The Work for Environmental Protection Task: A consequential web-based procedure for studying pro-environmental behavior

Florian Lange & Siegfried Dewitte

Behavioral Engineering Research Group, KU Leuven, Leuven, Belgium

Correspondence:

Florian Lange

BEE – Behavioral Engineering Research Group

KU Leuven

Naamsestraat 69

3000 Leuven

Belgium

florian.lange@kuleuven.be

ACCEPTED MANUSCRIPT

Acknowledgements

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

Declarations of interest: none.

This version of the manuscript has been accepted for publication in *Behavior Research Methods* on 5 May 2021.

Abstract

Some human behaviors have serious societal consequences, but these consequences tend to be neglected in online research on societally relevant behaviors. For example, human activities contribute to climate change and biodiversity loss, but pro-environmental behavior is often studied using inconsequential self-reports and hypothetical scenarios. Such measures can easily be administered online, but suffer from severe validity problems. To address these problems, we developed a multi-trial web-based procedure for the study of consequential pro-environmental behavior. On the Work for Environmental Protection Task (WEPT), participants can choose to exert voluntary extra efforts screening numerical stimuli in exchange for donations to an environmental organization. They thus have the opportunity to produce actual environmental benefits at actual behavioral costs (i.e., to show actual pro-environmental behavior). In a preregistered validation study (N = 209), we found WEPT performance to systematically vary with these consequences, that is, the implemented costs and benefits were large enough for participants to effectively take them into account. In addition, aggregated WEPT performance was found to be highly reliable and to be correlated to self-reports and objective observations of other pro-environmental behaviors and conceptually related measures. These findings support the validity of the WEPT as an online procedure for the study of actual pro-environmental behavior. We discuss how the WEPT can advance the experimental analysis of pro-environmental behavior, help to address problems of common-method variance in individual difference research, and be adapted for the consequential study of other societally relevant behaviors.

Keywords: pro-environmental behavior; online assessment; performance-based assessment; consequences; validity; reliability; Work for Environmental Protection Task

ACCEPTED MANUSCRIPT

The Work for Environmental Protection Task: A consequential web-based procedure for studying pro-environmental behavior

Consequences give meaning to human behavior. The same simple movement can classify as prosocial behavior (i.e., behavior that benefits others; Hinde & Groebel, 1991) or aggressive behavior (i.e., behavior directed at other people's harm; Krahé, 2020), depending on the consequences it produces. For example, a mouse click may confirm a donation to charity or send an insulting email, and the raising of an arm may assist or hurt another person. Similarly, behavior is labelled pro-environmental when it provides relative benefits to the natural environment (Steg & Vlek, 2009) or health behavior when it relates to positive health outcomes (Glanz, Rimer, & Viswanath, 2008). These examples also illustrate the often substantial societal relevance of such consequences, and thus of research into the behaviors that produce them. Reliable knowledge from the behavioral sciences is required to support human responses to social and environmental issues, from climate change (Lehman & Geller, 2004; Swim, Clayton, & Howard, 2011; van der Linden, Maibach, & Leiserowitz, 2015) to public health crises (IJzerman et al., 2020; van Bavel et al., 2020).

To obtain this knowledge, behavioral scientists across domains and disciplines examine the functional determinants and underlying mechanisms of a multitude of human behaviors. However, for practical reasons, many of them do not directly study the societally relevant behaviors to which they would ultimately like to generalize their results (Klein & Hilbig, 2019; Lange & Dewitte, 2019a). While the advantages of studying consequential behaviors in the field are widely acknowledged, the popularity of field studies seems limited by their cost and lack of experimental control (Cialdini, 2009; Lange & Dewitte, 2019a; Maner, 2016; Mortensen & Cialdini, 2010). For these reasons, many researchers aim to arrange laboratory or online situations in a way that allows studying human behavior in a more controlled and less expensive way. Unfortunately, however, the defining element of

ACCEPTED MANUSCRIPT

societally relevant behaviors (i.e., their consequences) tends to get lost when transferring behaviors from the field to more controlled settings.

Studying societally relevant behavior in controlled settings

When studying societally relevant behaviors in the lab or online, many researchers rely on participants' responses to self-report items or hypothetical scenarios (Baumeister, Vohs, & Funder, 2007; Klein & Hilbig, 2019; Sassenberg & Dietrich, 2019). Due to their lack of consequences, such responses might not be very informative for understanding everyday behaviors of societal relevance (Herziger & Hoelzl, 2017; Klein & Hilbig, 2019; Lewandowski & Strohmetz, 2009; Morales, Amir, & Lee, 2013). The external validity of lab and online results (e.g., whether the effect of an experimental manipulation on a self-report proxy of behavior generalizes to situations of actual societal relevance) depends on the degree to which arranged situations and everyday situations involve the same contingencies (or "critical theoretical parameters"; Schmuckler, 2001, p. 432). In many everyday situations, producing societal benefits (e.g., through donating blood, reducing private car use, or wearing face masks) relates to behavioral costs in terms of, for example, money, time, or comfort. When responding to items like "I am willing to donate blood," "I intend to reduce my car use," or "I would wear a face mask," participants do not have to incur such costs when choosing "rather agree" over "fully disagree." It thus seems questionable to generalize any results obtained with such items to everyday societally relevant behaviors (Lange & Iwasaki, 2020). Contingency differences between engaging in societally relevant behavior and reporting on this engagement may also partly explain why self-reports of behavior (or its proposed antecedents) often do not converge with observations of actual behavior (e.g., Kormos & Gifford, 2014).

The external validity problems related to the use of self-reports and hypothetical scenarios can be addressed by contriving situations or designing tasks that involve actual

ACCEPTED MANUSCRIPT

consequences. While behavioral scientists have found numerous creative solutions to create such consequential situations in the laboratory, these solutions are often difficult to implement in online study designs. The possibility to run behavioral studies online comes with a lot of methodological advantages (Birnbaum, 2004; Gosling & Mason, 2015; Hartshorne, de Leeuw, Goodman, Jennings, & O'Donnell, 2019), yet relating online behaviors to societally relevant consequences may not always be straightforward. In the following, we will illustrate this problem (and one way to address it) by drawing on research on pro-environmental behavior.

Pro-environmental behavior researchers are interested in studying behavior that "harms the environment as little as possible, or even benefits the environment" (Steg & Vlek, 2009, p. 309). In the laboratory, pro-environmental behavior has been studied, for example, by observing whether participants turned off the laboratory lights (Murtagh, Gatersleben, Cowen, & Uzzell, 2015) or the lab computer screen (Moussaoui, Desrichard, & Milfont, 2020) when leaving the room. Other researchers have measured the amount of water participants used while testing the quality of a towel (Liu, Geng, Ye, & Zhou, 2019) or the amount of scrap paper they used to complete a writing task (Cornelissen, Pandelaere, Warlop, & Dewitte, 2008). Recently, Lange, Steinke, and Dewitte (2018) have also developed the Pro-Environmental Behavior Task (PEBT) as a multi-trial procedure involving repeated choices between an environmentally harmful and an environmentally friendly response option. Choosing the environmentally friendly option prolonged the time participants had to wait in the laboratory, but choosing the environmentally harmful option turned on an additional set of lights and thus wasted energy. Environmental consequences like these (i.e., the waste of energy, water, or scrap paper) are difficult to implement online as are the behavioral costs that participants have to incur when doing something to benefit the environment. In a task similar to the one by Lange and colleagues (2018), Taube, Kibbe, Vetter, Adler, and Kaiser (2018)

ACCEPTED MANUSCRIPT

asked participants in an online study to choose between an environmentally friendly and an environmentally harmful option, with the latter involving shorter waiting times than the former. Yet when participants complete such tasks online, they can easily forgo such waiting time consequences (e.g., by pursuing enjoyable activities while waiting), which may limit generalization to everyday behaviors that do not involve this possibility (Lange & Dewitte, 2019a). With the present research, we aimed to address these limitations by developing and validating an online task for the study of consequential pro-environmental behavior.

