The value of multi-proxy experiments to study pro-environmental behavior

Millet, Kobe; Weijters, Bert

published in
Journal of Environmental Psychology
2025

DOI (link to publisher)
10.1016/j.jenvp.2024.102495

document version
Early version, also known as pre-print

citation for published version (APA)
Millet, K., & Weijters, B. (2025). The value of multi-proxy experiments to study pro-environmental behavior. *Journal of Environmental Psychology*, *101*, Article 102495. https://doi.org/10.1016/j.jenvp.2024.102495

The value of multi-proxy experiments to study pro-environmental behavior.

Kobe Millet[1] & Bert Weijters[2]

[1]Vrije Universiteit Amsterdam

[2]Ghent University

Preprint v3 (October 10, 2024).

[The final version is published as correspondence, including the tutorial in Appendix, in *Journal of Environmental Psychology* https://doi.org/10.1016/j.jenvp.2024.102495]

Author Note

Both authors contributed equally (author names are listed alphabetically).

Kobe Millet (kobe.millet@vu.nl), Marketing Department, School of Business and Economics, Vrije Universiteit Amsterdam, 1081 HV Amsterdam, Netherlands.

Bert Weijters (bert.weijters@ugent.be), Department of Work, Organization and Society, Faculty of Psychology and Pedagogical Sciences, Ghent University, Dunantlaan 2, B9000 Ghent, Belgium.

Acknowledgments: We thank Hans Baumgartner, Bob Fennis, Anneleen Van Kerckhove and Niels van de Ven for their friendly reviews of earlier versions of this manuscript.

Abstract

In the study of pro-environmental behavior, experimental studies are essential for identifying causal relationships and underlying psychological mechanisms. However, despite the multi-dimensional nature of the construct "pro-environmental behavior", experimental studies often adopt a single, fixed, proxy measure (e.g. one specific pro- environmental decision) to capture this construct. The problem herewith is that this approach ignores the idiosyncratic nature of the specific proxy used, undermining the reliability of study conclusions on pro-environmental behavior in general. In contrast herewith, the proposed multi-proxy experimental approach makes use of a wide variety of proxy measures to operationalize pro-environmental behavior: participants are randomly assigned to one out of a large set of different proxy measures (i.e. a variety of different pro-environmental decisions). This approach preserves the strengths of classic experimental design while increasing the reliability of study conclusions. Moreover, it allows for the collection of additional information about specific characteristics of the proxy measures used (also post hoc), which can then in turn be employed as moderating variables. This may not alone help identify the specificity of the experimental effects, but can also be used to test different conceptual models against each other. We provide a roadmap to implement the multi-proxy experimental approach.

When studying pro-environmental behavior (PEB), experimental studies are essential for identifying causal relationships and underlying psychological mechanisms. Experimental studies often adopt a single, fixed, proxy measure (e.g. one specific pro-environmental decision) to capture PEB; for instance, taking the bicycle instead of the car (Lange et al., 2018), switching off the light instead of leaving it on (Murtagh et al., 2015) or opting for a more expensive environmentally friendly option rather than a cheaper less eco-friendly alternative for a limited set of products like milk, pens, or mobile phones (Van der Werff et al., 2014). While studying specific PEBs is valuable and often necessary[1], researchers frequently use such study results to draw conclusions about PEB in general or unintentionally give this impression, thereby overlooking its multidimensional nature. Indeed, narrow operationalizations of constructs cannot simply be generalized to the construct level (Yarkoni, 2022). Moreover, the idiosyncratic properties of a specific proxy make it also challenging or even impossible to exclude confounding effects. For instance, the environmentally friendly choice of tap water over bottled water may be motivated by cost differences, taste preferences, practical issues, availability, etc. (Deltomme, 2023; Deltomme et al. 2023). Related to this ad hoc selection of proxies, often only a limited set of PEBs receive research attention, as illustrated by the lack of heterogeneity in published PEB studies in the flagship consumer research journals (Lembregts & Cadario, 2024)[2]. As a result, the conclusions of this literature only apply to this very limited set of PEBs.

We introduce **multi-proxy experiments** as a new approach to help addressing these issues. Instead of incorporating a fixed proxy as dependent measure, multi-proxy experiments make use of a large set of different proxies, each associated with the construct of interest, and

[1] For some research questions, generalization may not be appropriate as these questions are specific to particular choices or contexts.
[2] Note that these often turn out to be behaviors with low climate mitigation potential (Lembregts & Cadario, 2024).

participants are randomly assigned one of these proxies as dependent measure. As far as we know, this approach has only been used in two recent publications to address conceptual questions (Millet et al., 2022; Millet & Weijters, 2023) without giving much attention to the methodology used[3]. Our aim is to highlight the unique value and specific aspects of this approach as it differs considerably from traditional experiments and current practices.

