [-0.48, -0.40]
H2: within-subject effect Relative Impact	0.23 [0.20, 0.26]	0.32 [0.26, 0.39]	0.28 [0.21, 0.35]	0.18 [0.13, 0.23]	0.21 [0.17, 0.26]	0.25 [0.20, 0.29]	0.23 [0.18, 0.27]	0.24 [0.19, 0.28]
H3a: between-subjects effect Absolute Impact	0.19 [0.04, 0.34]	0.22 [-0.02, 0.47]	0.21 [-0.06, 0.48]	0.19 [-0.06, 0.45]	0.27 [0.06, 0.48]	0.11 [-0.11, 0.33]	-	-
H3b: between-subjects effect Donation Amount	0.01 [0.00, 0.02]	0.02 [-0.01, 0.04]	0.01 [-0.01, 0.03]	0.00 [-0.03, 0.02]	0.01 [-0.01, 0.03]	0.00 [-0.01, 0.02]	0.00 [-0.02, 0.03]	0.24 [0.01, 0.46]
H4: partial correlations with...								
...GEB	.06 [.00, .11]	.08 [-.02, .18]	.15 [.05, .25]	.07 [-.03, .16]	.09 [.01, .17]	.13 [.04, .20]	.13 [.05, .21]	.08 [-.01, .16]
...NEP	-.07 [-.13, -.01]	.10 [.00, .20]	.05 [-.05, .16]	-.02 [-.11, .08]	.09 [.01, .17]	.01 [-.07, .09]	.07 [-.02, .15]	.04 [-.05, .12]
...biospheric values	.26 [.21, .31]	.20 [.10, .29]	.17 [.07, .27]	.12 [.03, .22]	.19 [.11, .27]	.15 [.06, .22]	.13 [.04, .20]	.21 [.13, .29]
...climate change belief	.03 [-.03, .08]	.06 [-.05, .16]	.08 [-.03, .18]	.01 [-.09, .10]	.05 [-.03, .13]	.05 [-.03, .13]	.05 [-.03, .13]	.06 [-.02, .14]
...donation	.09 [.03, .14]	.14 [.04, .24]	.26 [.16, .35]	.03 [-.06, .13]	.17 [.09, .25]	.12 [.04, .20]	.10 [.02, .18]	.17 [.09, .25]
H5: moderation of within-subject effect Relative Impact by...								
...GEB	0.04 [0.00, 0.09]	0.04 [-0.05, 0.13]	0.05 [-0.05, 0.14]	-0.02 [-0.10, 0.06]	0.08 [0.02, 0.14]	0.01 [-0.06, 0.07]	0.05 [-0.01, 0.12]	0.03 [-0.03, 0.10]
...NEP	0.10 [0.05, 0.15]	0.11 [0.00, 0.23]	0.10 [0.01, 0.20]	0.02 [-0.08, 0.13]	0.10 [0.03, 0.18]	0.10 [0.03, 0.17]	0.11 [0.04, 0.17]	0.09 [0.01, 0.17]
...biospheric values	-0.01 [-0.03, 0.01]	0.02 [-0.03, 0.07]	0.00 [-0.04, 0.04]	-0.01 [-0.05, 0.02]	0.01 [-0.02, 0.04]	-0.03 [-0.06, 0.01]	-0.01 [-0.04, 0.02]	-0.01 [-0.05, 0.01]
...climate change belief	0.03 [-0.01, 0.07]	0.02 [-0.06, 0.11][i]	0.04 [-0.05, 0.12]	0.03 [-0.03, 0.09]	0.01 [-0.05, 0.07]	0.05 [0.00, 0.10]	0.06 [0.00, 0.11]	0.00 [-0.05, 0.06]
Split-half reliability	.92 [.87, .94]	.90 [.83, .92]	.91 [.83, .93]	.92 [.87, .94]	.92 [.87, .94]	.93 [.88, .94]	.92 [.87, .94]	.93 [.87, .94]

Note. Tests of Hypotheses 4 and 5 were controlled for Absolute Impact, except when the Absolute Impact conditions were analyzed separately (rightmost columns). H = Hypothesis, GEB = General Ecological Behavior Scale, NEP = New Environmental Paradigm Scale. [i]computed with independent covariance matrix because estimation with the exchangeable matrix failed