The Work for Environmental Protection Task

To this end, we followed Lange and colleagues (2018) in seeking to create choice situations in which individual consequences and environmental consequences conflict. This conflict is characteristic of many everyday choices of environmental relevance (Gifford, 2011; Steg, 2015): benefits for the environment (e.g., in terms of saved resources, species, or CO_2 emissions) often come at a cost for the individual (e.g., in terms of money, time, or comfort). We implemented this conflict by offering participants to exert voluntary efforts on a monotonous number-identification task. This approach is related to the use of real-effort tasks in experimental economics (Charness, Gneezy, & Henderson, 2018; Farrelly & Bennett, 2018; Farrelly, Moan, White, & Young, 2015). Importantly, for every page of numbers that participants completed, the researchers made a small donation to an environmental organization working for the protection of tropical forests. This way, participants were able to produce environmental benefits through their work on the number-identification task. To produce these benefits, participants had to incur actual behavioral costs (i.e., they had to spend more time on the online study without receiving additional monetary compensation). As the additional time was filled with an active but monotonous task (i.e., identifying target numbers), participants were not able to forgo the costs of pro-environmental behavior. Hence, the newly developed choice procedure (henceforth referred to as Work for Environmental

ACCEPTED MANUSCRIPT

Protection Task, WEPT) involves actual consequences for both the individual and the natural environment, and both these consequences can be implemented in online environments.

In contrast to the one-shot observations of pro-environmental behavior often used in the laboratory (e.g., assessing whether participants turn out the lights), the WEPT involves multiple choice trials (i.e., participants are asked repeatedly if they would like to complete an additional page of the number-identification task). This multi-trial structure of the WEPT is methodologically advantageous as it allows examining potential determinants of pro-environmental behavior in powerful within-subject designs. In addition, by observing repeated choices of environmental relevance, it is possible to study the temporal course of intervention effects on pro-environmental behavior (e.g., if the effect of an experimental manipulation declines over time; Lange & Dewitte, 2020). Finally, having multiple WEPT trials also allows computing aggregate scores of WEPT performance, which, if proven valid and reliable, can facilitate the study of individual differences in the propensity to behave pro-environmentally (e.g., Kaiser, 1998; Lange & Dewitte, 2019b; Weigel & Newman, 1976).

The present study

To determine if the WEPT is suitable for such research purposes, we conducted a preregistered (https://osf.io/fp4ey/?view_only=3b03ffd02a4b406b822cb176abe59560) validation study. On the one hand, we tested whether people's choice behavior on the WEPT is sensitive to variations in its behavioral costs and environmental benefits. As argued above, the costs of pro-environmental behavior on the WEPT are critical for it to be informative with regard to everyday pro-environmental behaviors. The fact that participants have to work to produce environmental benefits ensures that their support of environmental conservation is more than "cheap talk". However, this reasoning is only valid if the costs implemented in the WEPT are an effective impediment to behaving pro-environmentally. If the costs turn out to be negligible (e.g., because the number-identification task is not tedious enough), the

ACCEPTED MANUSCRIPT

acceptance of such costs to produce environmental benefits would be trivial and thus uninformative. To address this possibility, we independently manipulated the costs of pro-environmental behavior (i.e., how many numbers participants had to screen in order to produce a pro-environmental donation) and examined whether WEPT choice behavior varied as a function of this manipulation.

Hypothesis 1: The likelihood to complete a page of numbers on the WEPT decreases with increasing behavioral costs (i.e., with increasing page size, when more numbers have to be screened in order to generate a pro-environmental donation).

Along similar lines, it is possible that the environmental benefits implemented in the WEPT are too small for participants to take them into account. Through their behavior on the WEPT, participants can generate only small pro-environmental donations (€0.10 – €0.30) and they might not be willing to exert any effort for an environmental impact that small. In that case, they would not encounter an actual trade-off between personal and environmental consequences and the generalizability of WEPT results to everyday pro-environmental behaviors would seem questionable. To address this possibility, we tested if people "care" about the environmental consequences of their WEPT choices, that is, if their choice behavior is sensitive to variations of these consequences. To this end, we manipulated the environmental impact of participants' WEPT choices (i.e., the size of pro-environmental donations) and examined whether WEPT choice behavior varied accordingly.

Hypothesis 2: The likelihood to complete a page of numbers on the WEPT increases with increasing environmental impact (i.e., with increasing donation amount, when completing the page leads to larger pro-environmental donations).

On the other hand, we examined the relationship between aggregated WEPT performance (i.e., the number of completed WEPT pages), other pro-environmental behaviors, and individual difference variables known to be related to pro-environmental

ACCEPTED MANUSCRIPT

behavior (i.e., environmental concern and value orientations). Observations and self-reports of different pro-environmental behaviors and verbal statements regarding environmental protection are typically found to be moderately related to each other as indicators of an individual's general propensity to engage in pro-environmental behavior (Kaiser, 1998; Kaiser, Byrka, & Hartig, 2010). As we consider WEPT choice behavior to be a pro-environmental behavior, we expected individual differences in WEPT performance to correlate with individual differences in other indicators of this general propensity.

Hypothesis 3: The number of completed WEPT pages is positively related to a self-report measure of the propensity to show pro-environmental behavior in everyday life, to the amount of their payment that participants decide to donate to an environmental organization as well as to self-reported environmental concern and biospheric value orientation. The number of completed WEPT pages is also expected to be negatively related to a self-report measure of egoistic value orientation.

Through testing these hypotheses, we probed the psychometric characteristics of the WEPT as a consequential, web-based procedure for the study of pro-environmental behavior. By this means, we aimed to contribute to the methodological toolbox of behavioral researchers interested in studying societally relevant behaviors in online environments.

Methods

Participants

We collected data from 215 participants from the subject pool of the Faculty of Economics and Business at KU Leuven, Belgium. Six of these participants aborted the study before completing all trials of the WEPT and, as preregistered, were excluded from all analyses. This rendered a sample size of $N = 209$ (151 female, 52 male, 4 prefer not to say; 2 participants aborted the study before reaching the demographic questions). Self-reported age ranged from 18 to 65 years, $M = 24.78$ years, $SD = 6.35$ years. Participants reported 29

ACCEPTED MANUSCRIPT

different first languages, with the majority coming from a Dutch-speaking background (56%). Most participants indicated to be students (79%) and to complete the study while staying in Belgium (87%). Participants received €4 for their participation.

We determined the target sample size for this study based on power considerations regarding the correlational test of Hypothesis 3 because this test was likely to be the least powerful of our confirmatory analyses. In a previous study using the PEBT, Lange et al. (2018) found medium-sized correlations ($r \approx .30$) between environmentally relevant choice behavior and conceptually related variables similar to the ones we assessed in the present study. Based on this finding, we selected a target sample size that allowed detecting correlations of $r = .30$ with high statistical power (95%, given a corrected $\alpha = .01$, one-tailed test).

Power analysis (G*Power 3.1.9.2, Faul, Erdfelder, Lang & Buchner, 2007) shows that $N = 167$ participants are necessary to detect correlations of $r = .30$ with a statistical power of 95%. Note that this calculation assumes the computation of Pearson correlation coefficients. In order for our tests to be more robust against the influence of potential outliers, we decided to compute and test non-parametric correlation coefficients (Spearman's rho, r_S). In comparison to parametric correlation analysis, Spearman correlation analysis is less efficient (relative efficiency = 0.91) in detecting relationships when the normality assumptions of parametric correlation analysis are met (Hotelling & Pabst, 1936). We decided to increase our sample size by (1-.0.91)/0.91×100 = 10% to compensate for this potential loss of power. We thus aimed at a target sample size of $N = 184$.