A multi-proxy experiment implements many different proxies simultaneously, in contrast to a specific, fixed proxy as the experimental outcome. Thus, instead of operationalizing PEB by means of one proxy (e.g., turning out the lights), many different proxies are adopted (i.e. the proxy pool, e.g. not just turning out the lights, but also dozens of other pro-environmental behaviors, like sorting waste and buying carbon offsets). For each participant, a single proxy[4] is *randomly* drawn from this proxy pool (to represent PEB). Across participants and proxies, multilevel analysis can then be used to examine the overall effect. By doing so, observed effects will be much more generalizable than traditional experiments adopting a single fixed proxy to capture the construct of interest.

Perhaps even more interestingly, this approach also enables the opportunity to identify for which type of behavior an experimental effect is stronger, attenuated, or even inverts. It is possible to measure characteristics of the proxies not only upfront but also post hoc, and use the initially collected experimental data as secondary data for post hoc moderation analyses. For instance, the identity signaling value for each proxy used to operationalize PEB in Study 2 of Millet et al. (2022) was measured after the experiment had already been published (e.g., choosing to drink tap water instead of bottled water does not strongly express someone's

---

[3] While we focus on random assignment of proxies for the dependent variable given its relevance for the study of pro-environmental behavior, the use of random, instead of fixed, stimuli has already been advocated in e.g. psycholinguistics (Clark, 1973), social psychology (Judd et al., 2012) and cognitive science (Baribault et al., 2018).

[4] This prevents potential order effects that may occur when multiple proxies are presented sequentially to the same subject, which is especially relevant given the possibility of both positive (e.g., Lanzini & Thøgersen, 2014) and negative spillover effects (e.g., Thøgersen & Ölander, 2003).

identity, whereas opting to eat only plant-based food does signal one's identity; Millet & Weijters, 2023). Subsequently, the multilevel analysis from the initial multi-proxy experiment was extended by incorporating the identity signaling value of each proxy as a moderating variable (at the proxy level) to explore if the experimental effect changes depending on the proxy's identity signaling value. This post-hoc moderation analysis provided invaluable insights. Whereas no effect was found in the published experiment, a cross-over interaction emerged in the post hoc moderation analysis, showing opposite experimental effects depending on the identity signaling value of the pro-environmental choices (Millet & Weijters, 2023). This demonstrates that data from a multi-proxy experiment can still reveal meaningful insights post hoc, given the opportunity to explore the impact of potentially important proxy characteristics, even when these are identified in hindsight.

Compared to traditional experiments, there are also some challenges. First, since the setup of a multi-proxy experiment and its data analyses differ from traditional experiments, researchers may be less familiar with this approach. However, with the right guidance and support, adopting this method is quite straightforward. Therefore, we provide a **roadmap** in Appendix *how* to execute and analyze multi-proxy experiments and show how to use initially collected experimental data as secondary data by using the pro-environmental behavior studies and (publicly available) data from Millet et al. (2022) and Millet and Weijters (2023) that employed this approach.

Another challenge is the lack of clear guidelines regarding participant sample sizes. It is not yet fully clear how many subjects need to be sampled for multi-proxy studies to achieve sufficient statistical power. We offer some considerations and guidelines in Appendix. Although multi-proxy experiments typically require larger sample sizes than traditional experiments, it is important to note that multi-proxy experimental data can be reused as secondary data, thereby offering many opportunities for future research.

Additionally, developing a proxy pool to represent PEB is demanding: selecting viable proxies may turn out to be difficult, and extensive pretesting is required. However, the development of reliable and pretested proxy pools benefits all researchers interested in studying PEB. Other research teams may adopt existing proxy pools in their multi-proxy experiments, and the proxy features can be used as moderators. Furthermore, if new information on features of proxies in an existing proxy pool is obtained, post-hoc moderation analyses can be conducted for all experiments that made use of the same proxy pool.

In conclusion, the use of multi-proxy experiments not only enhances the generalizability of experimental research findings but also provides numerous additional benefits. The ability to leverage previous multi-proxy experiments and proxy sets (as well as its pretested features) to explore new research questions underscores the potential of this innovative experimental approach. We strongly believe the proposed paradigm holds the promise to facilitate a more efficient, collaborative, and standardized approach to investigate research questions and may therefore result in more effective experimental research on pro-environmental behavior.

REFERENCES

Baribault, B., Donkin, C., Little, D. R., Trueblood, J. S., Oravecz, Z., van Ravenzwaaij, D., White, C. N., De Boeck, P., & Vandekerckhove, J. (2018). Metastudies for robust tests of theory. *Proceedings of the National Academy of Sciences of the United States of America, 115*, 2607–2612.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12*, 335–359.

Deltomme, B. (2023). *The Measurement of Pro-Environmental Behavior* [Unpublished doctoral dissertation]. Ghent University.