3.2 Exploratory Analyses

In the following, we report the results of a number of exploratory analyses that may inform the interpretation of the confirmatory results reported above. First, inspection of Table 2 points to a number of additional potentially interesting relationships between WEPT performance and self-report measures. For example, the number of completed WEPT pages was positively related to participants' propensity to engage in prosocial behavior (as indexed by the SRA), r = .26, 95% CI, [.21, .31]. Accordingly, a larger number of completed WEPT pages related to stronger expressions of altruistic values as assessed by the value instrument by de Groot and Steg (2010), r = .26, 95% CI, [.20, .31]. Somewhat surprisingly, we found a correlation of similar size between WEPT performance and egoistic values, r = .30, 95% CI, [.25, .36]. Correlations were partly driven by between-sample differences (i.e., higher value scores and more completed WEPT pages in the South African subsample) and decreased when being examined in country-specific subsamples. Regressing WEPT page completion on all three types of values, controlled for study subsample and Absolute Impact, reveals that biospheric, b = 0.12, 95% CI = [0.05, 0.20], and egoistic values, b = 0.08, 95% CI = [0.02, 0.13], but not altruistic values, b = 0.04, 95% CI = [-0.06, 0.13], predicted a significant (α = .05) amount of unique WEPT variance.

Next, we focused on the effect of trial position on WEPT performance. Independent of Page Size and Relative Impact, participants were less likely to complete later WEPT pages than WEPT pages that occurred early in the trial sequence, b = -0.10, 95% CI = [-0.11, -0.09], *Exp(b)* = 0.90, Wald(1) = 540.53, p < .001, that is, the odds of a WEPT trial being completed decreased by 10% with every trial. The size of this effect was similar across countries (UK: b = -0.13, US: b = -0.10, SA: b = -0.11). This effect was not moderated by any of the variables related to the general propensity to protect the environment, nor by trait self-control (as assessed by the SCS), -0.01 < all bs < 0.02, all ps >.14.

Finally, we tested the robustness of our main analyses by repeating them using an accuracy-based criterion for considering a WEPT page as completed. Of the 7485 completed WEPT pages, 630 pages (or 8 %) were completed at an accuracy of less 80% (UK: 9 %, US: 8 %, SA: 8 %). The within-subject effects of Page Size (H1), b = -0.41, 95% CI = [-0.44, -0.38], and Relative Impact (H2), b = 0.23, 95% CI = [0.20, 0.26], as well as the between-subjects effect of Absolute Impact (H3a), b = 0.19, 95% CI = [0.04, 0.34], and Donation Amount (H3b), b = 0.00, 95% CI = [-0.01, 0.02], did not markedly change in size when requiring completed WEPT pages to be at least 80% correct. The same applies to the size of the partial correlations examined to test Hypothesis 4 (GEB: r = .06, 95% CI, [.00, .12], NEP: r = -.04, 95% CI, [-.10, .02], biospheric values: r = .24, 95% CI, [.19, .29], climate change beliefs, r = .03, 95% CI, [-.03, .09], donation: r = .09, 95% CI, [.03, .15]). Repeating the test of Hypothesis 5 revealed similar results as well: the interaction between NEP scores and Relative Impact remained significant, b = 0.10, 95% CI = [0.05, 0.15], all other interactions remained non-significant. Additional robustness checks can be found in the Supplementary Materials.

4. Discussion

The present findings support the validity of the WEPT in online samples from the UK, the US, and South Africa. WEPT efforts were revealed to be directed at the generation of environmental impact (as they were differentially attracted by increasing pro-environmental donations) and to be more than cheap talk (as they were differentially deterred by increasing effort requirements). This sensitivity to the consequences implemented in the WEPT environment closely replicates the findings from a Belgian population (Lange & Dewitte, 2022) and it was observed to be highly robust across study samples. The WEPT thus seems to involve an effective trade-off between valued environmental and individual consequences and

to be suited for the experimental analysis of pro-environmental behavior in the investigated populations.