To recruit this sample, we advertised the study to potential participants who were enlisted in the faculty's subject pool. Participants were given the possibility to register for the study using an online sign-up platform, which provided them with a link to the study. Based on previous studies using the same sampling procedure, we anticipated that not all

ACCEPTED MANUSCRIPT

participants registering via that platform would end up completing the study. Allowing for up to 20% attrition, we thus offered 230 testing slots. Participants who registered via the platform were reminded twice to complete the study and online recruitment efforts were stopped after three days. At this point, 209 participants had completed the study and, as preregistered, the additional data were kept for all analyses.

Procedure

All study materials were administered in English and participants were informed that a proficient level of English was required for participation in this study. In the recruitment email and on the informed consent form, participants were instructed to complete the study on a laptop or desktop computer. After providing informed consent, participants first completed the WEPT before proceeding to self-report scales of conceptually related variables, providing demographic information, and making a consequential donation decision (see below). The study was approved by the Social and Societal Ethics Committee of KU Leuven, Belgium (G-2020-2203).

The Work for Environmental Protection Task (WEPT). The WEPT started with the introduction of a number-identification task. Participants were presented with twenty two-digit numbers and instructed to identify all those numbers that consisted of an even first digit and an odd second digit (modelled after Jacobs, 2014). After participants had identified all target numbers on this familiarization trial (with automatic corrective feedback, if required), they were thanked and informed that they "now have the option to continue with the number-identification task for a little longer." The instructions continued:

> "In the following, we will ask you if you want to complete additional pages of the number-identification task. For each page that you correctly complete, we will make a donation to OroVerde, an environmental non-profit organization working on the

ACCEPTED MANUSCRIPT

preservation of tropical forests. OroVerde invests, for example, in reforestation and environmental education in Latin America.

It is up to you to decide how much time and effort you want to invest in the task. There are a maximum of 15 pages that you can complete. For each page, we will tell you how many numbers you will have to check (so that you can estimate the effort) and how much money we will donate if you complete the task. You can then decide, for each page separately, if you want to do this additional effort or not.

Doing this task is completely voluntary. If you want, you can decline checking the numbers (by clicking "no") every time and go directly to the next part of the study. However, please do not simply close the survey before you have reached the end of it (otherwise we do not know whom to pay for their participation)."

With OroVerde, we selected a receiving organization that we expected to be relatively unknown in the target population. Being based in Germany and not involved in any projects in Belgium (where this study was conducted), we deemed it unlikely that participants had encountered the organization before. Hence, we reasoned that participants' behavior towards that organization would be guided primarily by the information provided in the course of the study (rather than by prior associations with the particular organization).

After the instruction, participants were presented with the first choice trial, asking them whether they would like to check X numbers for a donation of Y euros. If participants decided to take on this task, they were presented with a page containing X numbers (see Figure 1). They could take as much time as they wanted to identify the correct target numbers before proceeding to the next choice trial (without corrective feedback). Participants were encouraged to work thoroughly as we would "only count pages that are at least 90% correct".

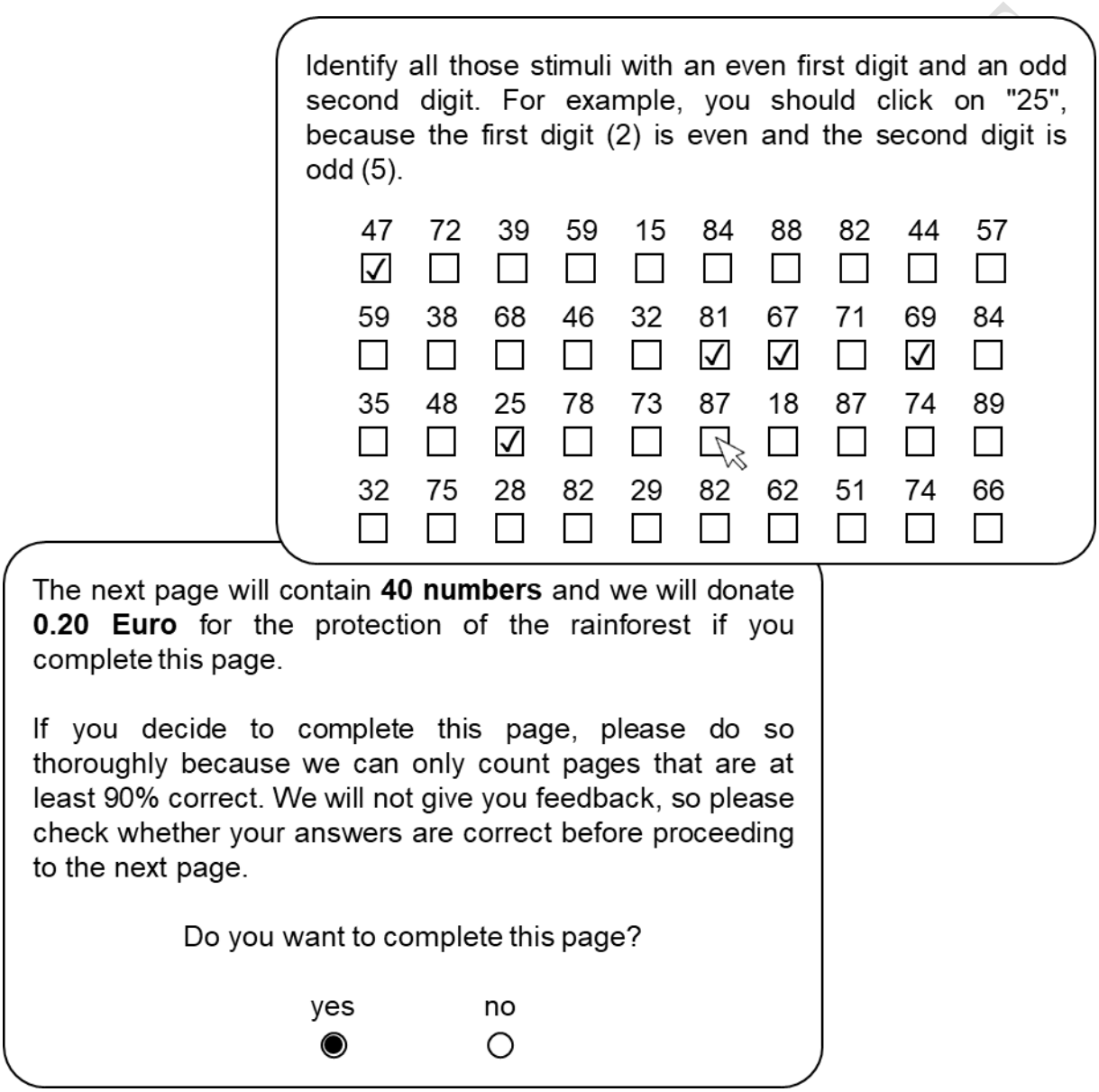
Across the 15 trials, the number X of numbers to be checked and the amount Y of money to be donated was varied to examine Hypotheses 1 and 2. Pages contained 40, 80, 120,

ACCEPTED MANUSCRIPT

160, or 200 numbers. Completing a page of the number-identification task led to a donation of €0.10, €0.20, or €0.30. Combination of these factors yielded 15 conditions, the order of which was randomly determined for each participant.