Deltomme, B., Gorissen, K., & Weijters, B. (2023). Measuring pro-environmental behavior: Convergent validity, internal consistency, and respondent experience of existing instruments. *Sustainability, 15*, 14484.

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: a new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology, 103*, 54.

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.

Lanzini, P., & Thøgersen, J. (2014). Behavioural spillover in the environmental domain: An intervention study. *Journal of Environmental Psychology, 40*, 381-390.

Lembregts, C., & Cadario, R. (2024). Consumer-driven climate mitigation: Exploring barriers and solutions in studying higher mitigation potential behaviors. *International Journal of Research in Marketing, 41*, 513-528.

Millet, K., Du, G., Cabooter, E., & Weijters, B. (2022). The limited impact of positive cueing on pro- environmental choices. *Journal of Environmental Psychology, 79*, 101732.

Millet, K., & Weijters, B. (2023). The behavioral intervention positive cueing: altering self-perception, increasing green awareness, or undermining the signaling value of costly green behavior? *Journal of Environmental Psychology, 87*, 101979.

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.

Thøgersen, J., & Ölander, F. (2003). Spillover of environment-friendly consumer behaviour. *Journal of Environmental Psychology, 23*, 225-236.

van der Werff, E., Steg, L., & Keizer, K. (2014). Follow the signal: When past pro-environmental actions signal who you are. *Journal of Environmental Psychology, 40*, 273-282.

Yarkoni, T. (2022). The generalizability crisis. *Behavioral and Brain Sciences, 45*, 1-78.

APPENDIX

The value of multi-proxy experiments to study pro-environmental behavior.

Table of Contents

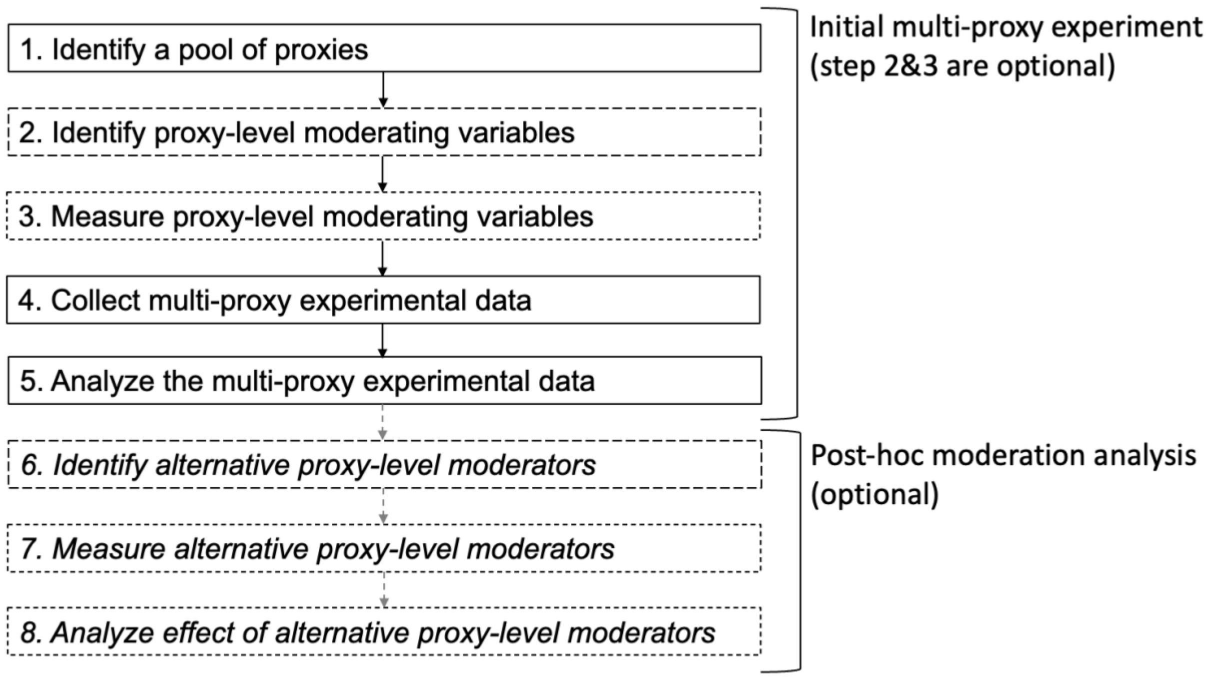
APPENDIX A

A step-by-step guide to conducting multi-proxy experiments

We provide a step-by-step outline of how to conduct a multi-proxy experiment, starting from the situation where the researcher has defined a traditional experiment with a control condition and an experimental condition that are hypothesized to affect PEB. Taking it from here, conducting a multi-proxy experiment requires the following steps (see Figure 1; steps 2 and 3 are optional): (1) identifying a pool of proxies for the dependent variable PEB, (2) identifying proxy-level moderating variables, (3) measuring proxy-level moderating variables, (4) collecting multi-proxy experimental data, (5) analyzing the multi-proxy experimental data. Optionally, follow-up studies can additionally (6) identify alternative proxy-level moderators, (7) measure the proxy-level moderators, and (8) analyze the effect of these moderators by merging the data from step 7 with those from step 4.