WEPT performance was found to be sensitive not only to within-subject variation of relative environmental impact, but also to impact differences in a between-subjects design. Shifting the scope of WEPT-related donations from small (1, 2, 3 pence) to larger amounts (10, 20, 30 pence) exerted a small effect on pro-environmental behavior with an effect size that was highly similar across the three independent study samples. This between-subjects effect indicates that people do not only scale their efforts relative to an arbitrarily created anchor (i.e., their performance on earlier WEPT trials), but generally work harder for objectively larger environmental outcomes. In other words, the within-subject effect observed here and in the study by Lange and Dewitte (2022) does not seem to exclusively reflect "coherent arbitrariness" (Ariely et al., 2003), but also truly scope-sensitive behavioral valuations of an environmental consequence.

To our knowledge, this is the first evidence for the scope sensitivity of effortful pro-environmental behavior. This findings contrasts with several between-subjects studies that found contributions to environmental goods to be scope insensitive (Baron & Greene, 1996; Desvouges et al., 1992; Hsee & Rottenstreich, 2004; Kahneman & Knetsch, 1992; see also Ojea & Loureiro, 2011, for review). Diverging results might be due to the implementation of actual consequences, the focus on real effort as a behavioral cost, the specific range of impact magnitudes we selected, or the increased sensitivity of our experimental design (involving a large sample size and multiple WEPT trials). In support of the last possibility, we did not find a between-subjects effect of impact size when we examined participants' behavior on the very first WEPT trial (Hypothesis 3b). This first-trial analysis was less powerful (based on fewer trials and fewer participants per cell) than the analysis of Hypothesis 3a and it may have been affected by a ceiling effect (as WEPT completion likelihoods were particularly high on the

first trial). These differences may account for the lack of convergence between our tests of Hypothesis 3a and 3b.

In addition, we observed small-to-medium correlations between WEPT performance, pro-environmental donation behavior, and self-report measures related to the general propensity to engage in pro-environmental behavior. Some of these correlations (e.g., the cross-culturally stable correlation with biospheric value orientation) confirm earlier findings from a Belgian sample (Lange & Dewitte, 2022), but their size is not sufficient to support the construct validity of the WEPT as a standalone measure of individuals' propensity to behave pro-environmentally. Instead, we propose that the WEPT can be considered a behavioral indicator (that can be captured in a highly objective and reliable way) of that propensity. Behavior in the WEPT is a specific pro-environmental behavior and the correlation between specific pro-environmental behaviors (or between pro-environmental behaviors and verbal propensity indicators) is typically small-to-medium in size (Lange & Dewitte, 2021; Weigel & Newman, 1976). Nonetheless, shared variance across specific behaviors can be used to obtain useful propensity measures for individual difference research (Kaiser, 1998; Kaiser & Wilson, 2000; Weigel & Newman, 1976). Though we see the main use of the WEPT in facilitating the experimental analysis of pro-environmental behavior and its determinants, we think that the WEPT could also play a role in the behavioral assessment of individual differences in people's pro-environmental propensity. To this end, the task would probably need to be combined with other pro-environmental behavior paradigms (e.g., Berger & Wyss, 2021a; Klein et al., 2017; Lange et al., 2018), thus paralleling, for example, the multi-task assessment of individual differences in executive functioning (Miyake et al., 2000; Miyake & Friedman, 2012).

Not all of the correlations reported in the original WEPT validation study were replicated in the present study. Perhaps most notably, no correlation was found between

WEPT performance and NEP scores. This relationship might have been attenuated by differences in item understanding within and across the present samples, leading to low reliability of the NEP in our South African sample in particular. The use of language-related Prolific criteria might not be sufficient to ensure that instruments using complex questionnaire items (such as the NEP) retain their psychometric properties across populations. In contrast, it should be noted that the reliability and consequence sensitivity of WEPT performance was very similar across samples, which illustrates the potential use of the task for cross-cultural research.

In view of the poor reliability of the NEP, the interaction observed between NEP scores and within-subject impact variation should be interpreted with caution. WEPT performance of the participants who scored highest on the NEP was most sensitive to environmental impact (i.e., to the relative size of pro-environmental donations), but those participants also completed the lowest number of WEPT pages. This pattern and the lack of significant interactions between environmental impact and other propensity measures is in contrast with findings obtained on the Carbon Emission Task (Berger & Wyss, 2021b; Wyss et al., 2022). Those authors found peoples' monetary contributions to climate change mitigation to increase with the amount of avoided carbon emissions and people's sensitivity to emission amounts to depend on their climate change beliefs (Berger & Wyss, 2021b) and NEP scores (Wyss et al., 2022). Next to the questionable reliability of NEP scores in our study, a less direct connection of the WEPT to climate change mitigation or other methodological differences between the behavioral paradigms might account for this difference in results.