Figure 1

The Work for Environmental Protection Task (WEPT)



Conceptually related measures. After the last WEPT trial, we administered a set of self-report scales, presented in the same fixed order for all participants. These measures had in common that we considered them related to individuals' propensity to engage in pro-

environmental behavior. Participants first completed a 50-item version of the General Ecological Behavior scale (GEB, Kaiser & Wilson, 2004), a well-established scale for the assessment of individual differences in the propensity to show pro-environmental behavior (for an overview of its psychometric features, see Kaiser et al., 2010; Lange & Dewitte, 2019a). On 32 of the items, participants indicated how frequently they engaged in various pro-environmental behaviors (e.g., "*I buy seasonal produce.*") using a 5-point scale (*never* to *very often*). The remaining 18 items used a dichotomous response scale (e.g., "*I am a vegetarian.*"; *yes* = 1, *no* = 0). Participants were instructed to choose the response option *not applicable* when a question "did not apply to [their] current life situation". In line with established GEB practice, we recoded the 32 items with a frequency response format into a binary format (*very often*, *often* = 1; *never*, *sometimes*, *occasionally* = 0) before calibrating the scale. GEB data were analyzed using a dichotomous Rasch model implemented in the eRm package (Mair, Hatzinger, Maier, Rusch, & Debelak, 2020) in R 4.0.2 (R Core Team, 2020). Person parameters (n = 209, M = 0.02, SD = 0.73) were estimated with reasonable separation reliability (*rel* = .73). One item had to be excluded for this estimation due to constant values (i.e., all participants reported to reuse their shopping bags).

Next, we administered a self-report instrument for the assessment of value orientations (De Groot & Steg, 2010). Participants were asked to rate the importance of 13 values for their lives on a 9-point scale (ranging from -1 = *opposed to the value*, 0 = *not at all important* to 7 = *of supreme importance*). These values formed three scales of egoistic (five items, α = .72), biospheric (four items, α = .87), and altruistic value orientations (four items, α = .76). Participants' average scores for the egoistic (n = 208, M = 3.79, SD = 1.25) and biospheric value orientation scale (n = 208, M = 5.10, SD = 1.28) were computed for preregistered correlation analyses and the altruistic value orientation scale (n = 208, M = 5.63, SD = 1.07) was analyzed as an exploratory measure.

ACCEPTED MANUSCRIPT

Participants then completed the revised New Environmental Paradigm Scale (NEP, Dunlap, Liere, Mertig, & Jones, 2000), a 15-item measure of environmental concern. Participants responded on a 5-point Likert scale (ranging from *strongly disagree* to *strongly agree*) and we computed the average items score across all items (n = 207, M = 3.70, SD = 0.52). Reliability of the NEP in our study was $\alpha = .79$.

Finally, participants had the opportunity to donate a part of their payment to an environmental organization. Using sliders, they decided how much of the €4 they wanted to keep, how much they wanted to donate to OroVerde (i.e., the environmental organization utilized in the WEPT), and how much they wanted to donate to Carbon+Alt+Delete, another environmental organization focusing on the compensation of carbon gas emissions. We offered this second donation opportunity to prevent that OroVerde-specific carry-over effects reduce the total donation and thus attenuate the correlation between WEPT performance and environmental donations (i.e., participants who have already donated a lot of effort to OroVerde might be less willing to also make a monetary donation). Participants were informed (truthfully) that we would add 50% to every donation. Participants donated an average of €0.42 (SD = €0.86) to OroVerde and €0.28 (SD = €0.60) to Carbon+Alt+Delete. Donations to the two organizations were positively correlated, $r_s = .72$, and summed up to a composite score of pro-environmental donations (n = 207, M = €0.71, SD = €1.28).

Preregistered analyses

To test Hypotheses 1 and 2, we ran a logistic regression analysis (using an unstructured working correlation matrix) with the continuous within-subject factors Page Size (40, 80, 120, 160, 200 numbers) and Donation Amount (€0.10, €0.20, €0.30) and the outcome measure Page Completion (no = 0, yes = 1). A WEPT page was counted as completed when the amount of time a participant spent on the page was not more than two standard deviations below the sample mean for that page. This preregistered criterion was used to ensure that

ACCEPTED MANUSCRIPT

participants actually endured the behavioral costs of choosing to complete a WEPT page and we preferred it over an accuracy-based criterion to avoid discounting the efforts of participants less skilled in identifying target numbers. Analysis scripts for the application of this time-based filter criterion and alternative accuracy-based criteria can be found at https://osf.io/gpzy3/?view_only=d09ffc37003343ed824c96b0e47231a5.

We preregistered to regard a negative regression coefficient for the factor Page Size with a 95% Wald confidence interval excluding zero as support for Hypothesis 1. To provide support for Hypothesis 2, regression analysis needed to yield a positive coefficient for the factor Donation Amount with a 95% Wald confidence interval (CI) excluding zero. We did not specify the use of an unstructured working correlation matrix in our preregistration. Results did not change qualitatively (i.e., CIs including/excluding zero continued to include/exclude zero) when we used an independent correlation matrix instead.

To examine Hypothesis 3, Spearman correlation coefficients were calculated for the associations between the total number of completed WEPT pages and five potential correlates: 1) participants' GEB person parameters 2) participants' average score on the NEP, 3) participants' average score on the biospheric value orientation scale, 4) participants' average score on the egoistic value orientation scale, and 5) the total amount participants donated to the two environmental organizations. The level of significance was set to .05/5 = .01 to adjust for the number of examined correlations. We preregistered to regard a negative correlation with the egoistic value orientation scale and positive correlations with the other four measures as support for Hypothesis 3.

Results

Overall, participants decided to complete a total of 1230 WEPT pages (39% of all possible pages). Only one of these pages was not counted as completed due to the completion time criterion specified above (i.e., one participant spent only 43 seconds on a page with an

ACCEPTED MANUSCRIPT

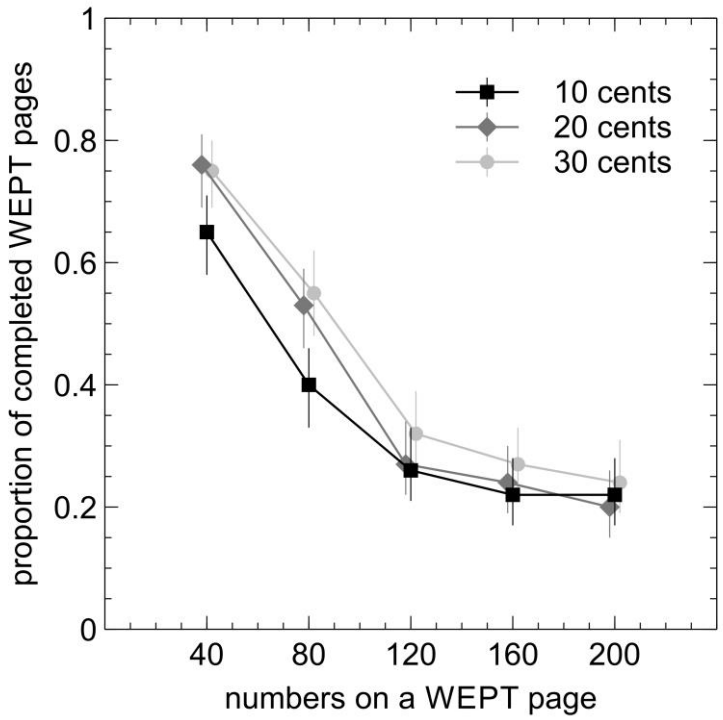
average completion time of 197 s, *SD* = 71 s). Average completion times of the remaining pages closely tracked pages sizes (40 numbers: 42 s, min = 18 s; 80 numbers: 96 s, min = 37 s; 120 numbers: 117 s, min = 44 s; 160 numbers: 172 s, min = 80 s; 200 numbers: 206 s, min = 110 s). Through their completion of WEPT pages, participants generated pro-environmental donations of €253.70.