Figure 1.

Roadmap of a multi-proxy experimental approach



As a running example used to illustrate what a multi-proxy experiment might look like, we summarize the approach used in a multi-proxy experimental project on PEB consisting of two studies (Study A and Study B) reported by resp. Millet et al. (2022) and Millet and Weijters (2023). Study A corresponds to the preregistered study 2 by Millet et al. (2022) and used a multi-proxy approach to test the generalizability of positive cueing (i.e. a specific social influence technique) effects to a set of product choices (a variety of proxies to represent PEB). Positive cueing entails that when common environmentally beneficial behaviors (e.g., not littering, reusing grocery bags) are cued as pro-environmental, consumers increasingly view themselves as concerned with the environment, which subsequently results in more environmentally friendly choices. The study used a 2 (fixed factor) by 81 (random factor) between-subjects design. N = 1615 participants were randomly assigned to either a positive cueing vs. control condition (i.e., the fixed between-subjects factor). In both conditions, participants rated the frequency with which they engage in each of five behaviors. In the positive cueing condition, the question read: “I usually engage in the following pro-environmental behavior,” and the behaviors were green behaviors that people regularly engage in (turning off the light, reusing grocery bags, not littering, sorting trash, and using reusable water bottles/cups/coffee mugs). In the control condition, the question read: “I usually engage in the following behavior,” and the behaviors were common behaviors that have no clear relation to environmentalism (watching Netflix series, listening to music, reading news articles, chatting with friends, and cooking). Next, participants received a choice task. Crucially, for this choice task, respondents were randomly assigned to one of 81 pretested choice pairs (i.e., the proxy pool corresponding to the random between-subjects factor in the design). Each choice pair included a sustainable and non-sustainable option (e.g., use recycled toilet paper vs. non-recycled toilet paper). Participants indicated preferences for the sustainable option vs. the non-sustainable option on a rating scale; the resulting response was used as the dependent variable. After data were collected in this multi-proxy experiment (Study A), the characteristic "identity signaling value" of the different proxies was measured (in an independent sample of respondents; Millet & Weijters, 2023) and combined with Study A's initially collected experimental data to run a post hoc moderation analysis (Study B).

Step 1: identifying a pool of proxies

In the first stage, proxies need to be identified to be used in the experiment as outcome variable. Until further methodological guidelines are developed (as the number of proxies that is needed depends on several factors, including the effect size and the heterogeneity of the experimental effect of interest across proxies), we propose that researchers formulate at least 30 such proxies, as this is a commonly recommended lower boundary for the number of clusters in a multi-level analysis (Hox & McNeish, 2020). There is no clear upper limit to the number of proxies to be generated, but it seems unlikely that adding many extra proxies over a hundred would add much value. One can use secondary or primary sources for this.

Secondary sources. In the scenario where the outcome of interest has been extensively studied before, it is probably meaningful to identify proxies in secondary sources like the extant literature and/or the world-wide web. For example, the proxy pool for the illustrative positive cueing study (A) consisted of a set of 81 choice pairs (see Appendix A.S1) based on Moran et al. (2020), who studied the carbon emission impact of a broad portfolio of consumer behavior choices that are environmentally relevant. Alternatively, due to the recent rise of Generative Artificial Intelligence tools (like ChatGPT) a particularly efficient way to harvest proxies from the web involves simply asking a Generative Artificial Intelligence tool, using probes like the following: "Come up with a series of approximately 80 binary choices that pitch a non-sustainable option vs. an environmentally sustainable option. The choices should be everyday choices that consumers typically face on a regular basis, e.g., drink water from the tap vs. bottled water, go to work by car vs. by bicycle." Appendix A.S2 has an example of proxies generated in this way.

Primary sources. For outcome variables that have not been extensively studied in extant research or if the set of proxies used in the literature to represent the construct of interest is quite limited, it might be necessary to identify proxies based on primary sources. For instance, also for PEB, only a single proxy set has been created. It is important to be aware that this set, as used in Millet et al. (2022) is based on a single paper of Moran et al. (2020) who focused on a portfolio of green behavior changes that may reduce CO_2 emissions. Still, PEB is much broader than that (e.g. think about choices to protect biodiversity, reduce pollution, promote air quality), or could be extended in particular domains (e.g. transport, food or energy). Moreover, it has recently been pointed out that research attention is typically focused on a limited set of PEBs (Lembregts & Cadario, 2024). In such a scenario, researchers could use qualitative research techniques (focus groups, interviews, etc.) among relevant subject populations, asking them to come up with examples of PEB that they are familiar with. Similarly, quantitative research among the relevant subject population could involve surveys with open questions, asking a similar question (e.g., "Please list some choices that you and other people face in daily life where you have an option that is good for the environment versus an option that is bad for the environment."). Interviews and surveys among experts could then be used to validate the proxies that respondents from the general population came up with (e.g., asking them whether the proposed proxies indeed pitch a pro-environmental option against an option that harms the environment).