Several exploratory findings warrant additional discussion. While the validity and between-subjects scope sensitivity of the WEPT was supported in all three country-specific subsamples, participants from South Africa completed more WEPT pages than participants

from the UK or the US. Given the current evidence, it can only be speculated if this difference is due to South African participants placing higher value on WEPT-related environmental consequences, caring less about the associated cost, or interpreting the WEPT task instructions differently. It is also possible that Prolific reaches a different segment of the population in South Africa as compared to the UK or the US. We also found the likelihood for a WEPT page to be completed to decrease across trials. Although being transparently informed about the effort requirements of the task, people may discount these costs until they have actually experienced them. The fact that WEPT participants actually experience the behavioral costs of their choices might offer advantages and additional research opportunities in comparison to one-shot behavioral paradigms of pro-environmental behavior. In addition, we found WEPT performance to be positively correlated not only to biospheric and altruistic value orientations, but also to egoistic value orientations. This contrasts with other studies finding the likelihood to engage in specific pro-environmental behaviors to be negatively related or uncorrelated to egoistic value orientations (e.g., Lange et al., 2018; Steg et al., 2014). The divergent finding in our study might reflect that people with strong achievement goals score high on (some of) the egoistic value items and are motivated to complete a lot of pages of the number-identification task. Finally, while we have seen that the scope of WEPT-related donations affected absolute WEPT performance, validity support was largely independent of this task parameter as well as of the chosen receiving organization (see Table 3). Hence, our study not only confirms the validity of the WEPT for use in samples from other countries, but also the validity of alternative task variants. The validity support for the low-impact task variant might be of particular interest to some researchers as it points to the possibility to study consequential pro-environmental behavior in online studies while keeping donation costs low.

The use of English-speaking Prolific samples from the UK, US, and South Africa allowed to run very close replications of the same study in different countries. Residents from all selected countries were exposed to the exact same materials and procedures, at the same day and local time of the day. Yet this sampling method also comes with limitations. Prolific populations from countries that do not include English as one of their official languages are generally small and we were not able to expand our study to a more diverse set of countries. In addition, it was not possible to randomly draw participants from the pool of all Prolific users, but participants were sampled on a first-come, first-serve basis. These issues notwithstanding, we think that the provided information on the validity of the WEPT (and the administered self-report measures) will be of interest to many researchers in the field given the size and accessibility of the investigated populations.

Our findings should not be taken to suggest that a WEPT-based experiment will have universal external validity. The results of WEPT-based experiments can be expected to generalize to naturally occurring situations of environmental relevance only if 1) these situations involve the same trade-off between environmental and individual consequences and 2) the experimental manipulation modulates these consequences in the same way as it would in the naturally occurring situation of interest. Different classes of situations and pro-environmental behaviors will likely require different behavioral paradigms to be adequately studied under controlled conditions. The WEPT can be adapted to model other situations (e.g., by changing the donation recipient or donation contingency or by requiring other types of effort) and such adaptations will likely be necessary to capture the diversity of pro-environmental behaviors.

References

Ariely, D., Loewenstein, G., & Prelec, D. (2003). "Coherent arbitrariness": Stable demand curves without stable preferences. *The Quarterly Journal of Economics*, *118*(1), 73-106.

Baron, J., & Greene, J. (1996). Determinants of insensitivity to quantity in valuation of public goods: Contribution, warm glow, budget constraints, availability, and prominence. *Journal of Experimental Psychology: Applied*, *2*(2), 107-125.

Berger, S. & Wyss, A. M. (2021a). Measuring pro-environmental behavior using the carbon emission task. *Journal of Environmental Psychology*, 75, 101613.

Berger, S., & Wyss, A. M. (2021b). Climate change denial is associated with diminished sensitivity in internalizing environmental externalities. *Environmental Research Letters*, *16*, 074018.

Caruso, J. C. (2000). Reliability generalization of the NEO personality scales. *Educational and Psychological Measurement*, *60*(2), 236-254.

De Groot, J. I., & Steg, L. (2010). Relationships between value orientations, self-determined motivational types and pro-environmental behavioural intentions. *Journal of Environmental Psychology*, *30*(4), 368-378.