The odds of completing a page decreased with increasing page size, b = -0.44, 95% CI = [-0.51, -0.38], *Exp(b)*= 0.64, Wald(1) = 171.93, $p < .001$, 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 13% with every 10 additional cents of pro-environmental donations, b = 0.13, 95% CI = [0.06, 0.19], *Exp(b)*=1.13, Wald(1) = 14.12, $p < .001$ (see Figure 2). The size of these effects was slightly changed by adding the interaction of the two within-subject factors, Page Size: b = -0.38, 95% CI = [-0.49, -0.27], Donation Amount: b = 0.22, 95% CI = [0.08, 0.36]. The interaction itself was not significant, b = -0.03, 95% CI = [-0.08, 0.01]. Hence, both Hypothesis 1 and Hypothesis 2 received substantial support from the data.

ACCEPTED MANUSCRIPT

Figure 2

WEPT performance as a function of Page Size and Donation Amount



Note. WEPT = Work for Environmental Protection Task. The proportion of participants completing a WEPT page is displayed as a function of Page Size (i.e., how many numbers were displayed on a WEPT page, 40 – 200) and Donation Amount (i.e., how many cents were donated for a completed WEPT page, 10 – 30). Vertical bars are 95% confidence intervals.

On average, participants completed 5.88 WEPT pages (*SD* = 4.64, Median = 5). Correlations between WEPT performance and conceptually related variables are displayed in Table 1 and Figure 3. Participants who completed more WEPT pages scored significantly higher on the GEB, $r_s = .18$, $p = .004$, and the NEP, $r_s = .20$, $p = .002$. WEPT performance was also positively related to biospheric value orientation, $r_s = .22$, $p < .001$, whereas the correlation with egoistic value orientation was close to zero, $r_s = -.05$, $p = .248$. Finally, WEPT performance was most strongly correlated to our second observation of actual pro-environmental behavior, that is, to participants' donations to an environmental organization, r_s

ACCEPTED MANUSCRIPT

= .33, $p < .001$. Yet, this result should be interpreted with caution due to the large proportion of participants donating €0 (see Figure 3). An alternative analysis of Kendall's tau (not pre-registered) yielded a slightly smaller coefficient of $\tau_b = .26$. WEPT performance was also found to be positively related to altruistic value orientations, $r_s = .21$, $p < .001$ (not pre-registered). With the exception of the non-significant correlation between WEPT performance and egoistic value orientation, these findings largely support Hypothesis 3.

Supplementary analyses

We used the REL_{EX} tool developed by Steinke and Kopp (2020) to estimate the split-half reliability of the WEPT. By this means, we examined how consistently participants were ranked in distinct sets of WEPT trials (i.e., how well their behavior in one half of the trials corresponded to their behavior in the other half). Split-half reliability of the WEPT is likely to differ substantially depending on how the 15 WEPT trials (differing in Page Size and Donation Amount) are split into halves. REL_{EX} addresses this problem by randomly sampling splits from the pool of all possible splits. For the present dataset, reliability sampling using 1,000 iterations revealed a median Spearman-Brown corrected split-half parallel reliability coefficient of $\rho_{SP} = .92$, with 95% of the sampled reliability coefficients lying between $\rho_{SP} = .87$ and $\rho_{SP} = .94$. These results indicate that the WEPT had good split-half reliability in the sample investigated in the present study.

Although participants were explicitly encouraged to complete the study on a computer, 19 participants indicated that they had used a smartphone. While self-reported computer users ($n = 183$; the remaining cases used a tablet, $n = 5$, or were missing values, $n = 2$) completed a mean of 6.03 WEPT pages ($SD = 4.74$), self-reported smartphone users only completed 4.53 WEPT pages on average ($SD = 2.89$). The effects of Page Size, $b = -0.42$, 95% CI = [-0.48, -0.35], and Donation Amount, $b = 0.13$, 95% CI = [0.06, 0.19], tested under Hypotheses 1 and 2 changed only slightly when smartphone users were excluded. The same applies to the

ACCEPTED MANUSCRIPT

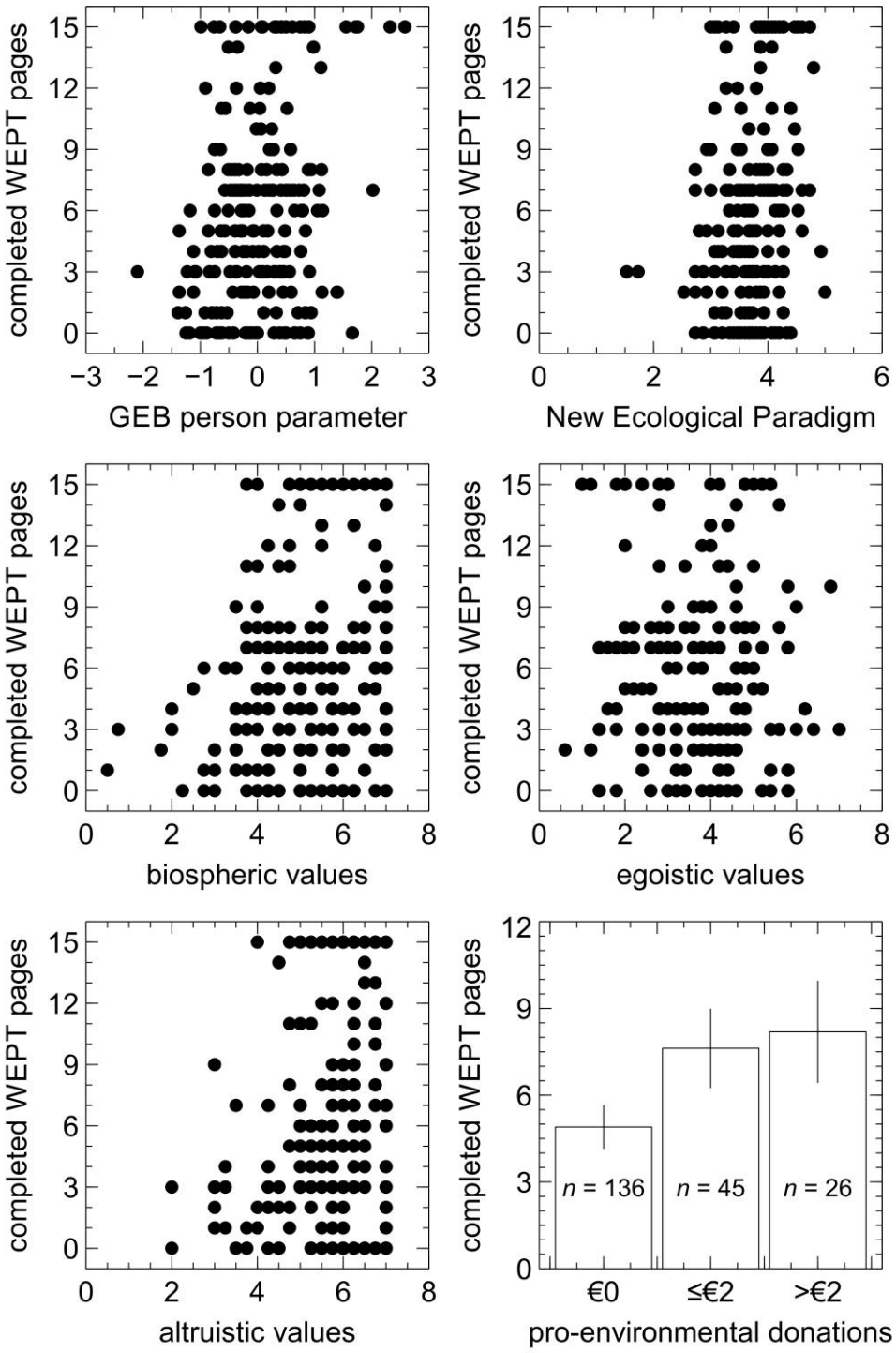
strength of the five correlations tested under Hypothesis 3 (GEB: r_s = .20, NEP: r_s = .21, biospheric values: r_s = .26; egoistic values: r_s = -.06; donations: r_s = .36).