If an initial pool of proxies has been self-generated, researchers need to verify that the proxies are representative of the construct of interest (our focus is on PEB here). If the secondary source the proxies are coming from has already verified this, this step could be skipped (or kept rudimentary). Otherwise, it is valuable to assess face validity by running a survey among a relevant sample of respondents asking for each of the proxies to what extent they are instances of the construct of interest. For instance, the survey could ask participants for each proxy: "to what extent is choosing option A over option B an example of pro-environmental behavior?", measured on a five-point scale with 1= no good example at all, 2= no good example, 3= neither a good example nor a bad example, 4= a good example, 5= a very good example. Proxies with scores not significantly higher than the midpoint 3 should probably be dropped from the pool.

Step 2: identifying proxy-level moderating variables

In the second stage, proxy-level moderating variables may be identified, if considered relevant. Note that a multi-proxy experiment can also be run without proxy-level moderators (in case no moderator will be included, the reader can skip steps 2 and 3, and proceed to step 4). For identifying potentially relevant moderators, we distinguish between inductive vs. deductive approaches.

Inductive approach. Primary sources are typically used in an inductive approach, i.e. when starting with specific observations or data and moving toward broader, more general conclusions or theories. If the topic of investigation is relatively new and there are no established conceptual frameworks at hand, researchers might want to identify proxy-level moderators by further investigating variation in proxies they identified in step 1. This investigation will entail a study in and of itself, and involves the collection of new primary data, and applying an inductive approach in which the proxies from step 1 act as the objects under study. More specifically, research could use qualitative techniques such as triadic sorting, in which respondents are asked to sort a random set of three proxies into a pair that shares a common characteristic versus a stand-out proxy that does not. For instance, respondents might say that switching off the light and turning down the heating both save money and energy, whereas buying carbon offsets costs money (and does not necessarily save energy), identifying monetary cost as a PEB attribute (and potential moderator). Quantitative techniques that serve a similar purpose include perceptual mapping and multidimensional scaling based on quantitative surveys that include similarity ratings.

Deductive approach. A deductive approach starts with a general theory or hypothesis and will typically be based on secondary data. Conceptually relevant moderators could be identified to test specific hypotheses that logically follow from conceptual frameworks and existing theories. Useful secondary sources to draw some conceptual input from include literature reviews, the identification of moderators in extant studies (experimental or observational), and meta-analyses. For instance, in our running example identity signaling value of a choice was a moderator of the positive cueing effect; for the underlying conceptual rationale, please refer to Millet and Weijters (2023).

Step 3: measuring proxy-level moderating variables

In the third step, all proxies from the pool (step 1) are assessed on a specific potentially relevant feature, which will be used as proxy-level moderating variable (if a moderator is included; otherwise skip to the next step). To assess proxy features, researchers can make use of primary sources, such as quantitative surveys among either a sample of target participants or experts (or if the feature is directly observable and self-evident, the researchers can even code the feature themselves). More specifically, an online survey could be conducted in which respondents rate all the proxies on the moderating variable of interest. To reduce respondent fatigue and enhance data quality, it is useful to implement planned missingness designs (i.e. intentionally omitting questions for respondents based on a predefined strategy, Graham et al., 2006).

For instance, in the illustrative positive cueing study from our running example, data were collected among Prolific participants (N = 466) to evaluate proxies. Each participant received a random subset of nine choice pairs sampled from the set of 81 choice pairs. For each of these pairs, participants were presented with the following instructions: "Suppose someone had the following two options, and decided to do A (not B) (followed by A: [description of sustainable choice] and B: [description of non-sustainable choice])." Option order (i.e., whether A or B was the sustainable option) was randomized. Participants then answered the following two questions on identity signaling: (a) "How much does the choice for one of the options contribute to self-expression (i.e., a person's ability to express their identity)?“, (b) "How much do people use the choice for one of the options in this decision context to make inferences about others (i.e., people think they know a lot about a person based on their choice in this particular choice dilemma)?". Using the 81 choice pairs as the unit of analysis, the mean rating of these two items ($r = 0.85$) was calculated as a measure of identity signaling of each proxy or choice pair (N = 81). For example, on a scale from 1 to 7, 'drinking tap water vs. bottled water' has a mean identity signaling score of 3.68; 'only eat plant-based foods vs. meat' has a mean score of 5.43 (see Appendix A.S1 for all values).