Desvouges, W., Johnson, F. R., Dunford, R. W., Boyle, K. J., Hudson, S. P., & Wilson, K. N. (1992). Measuring Nonuse Values Using Contingent Valuation: An Experimental Evaluation of Accuracy. Research Triangle Park, NC: RTI International.

Dunlap, R. E., Van Liere, K. D., Mertig, A. G., & Jones, R. E. (2000). New trends in measuring environmental attitudes: measuring endorsement of the new ecological paradigm: a revised NEP scale. *Journal of Social Issues*, *56*(3), 425-442.

643 Flake, J. K., & Fried, E. I. (2020). Measurement schmeasurement: Questionable measurement
644 practices and how to avoid them. *Advances in Methods and Practices in Psychological*
645 *Science*, *3*(4), 456-465.
646 Gifford, R. (2011). The dragons of inaction: psychological barriers that limit climate change
647 mitigation and adaptation. *American Psychologist*, *66*, 290-302.
648 Gifford, R. (2014). Environmental psychology matters. *Annual Review of Psychology*, *65*,
649 541-579.
650 Heath, Y., & Gifford, R. (2006). Free-market ideology and environmental degradation: The
651 case of belief in global climate change. *Environment and Behavior*, *38*(1), 48-71.
652 Hsee, C. K., & Rottenstreich, Y. (2004). Music, pandas, and muggers: on the affective
653 psychology of value. *Journal of Experimental Psychology: General*, *133*(1), 23-30.
654 Kahneman, D., & Knetsch, J. L. (1992). Valuing public goods: the purchase of moral
655 satisfaction. *Journal of Environmental Economics and Management*, *22*(1), 57-70.
656 Kaiser, F. G. (1998). A general measure of ecological behavior. *Journal of Applied Social*
657 *Psychology*, *28*(5), 395-422.
658 Kaiser, F. G. (2021). Climate change mitigation within the Campbell paradigm: doing the
659 right thing for a reason and against all odds. *Current Opinion in Behavioral Sciences*,
660 *42*, 70-75.
661 Kaiser, F. G., & Lange, F. (2021). Offsetting behavioral costs with personal attitude:
662 Identifying the psychological essence of an environmental attitude measure. *Journal of*
663 *Environmental Psychology*, *75*, 101619.
664 Kaiser, F. G., & Wilson, M. (2000). Assessing People's General Ecological Behavior: A
665 Cross-Cultural Measure. *Journal of Applied Social Psychology*, *30*(5), 952-978.

Kaiser, F. G., Merten, M., & Wetzel, E. (2018). How do we know we are measuring environmental attitude? Specific objectivity as the formal validation criterion for measures of latent attributes. *Journal of Environmental Psychology*, *55*, 139-146.

Kaiser, F. G., & Wilson, M. (2004). Goal-directed conservation behavior: The specific composition of a general performance. *Personality and Individual Differences*, *36*, 1531-1544.

Karlan, D., & List, J. A. (2007). Does price matter in charitable giving? Evidence from a large-scale natural field experiment. *American Economic Review*, *97*(5), 1774-1793.

Klein, S. A., & Hilbig, B. E. (2019). On the lack of real consequences in consumer choice research. *Experimental Psychology*, *66*, 68-76.

Klein, S. A., Hilbig, B. E., & Heck, D. W. (2017). Which is the greater good? A social dilemma paradigm disentangling environmentalism and cooperation. *Journal of Environmental Psychology*, *53*, 40-49.

Kopp, B., Lange, F., & Steinke, A. (2021). The reliability of the Wisconsin card sorting test in clinical practice. *Assessment, 28*(1), 248-263.

Lange, F. (2019). Are difficult-to-study populations too difficult to study in a reliable way?. *European Psychologist*, *25*, 41-50.

Lange, F. (2022). Behavioral paradigms for studying pro-environmental behavior: A systematic review. *Behavior Research Methods*. https://doi.org/10.3758/s13428-022-01825-4

Lange, F., & Dewitte, S. (2019). Measuring pro-environmental behavior: Review and recommendations. *Journal of Environmental Psychology*, *63*, 92-100.

Lange, F., & Dewitte, S. (2021). Test-retest reliability and construct validity of the Pro-Environmental Behavior Task. *Journal of Environmental Psychology*, *73*, 101550.