Finally, we repeated our analyses using accuracy-based criteria for considering a WEPT page as completed. On 205 of the 1230 pages participants attempted to complete, participants identified less than 90% of the target numbers. Repeating our analyses while considering these low-accuracy pages as non-completed did not majorly change the results, within-subject effects: Page Size, b = -0.41, 95% CI = [-0.48, -0.34], Donation Amount, b = 0.16, 95% CI = [0.09, 0.23]; correlations: GEB: r_s = .18, NEP: r_s = .17, biospheric values: r_s = .21; egoistic values: r_s = -.05; donations: r_s = .30). Similar results were obtained when we applied an 80%-accuracy criterion (leading to the exclusion of 51 of the attempted pages) instead, within-subject effects: Page Size, b = -0.42, 95% CI = [-0.48, -0.35], Donation Amount, b = 0.12, 95% CI = [0.05, 0.18]; correlations: GEB: r_s = .19, NEP: r_s = .17, biospheric values: r_s = .21; egoistic values: r_s = -.05; donations: r_s = .31). If anything, correlations with conceptually related measures appeared to be slightly larger when using a time-based rather than accuracy-based criterion for considering WEPT pages as completed.

ACCEPTED MANUSCRIPT

Figure 3

Relationships between WEPT performance and conceptually related variables



Note. WEPT = Work for Environmental Protection Task, GEB = General Ecological Behavior scale. Scatter plots display how many of the 15 WEPT pages participants completed as a function of their average scores on conceptually related measures. In case of the GEB (top left), person parameters derived from a Rasch model were used instead of average scores. In case of pro-environmental donations (bottom right), donations to two environmental organizations were summed up and data points were collapsed into bins for display purposes only. Vertical bars are 95% confidence intervals.

Table 1

Spearman correlation coefficients for the relationship between WEPT performance and conceptually related variables

		1	2	3	4	5	6
1	WEPT	-	**[.05, .31]**	**[.09, .35]**	**[-.19, .09]**	**[.07, .33]**	**[.20, .45]**
2	GEB	**.18**	-	[.31, .54]	[-.33, -.07]	[.30, .53]	[-.04, .23]
3	Biospheric values	**.22**	.43	-	[.01, .28]	[.28, .51]	[-.03, .24]
4	Egoistic values	**-.05**	-.20	.15	-	[-.40, -.15]	[-.26, .01]
5	NEP	**.20**	.42	.40	-.28	-	[.00, .27]
6	Donations	**.33**	.10	.11	-.13	.14	-

Note: Correlation coefficients are presented below the diagonal, 95% confidence intervals are presented above the diagonal. Correlations of interest for confirmatory hypothesis testing are highlighted in bold. WEPT = Work for Environmental Protection Task, GEB = General Ecological Behavior scale, NEP = New Environmental Paradigm.

Discussion

ACCEPTED MANUSCRIPT

On the WEPT, participants can generate actual environmental benefits by exerting voluntary extra effort on a monotonous number-identification task. Our results suggest the required extra effort to be an effective impediment to pro-environmental behavior: the more numbers participants had to screen, the less likely they were to complete a page of numbers in exchange for pro-environmental donations. These donations, in turn, were of sufficient size for participants to take them into account when deciding whether or not to make an extra effort: participants were more likely to complete a page of numbers when this had a larger environmental impact. WEPT participants thus have to make repeated trade-offs between two valued consequences: the behavioral costs and the environmental benefits of their behavior. Individual differences in how they make these trade-offs were found to correlate with individual differences in other pro-environmental behaviors and conceptually related measures. These findings support the validity of the WEPT as an online procedure for the study of actual pro-environmental behavior.

Contributions to the experimental analysis of pro-environmental behavior

In view of the present results, we think that online (or laboratory) research using the WEPT can make substantial contributions to our understanding of pro-environmental (and other societally relevant) behaviors. In particular, the task may further the analysis of the functional determinants and underlying mechanisms of pro-environmental behavior. This analysis is difficult to conduct in field studies (that often lack experimental control) and in hypothetical scenario experiments (that often lack external validity). In contrast to hypothetical scenarios, the WEPT has been demonstrated to involve non-trivial costs and actual environmental benefits. These consequences render WEPT performance informative with regard to the understanding of other pro-environmental behaviors that are affected by similar consequences (Lange & Dewitte, 2019a). If, for example, a systematic experimental analysis reveals pro-environmental behavior on the WEPT to vary as a function of a particular

ACCEPTED MANUSCRIPT

manipulation, we would expect this result to generalize to other situations of environmental relevance to the degree to which those situations involve the same conflict between environmental and individual consequences.

To be sure, we suggest such an experimental analysis of WEPT performance to complement rather than to replace field studies on pro-environmental behavior. Both field studies and experiments under more controlled conditions are critical for understanding societally relevant behaviors and generating evidence for behavior change solutions. Yet they fulfill different functions within the process of evidence generation (IJzerman et al., 2020; Mortensen & Cialdini, 2010). Many behavioral scientists may be motivated by the prospect to change societally relevant behavior in everyday situations and, in the end, the effectiveness of every behavior change intervention needs to be demonstrated in the field (just as the value of a newly developed drug has to be demonstrated in clinical practice). We would argue, however, that such interventions stand a greater chance of being effective if they are developed based on a thorough understanding of the societally relevant behavior in question. One way to obtain such an understanding is to subject the behavior to a systematic experimental analysis, and such an analysis can best be conducted under controlled conditions (Lange & Dewitte, 2019a). When conducting laboratory or online studies, experimenters can exert more experimental control than in the field. They can randomize participants to different conditions, precisely control the events to which participants are exposed, use manipulation and attention checks, and relate observations of participants' societally relevant behavior to individual-level background data. In comparison to field studies (which often focus on the comparison of, e.g., one experimental and one control condition), they can also run more elaborated designs, involving, for example, parametric manipulations or the study of interactions between multiple factors. These experimental possibilities are illustrated by the analysis of the within-subject effects of behavioral costs and environmental impact in the

ACCEPTED MANUSCRIPT

present study. They facilitate the study of functional determinants and underlying mechanisms of societally relevant behaviors, and they may also help to pretest and optimize possible intervention techniques before applying them in the field. When looking for novel interventions to change societally relevant behavior (e.g., to promote pro-environmental behavior), controlled laboratory or online studies (e.g., with the WEPT as an outcome measure) can be used to establish causality, robustness, generalizability across stimuli, and possible side effects of the intervention (see also IJzerman et al., 2020). By this means, behavioral scientists can build the knowledge necessary to apply a particular intervention to the real target behavior and to make sure that only the most promising interventions are evaluated in costly field studies.