Step 4: collecting multi-proxy experimental data

Once a pool of proxies has been created (note that step 2 and 3 can also be executed post hoc), the actual experiment can take place. A multi-proxy experiment is similar to a traditional experiment, randomly assigning participants to between-subjects conditions (manipulating the independent variable of interest), and then measuring its effect on the proxy for a dependent variable. The only difference is that proxies will be randomly drawn from a pool of proxies, so that every subject receives one randomly sampled proxy as dependent measure, instead of all participants receiving one fixed proxy as dependent measure (as in a traditional approach). Thus, except for the random assignment of proxies to measure the construct of interest, the experiment itself is very comparable to a traditional experiment, enabling the opportunity to include other moderators as well (e.g. individual difference measurements).

Step 5: analyzing the multi-proxy experimental data

Once the experimental data have been collected, it is time for the data analysis. At this point, it is useful to visualize how the multi-proxy approach compares to a traditional experiment, again making use of the positive cueing experiment from our running example. Figure 2a presents a diagram how an experimental manipulation x_i (the positive cueing manipulation for individual i) causes an effect on a dependent variable y_i (eco-choice for individual i), where the dependent variable for this illustration is the choice between drinking tap water vs. bottled water. Figure 2b presents a diagram how the same experimental manipulation x_i (the positive cueing manipulation) causes an effect on a dependent variable y_{ij} (eco-choice), but here each individual i is nested in proxy j and proxy j has a given level z_j of variable z related to it. Variable z moderates the effect of x on y. Eco-choice in the traditional experiment relates to one specific instance of a choice, e.g., tap water vs. bottled water. By contrast, in the multi-proxy experiment, the choice between tap water vs. bottled water is only one of many possible proxies used to gauge PEB.

Figure 3a shows how the experimental positive cueing effect is different for different proxies of PEB and even ranges from negative to positive. In the illustrative positive cueing research, a multi-level analysis showed that approximately one third of the variance in the dependent variable (eco-choice, i.e., rated preference for the sustainable vs. non-sustainable option) was situated at the between-proxy level (ICC = .354). Note that this variance could only be observed because of the use of many proxies (if there is only one outcome proxy, there cannot be variance among proxies). This is important, since it side-steps the limitation of previous experimental research that uses a single proxy and cannot observe variation across proxies by design. Using multiple proxies allows for deeper exploration of how the effect of the behavioral intervention varies depending on specific proxy characteristics.

As an important extra aspect, the proxies can be treated as instantiations of different levels of the moderator variable z (identity signaling value). To visualize this moderating effect, Figure 3b shows the experimental effect (i.e. positive cueing effect) as a function of the moderating variable identity signaling. It now becomes clear that the positive cueing effect does not simply randomly vary across proxies, but is concomitant on a proxy's identity signaling value: proxies with a low identity signaling value (e.g., drinking tap water) show a positive cueing effect, but the effect reverses for proxies with a high signaling value (like sharing leftovers in an online community vs. throwing them away).

Figure 2a.

Diagram of a traditional experiment

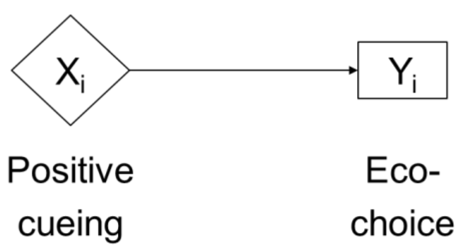


Figure 2b

Diagram of the newly proposed multi-proxy experimental approach

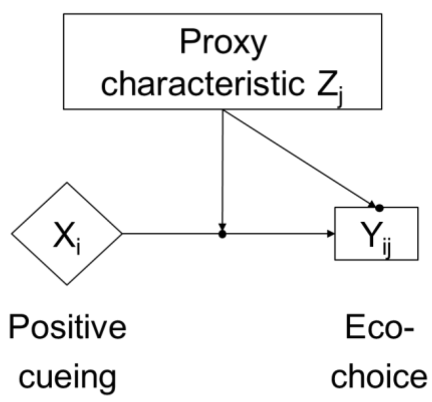
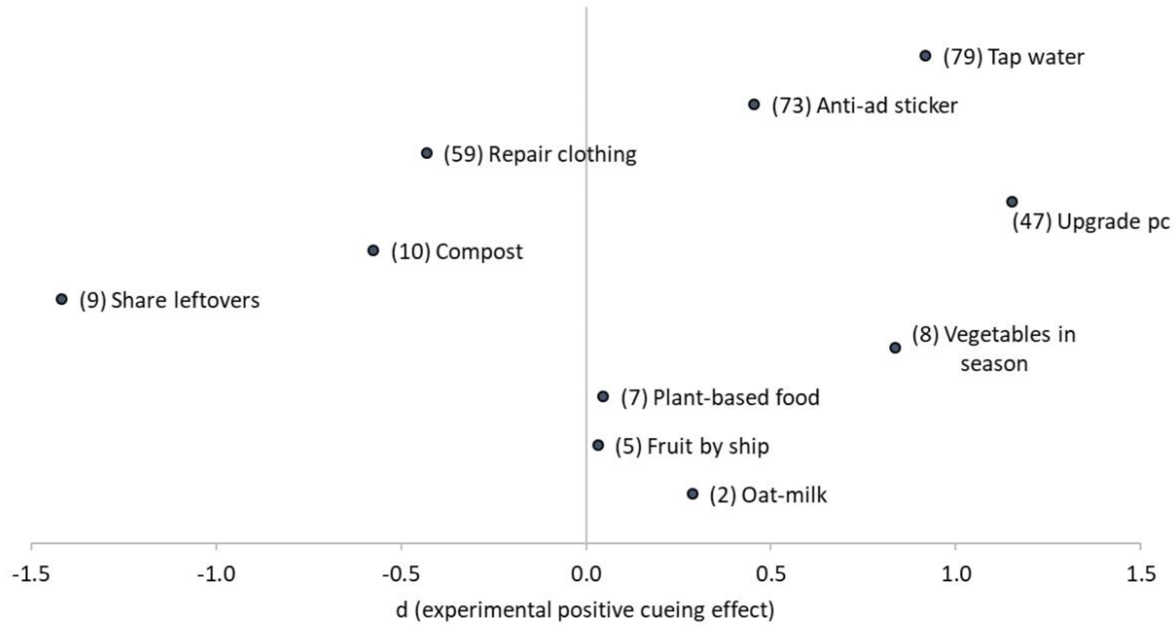