690 Lange, F., & Dewitte, S. (2022). The Work for Environmental Protection Task: A
691 consequential web-based procedure for studying pro-environmental behavior.
692 *Behavior Research Methods, 54*(1), 133-145.
693 Lange, F., Steinke, A., & Dewitte, S. (2018). The Pro-Environmental Behavior Task: A
694 laboratory measure of actual pro-environmental behavior. *Journal of Environmental*
695 *Psychology, 56*, 46-54.
696 Lange, F., & Truyens, M. (2022). Mixed evidence for the effect of virtual nature exposure on
697 effortful pro-environmental behavior. *Journal of Environmental Psychology*, *81*,
698 101803.
699 Lewandowski Jr, G. W., & Strohmetz, D. B. (2009). Actions can speak as loud as words:
700 Measuring behavior in psychological science. *Social and Personality Psychology*
701 *Compass, 3*(6), 992-1002.
702 Linder, N., Rosenthal, S., Sörqvist, P., & Barthel, S. (2021). Internal and external factors'
703 influence on recycling: Insights from a laboratory experiment with observed behavior.
704 *Frontiers in Psychology, 12*, 699410.
705 Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences
706 in executive functions: Four general conclusions. *Current Directions in Psychological*
707 *Science, 21*(1), 8-14.
708 Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D.
709 (2000). The unity and diversity of executive functions and their contributions to
710 complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1),
711 49-100.
712 Ojea, E., & Loureiro, M. L. (2011). Identifying the scope effect on a meta-analysis of
713 biodiversity valuation studies. *Resource and Energy Economics, 33*(3), 706-724.

Rushton, J. P., Chrisjohn, R. D., & Fekken, G. C. (1981). The altruistic personality and the self-report altruism scale. *Personality and Individual Differences*, *2*(4), 293-302.

Schmuckler, M. A. (2001). What is ecological validity? A dimensional analysis. *Infancy*, *2*, 419-436.

Steg, L., & Vlek, C. (2009). Encouraging pro-environmental behaviour: An integrative review and research agenda. *Journal of Environmental Psychology*, *29*, 309-317.

Steg, L., Perlaviciute, G., Van der Werff, E., & Lurvink, J. (2014). The significance of hedonic values for environmentally relevant attitudes, preferences, and actions. *Environment and Behavior*, *46*(2), 163-192.

Steinke, A., & Kopp, B. (2020). RELEX: An Excel-based software tool for sampling split-half reliability coefficients. *Methods in Psychology*, *2*, 100023.

Stern, P. C. (2000). New environmental theories: toward a coherent theory of environmentally significant behavior. *Journal of Social Issues*, *56*, 407-424.

Tam, K. P., & Milfont, T. L. (2020). Towards cross-cultural environmental psychology: A state-of-the-art review and recommendations. *Journal of Environmental Psychology*, *71*, 101474.

Tangney, J. P., Baumeister, R. F., & Boone, A. L. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. Journal of Personality, 72, 271–322.

Taube, O., & Vetter, M. (2019). How green defaults promote environmentally friendly decisions: Attitude-conditional default acceptance but attitude-unconditional effects on actual choices. *Journal of Applied Social Psychology*, *49*(11), 721-732.

Taube, O., Kibbe, A., Vetter, M., Adler, M., & Kaiser, F.G. (2018). Applying the Campbell Paradigm to sustainable travel behavior: Compensatory effects of environmental

attitude and the transportation environment. *Transportation Research Part F: Traffic Psychology and Behaviour*, *56*, 392-407.

Weigel, R.H., & Newman, L.S. (1976). Increasing attitude-behavior correspondence by broadening the scope of the behavioral measure. *Journal of Personality and Social Psychology*, *33*, 793-802.

Wille, F., & Lange, F. (2022). Potential Contributions of Behavior Analysis to Research on Pro-environmental Behavior. *Frontiers in Psychology*, *13*, 685621.

Wyss, A. M., Knoch, D., & Berger, S. (2022). When and how pro-environmental attitudes turn into behavior: The role of costs, benefits, and self-control. *Journal of Environmental Psychology*, *79*, 101748.

Zhang, Y., Jiang, J., Sun, Y., Gu, D., & Jiang, W. (2021). Engagement in cause-related marketing reduces pro-environmental behaviors. *Environment and Behavior*, *53*(10), 1047-1069.