Contributions to individual difference research

In addition to its use for experimental designs and behavioral intervention research, the WEPT may also help advance the study of individual differences in the propensity to behave pro-environmentally. Given the high reliability of the task and the positive correlations to other measures and indicators of that propensity, aggregated WEPT performance may play a role in the assessment of that propensity. Yet the size of these correlations suggests that the WEPT should not be used as a standalone propensity measure (see also Lange & Dewitte, 2021). Completing pages on the WEPT is a specific pro-environmental behavior (just like donating for an environmental organization or engaging in one of the behaviors assessed by the GEB). Individual differences in specific pro-environmental behaviors seem to overlap to a certain extent (Kaiser, 1998; Weigel & Newman, 1976; see also the correlation between WEPT performance and pro-environmental donation in the present study), but this does not imply that observing a specific pro-environmental behavior would allow distinguishing individuals regarding their tendency to show other pro-environmental behaviors. The overlap can be utilized, however, to obtain

ACCEPTED MANUSCRIPT

reliable composite measures reflecting a person's general propensity to engage in pro-environmental behavior (Kaiser, 1998; Weigel & Newman, 1976). In other words, while completing WEPT pages cannot be regarded as a valid standalone measure of that propensity, it can contribute to propensity measurement as a valid and useful propensity indicator. Previous approaches to develop propensity measures have largely relied on self-reported propensity indicators (e.g., Kaiser, 1998; Kaiser & Wilson, 2000; Markle, 2013). This is not surprising given the difficulty to objectively observe actual pro-environmental behavior in a way that allows relating these observations to individual differences in other propensity indicators (Lange et al., 2018). The reliance on self-report-based propensity measurement is problematic as potential correlates of that propensity (e.g., attitudes or personality traits) are typically assessed via self-report as well. As a result, observed correlations in individual difference research on pro-environmental behavior may be systematically inflated by common-method bias (Lange & Dewitte, 2019a). This problem can be mitigated by observing actual pro-environmental behavior on the WEPT as a valid and easily accessible indicator of the propensity to show pro-environmental behavior. If a sufficient number of validated behavioral indicators (such as the WEPT; the PEBT, Lange et al., 2018; the Greater Good Game, Klein, Hilbig, & Heck, 2017; or the Carbon Emission Task, Berger & Wyss, 2021) are available, behavior can be aggregated across tasks to obtain a reliable and valid performance-based propensity measure.

An important open question in this respect is whether WEPT performance relates specifically to the propensity to show pro-environmental behavior or more generally to the propensity to show non-selfish (or prosocial) behavior. Given that these propensities may partly overlap (Kaiser & Byrka, 2011; Klein et al., 2017), evidence regarding the discriminant validity of the WEPT would be desirable. In the present study, we found WEPT performance to be correlated not only to self-reported biospheric value orientation (which is arguably

ACCEPTED MANUSCRIPT

related to the propensity to show pro-environmental behavior), but also to altruistic value orientation (related to the propensity to show prosocial behavior). Yet the mapping of these value orientations to the two propensities in question may not be very strong and future studies on the specificity of the WEPT should involve a more comprehensive assessment of prosocial and pro-environmental behaviors.

Applicability to other behaviors

Studies on the relationship between pro-environmental and prosocial behavior may also benefit from the versatility of the WEPT. By changing the organization receiving the donations that participants produce through their task performance, the WEPT can easily be transformed into a procedure for the study of prosocial behavior. As all other features of the task can remain constant, changing the receiving organization can allow for ceteris paribus comparisons between behaviors directed at different societally relevant consequences. This does not only allow for contrasting pro-environmental and prosocial behavior, but also for studying commonalities and differences between pro-environmental behaviors directed at different environmental consequences (i.e., if these consequences can be produced through transferring money). For example, adaptations of the WEPT could be completed to the benefit of river cleanups, wildlife conservation, or political campaigns for environmental protection. Alternatively, in line with recent work in experimental economics (e.g., Ockenfels, Werner, & Edenhofer, 2020, see also Berger & Wyss, 2021), WEPT participants could work for the retirement of CO_2 emission certificates.

Through the receiving organization, it is also possible to manipulate factors that may moderate the influence of social or environmental consequences (e.g., their spatial distance). Along similar lines, the WEPT can be altered to yield more or less impactful behaviors by the changing the donation amount per completed page. This might be particularly relevant as high-impact behaviors are often assumed to have different determinants than low-impact

ACCEPTED MANUSCRIPT

behaviors (e.g., Nielsen et al., 2021). The highest donation amount in the present study (i.e., €0.30) was probably too low to speak of a high-impact behavior, but the parameter range can easily be enlarged. The same applies to the costs of the societally relevant behavior (i.e., the size of the page to be screened to produce a donation). By comparing low-cost and high-cost task versions that differ in nothing but their costs, hypotheses about the differential determinants of low-cost and high-cost pro-environmental behavior (e.g., De Groot & Steg, 2009) can be put to a powerful empirical test.

Limitations and concluding remarks

As a limitation of the present validation study, it has to be noted that we recruited a non-random sample from a largely student-based participant pool. Cross-cultural validation studies are required to test if results generalize to other populations. In particular, replications in samples drawn from popular recruitment platforms (e.g., Amazon Mechanical Turk or Prolific) would be desirable to support the validity of the WEPT for use in online research using such platforms. Moreover, while we argue that the WEPT can be used for conducting externally valid experiments on pro-environmental behaviors, we do not expect that the results obtained from any WEPT experiment would generalize to every pro-environmental behavior. As outlined in the Introduction, external validity depends on the degree to which experimentally arranged situations and everyday situations of environmental relevance involve the same contingencies. Some pro-environmental behaviors may involve contingencies other than the ones implemented in the WEPT (e.g., they may not involve any effort or relate to financial benefits for the individual). To ensure that experimental results using the WEPT or other procedures generalize to the behavior of interest, we suggest that researchers first analyze the contingencies involved in that behavior and then adapt the experimental procedure in a way that reflects these contingencies. Relatedly, future studies could test if individual differences in WEPT performance specifically relate to individual

ACCEPTED MANUSCRIPT

differences in particular pro-environmental behaviors. Here, we focused on relationships with an aggregate measure of the propensity to show pro-environmental behavior (i.e., the GEB) and with individual differences in one specific behavior (i.e., pro-environmental donations). Examining correlations with pro-environmental behaviors that involve different contingencies may shed new light on the structure of individual differences in pro-environmental behavior (Lange & Dewitte, 2021).

The effort-based approach of the WEPT is not the only way to study consequential societally relevant behavior online. One alternative would be to endow participants with a monetary budget that they can distribute among themselves and, for example, an environmental organization. The donation measure we used in the present study is an example of this approach, but validated money-based tasks (such as the Greater Good Game, Klein et al., 2017; or the Carbon Emission Task, Berger & Wyss, 2021) can also be used in online settings. We do not think that the WEPT is generally superior to such money-based approaches, although some of the versatility advantages outlined above may be difficult to realize when using typical donation measures. We do think, however, that an exclusive focus on money-based tasks would be problematic and that research on societally relevant behaviors can benefit from a more diverse set of methodological tools (Farrelly & Bennett, 2018; Farrelly et al., 2015). Societally relevant behaviors and their determinants may critically differ depending on the type of costs they involve (e.g., money vs. effort vs. time). For some research questions, it is also possible that the type of costs does not matter, but this is an empirical question that can only be answered when we have tools involving different consequences at our disposal. As a validated and versatile procedure, the WEPT can add to the set of available behavioral research tools, complement existing approaches, and offer new opportunities for the study of pro-environmental and other societally relevant behaviors.

Open Practices Statement

ACCEPTED MANUSCRIPT

The study was preregistered and run in accordance with the preregistration. All data and analysis scripts as well as the Qualtrics file of the WEPT version used in this study can be found at https://osf.io/gpzy3/?view_only=d09ffc37003343ed824c96b0e47231a5. We confirm that we have reported all measures, conditions, data exclusions, and how we determined our sample size.

ACCEPTED MANUSCRIPT

References

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science, 2*, 396-403.

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

Birnbaum, M. H. (2004). Human research and data collection via the Internet. *Annual Review of Psychology*, *55*, 803-832.

Charness, G., Gneezy, U., & Henderson, A. (2018). Experimental methods: Measuring effort in economics experiments. *Journal of Economic Behavior & Organization*, *149*, 74-87.

Cialdini, R. B. (2009). We have to break up. *Perspectives on Psychological Science*, *4*, 5-6.

Cornelissen, G., Pandelaere, M., Warlop, L., & Dewitte, S. (2008). Positive cueing: Promoting sustainable consumer behavior by cueing common environmental behaviors as environmental. *International Journal of Research in Marketing*, *25*, 46-55.

De Groot, J. I., & Steg, L. (2009). Mean or green: which values can promote stable pro-environmental behavior? *Conservation Letters*, *2*, 61-66.