Figure 3a.

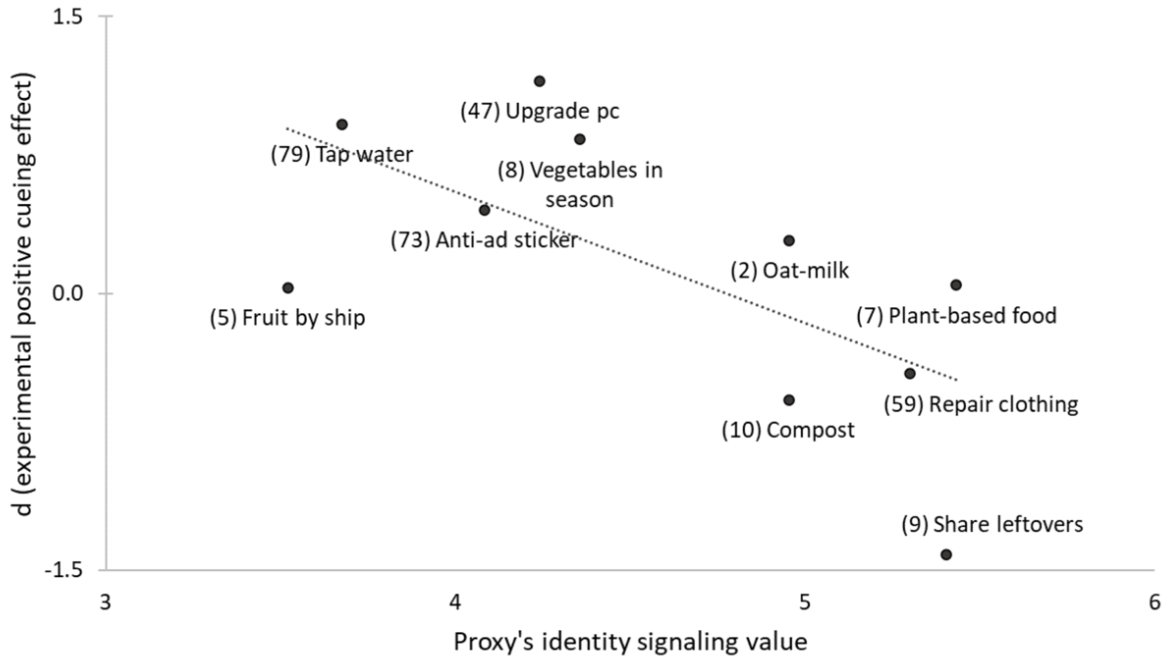
Positive cueing as a multi-proxy random effect



Note. Experimental positive cueing effect for selected proxies (numbers between brackets refer to the proxy's sequence number in Appendix A.S1). The X-axis shows the strength and direction of the positive cueing effect. Proxies are ranked arbitrarily (in order of the list in Appendix A.S1) on the Y-axis.

Figure 3b.

Positive cueing effect (Y) moderated by proxies' identity signaling value (X)



Note. Experimental positive cueing effect (Y) as a function of the proxy's identity signaling value (X) for selected proxies (numbers between brackets refer to the proxy's sequence number in Appendix A.S1).