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*, 368-378.

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*, 425-442.

ACCEPTED MANUSCRIPT

Farrelly, D., & Bennett, M. (2018). Empathy leads to increased online charitable behaviour when time is the currency. *Journal of Community & Applied Social Psychology*, *28*, 42-46.

Farrelly, D., Moan, E., White, K., & Young, S. (2015). Evidence of an alternative currency for altruism in laboratory-based experiments. *Europe's Journal of Psychology*, *11*, 100-111.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191.

Gifford, R. (2011). The dragons of inaction: psychological barriers that limit climate change mitigation and adaptation. *American Psychologist*, *66*, 290-302.

Glanz, K., Rimer, B. K., & Viswanath, K. (Eds.). (2008). *Health behavior and health education: theory, research, and practice*. John Wiley & Sons.

Gosling, S. D., & Mason, W. (2015). Internet research in psychology. *Annual Review of Psychology*, *66*, 877-902.

Hartshorne, J. K., de Leeuw, J. R., Goodman, N. D., Jennings, M., & O'Donnell, T. J. (2019). A thousand studies for the price of one: Accelerating psychological science with Pushkin. *Behavior Research Methods*, *51*, 1782-1803.

Herziger, A., & Hoelzl, E. (2017). Underestimated Habits: Hypothetical Choice Design in Consumer Research. *Journal of the Association for Consumer Research*, *2*, 359-370.

Hinde, R. A., & Groebel, J. (Eds.). (1991). *Cooperation and prosocial behaviour*. Cambridge University Press.

Hotelling, H., & Pabst, M. R. (1936). Rank correlation and tests of significance involving no assumption of normality. *The Annals of Mathematical Statistics*, *7*, 29-43.

ACCEPTED MANUSCRIPT

IJzerman, H., Lewis, N. A., Przybylski, A. K., Weinstein, N., DeBruine, L., Ritchie, S. J., ... & Anvari, F. (2020). Use caution when applying behavioural science to policy. *Nature Human Behaviour*, *4*, 1092-1094.

Jacobs, B. (2014). Analyse von Testgütekriterien und Übungseffekten zweier Online-Konzentrationstests [Analysis of test quality and practice effects of two online-concentration tests]. http://hdl.handle.net/20.500.11780/3371

Kaiser, F. G. (1998). A general measure of ecological behavior. *Journal of Applied Social Psychology*, *28*, 395-422.

Kaiser, F. G., & Byrka, K. (2011). Environmentalism as a trait: Gauging people's prosocial personality in terms of environmental engagement. *International Journal of Psychology*, *46*, 71-79.

Kaiser, F. G., & Wilson, M. (2000). Assessing People's General Ecological Behavior: A Cross-Cultural Measure 1. *Journal of Applied Social Psychology*, *30*, 952-978.

Kaiser, F. G., Byrka, K., & Hartig, T. (2010). Reviving Campbell's paradigm for attitude research. *Personality and Social Psychology Review*, *14*, 351-367.

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

Klein, S. A., & Hilbig, B. E. (2019). On the lack of real consequences in consumer choice research: And its consequences. *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.

Kormos, C., & Gifford, R. (2014). The validity of self-report measures of proenvironmental behavior: A meta-analytic review. *Journal of Environmental Psychology*, *40*, 359-371.

ACCEPTED MANUSCRIPT

Krahé, B. (2020). *The social psychology of aggression*. Routledge.

Lange, F., & Dewitte, S. (2019). Cognitive flexibility and pro-environmental behaviour: A multimethod approach. *European Journal of Personality*, *33*, 488-505.

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

Lange, F., & Dewitte, S. (2020). Positive Affect and Pro-environmental Behavior: A preregistered experiment. *Journal of Economic Psychology*, 102291.

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

Lange, F., & Iwasaki, S. (2020). Validating the Pro-Environmental Behavior Task in a Japanese Sample. *Sustainability*, *12*, 9534.

Lange, F., Steinke, A., & Dewitte, S. (2018). The Pro-Environmental Behavior Task: A laboratory measure of actual pro-environmental behavior. *Journal of Environmental Psychology*, *56*, 46-54.

Lehman, P. K., & Geller, E. S. (2004). Behavior analysis and environmental protection: Accomplishments and potential for more. *Behavior and Social Issues*, *13*, 13-33.

Lewandowski Jr, G. W., & Strohmetz, D. B. (2009). Actions can speak as loud as words: Measuring behavior in psychological science. *Social and Personality Psychology Compass*, *3*, 992-1002.

Liu, T., Geng, L., Ye, L., & Zhou, K. (2019). "Mother Nature" enhances connectedness to nature and pro-environmental behavior. *Journal of Environmental Psychology*, *61*, 37-45.

Mair, P., Hatzinger, R., Maier, M., Rusch, T., & Debelak, R. (2011). Package eRm (extended Rasch modeling). *Reference manual. R package version 0.14-0.* http://cran.r-project.org/web/packages/eRm/index.html

ACCEPTED MANUSCRIPT

Maner, J. K. (2016). Into the wild: Field research can increase both replicability and real-world impact. *Journal of Experimental Social Psychology*, *66*, 100-106.

Markle, G. L. (2013). Pro-environmental behavior: Does it matter how it's measured? Development and validation of the Pro-Environmental Behavior Scale (PEBS). *Human Ecology*, *41*, 905-914.

Morales, A. C., Amir, O., & Lee, L. (2017). Keeping it real in experimental research—Understanding when, where, and how to enhance realism and measure consumer behavior. *Journal of Consumer Research*, *44*, 465-476.

Mortensen, C. R., & Cialdini, R. B. (2010). Full-cycle social psychology for theory and application. *Social and Personality Psychology Compass*, *4*, 53-63.

Moussaoui, L. S., Desrichard, O., & Milfont, T. L. (2020). Do environmental prompts work the same for everyone? A test of environmental attitudes as a moderator. *Frontiers in Psychology*, *10*, 3057.

Murtagh, N., Gatersleben, B., Cowen, L., & Uzzell, D. (2015). Does perception of automation undermine pro-environmental behaviour? Findings from three everyday settings. *Journal of Environmental Psychology*, *42*, 139-148.

Nielsen, K. S., Clayton, S., Stern, P. C., Dietz, T., Capstick, S., & Whitmarsh, L. (2021). How psychology can help limit climate change. *American Psychologist*, 76, 130-144.

Ockenfels, A., Werner, P., & Edenhofer, O. (2020). Pricing externalities and moral behaviour. *Nature Sustainability*, *3*, 872-877.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Sassenberg, K., & Ditrich, L. (2019). Research in social psychology changed between 2011 and 2016: Larger sample sizes, more self-report measures, and more online studies. *Advances in Methods and Practices in Psychological Science*, *2*, 107-114.

ACCEPTED MANUSCRIPT

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

Steg, L. (2015). Environmental psychology and sustainable consumption. In: L.A. Reisch & J. Thøgersen (eds.), *Handbook of Research in Sustainable Consumption* (pp. 70-83). Cheltenham: Edward Elgar.

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

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

Swim, J. K., Clayton, S., & Howard, G. S. (2011). Human behavioral contributions to climate change: Psychological and contextual drivers. *American Psychologist*, *66*, 251-264.

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.

Van Bavel, J. J., Baicker, K., Boggio, P. S., Capraro, V., Cichocka, A., Cikara, M., ... & Drury, J. (2020). Using social and behavioural science to support COVID-19 pandemic response. *Nature Human Behaviour*, *4*, 460-471.

Van der Linden, S., Maibach, E., & Leiserowitz, A. (2015). Improving public engagement with climate change: Five "best practice" insights from psychological science. *Perspectives on Psychological Science*, *10*, 758-763.

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.

ACCEPTED MANUSCRIPT