Let us now provide some step-by-step guidance for analyzing the data of a multi-proxy experiment. If a moderator variable is hypothesized and measured, the experimental data from step 4 first need to be merged with the proxy-level measurement data from step 3. If no moderator is included, the reader can skip to the next paragraph. The experimental dataset from step 4 has N individual participants as its row or unit of analysis, nested in K proxies (for an example, see the first four columns in the data file on ResearchBox #2258).

The moderator dataset from step 3 has K proxies as its row or unit of analysis[5] (for an example, see the first four columns of the table in Appendix A.S1). The result should be one dataset with N individual participants as the rows (or units of analysis), enriched with the moderator variables from step 3 (for an example, see the full data file on ResearchBox #2258).

The main analysis uses a multi-level model with individuals at level 1 and stimuli at level 2. The Mplus syntax for this analysis is included in Appendix A.S3, the R syntax is included in Appendix A.S4. Both syntax files and the dataset used can be found on ResearchBox (#2258). In what follows, we consider two models. If no moderator is included, only model 1 can be estimated. If proxy moderator data are available, researchers can estimate both model 1 and model 2 (in sequence).

In model 1, at level 1 (the individual level), the dependent variable y for individual i and proxy j is modeled as a function of experimental manipulation dummy x as follows:

$$y_{ij} = \beta_{0j} + \beta_{1j} * x_{ij} + \varepsilon_{ij} \quad (1)$$

Where:

- β_{0j} is the random intercept at Level 2 (proxy level).

- β_{1j} is the random slope for the experimental effect at Level 2 (proxy level).

- ε_{ij} is the individual-level residual error.

[5] As a practical note, for merging two such datasets, multiple software packages can be used. In SPSS, researchers can use the menu ‘data > merge files > add variables’. In Excel, researchers can use the VLOOKUP function. In R, researchers can use the merge command from the dplyr package (but many alternative functions and packages are available in R with similar functionality).

At level 2 (proxy Level):

The random effects are specified as follows:

$$\beta_{0j} \sim N(\gamma_{00}, \tau_{00})$$

$$\beta_{1j} \sim N(\gamma_{10}, \tau_{11})$$

Where:

- γ_{00} is the mean of the random intercept.

- τ_{00} is the variance of the random intercept.

- γ_{10} is the mean of the random slope for the effect of x

- τ_{11} is the variance of the random slope for the effect of x

In words, the experimental effect (i.e., the effect of x on y) is specified as a random effect; that is, the effect has a mean but also a variance. Put simply, the experimental effect varies in strength (and possibly even direction) across proxies. This point is illustrated in the last column in the table in Appendix A.S1, which shows the experimental positive cueing effect observed for each proxy (for illustrative purposes operationalized as the difference in eco-choice score between the control condition and the positive cueing condition, as measured on a five-point scale). In model 1, a parameter of key interest in the results is the mean of the random effect (γ_{10}). If this parameter's confidence interval does not include zero, it is very likely that on average (across the different proxies) the experimental manipulation has an effect on the dependent variable.

Model 2 extends Model 1 by adding a moderator variable z at level 2 (proxy level). This moderator affects both the random intercept and the random slope. Level 1 (individual level) remains the same, but at level 2 (proxy level), we add effects of variable z on the random intercept β_{0j} and the random slope β_{1j}:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} * z_j + \zeta_{0j} \quad (2)$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} * z_j + \zeta_{1j} \quad (3)$$

Where:

- γ_{01} is the effect of the moderator z on the random intercept β_{0j}.

- ζ_{0j} is the residual term in the random intercept.

- γ_{11} is the effect of the moderator z on the random slope β_{1j}.

- ζ_{1j} is the residual term in the random slope.

In words, model 2 uses a specification that extends model 1 by adding a moderator variable at the between-proxy level (level 2). This moderator variable z functions as an antecedent of both dependent variable y's random intercept and the random effect. The moderation between z and x is a cross-level interaction that quantifies how much the proxy-level moderator strengthens (or weakens if it is negative) the experimental effect. For instance, if x is presence of positive cueing, y is PEB, and z is the identity signaling value of engaging in PEB (which varies as a function of the specific outcome stimulus, i.e., the specific PEB choice used as the outcome), this cross-level interaction might indicate that the experimental effect is stronger/attenuated/inverted depending on the signaling value of the PEB. Appendix A.S3 and A.S4 also contain an excerpt of respectively the Mplus and R output with annotations.

Steps 6 and 7: identify and measure alternative proxy-level moderators

Steps 1 through 5 result in a full-blown multi-proxy experiment. Once such an experiment has been conducted, we strongly recommend not only publishing its result, but also making available all the anonymized data online (including proxies, moderator measurements, experimental data, and analysis syntax) on platforms like www.osf.io or https://researchbox.org/, paying special attention to the so-called FAIR principles (i.e., principles of findability, accessibility, interoperability, and re-usability for both humans and machines).

For instance, the data (N=1615) of the positive cueing study in our running example (study A) is publicly available (ResearchBox #329), so that independent research teams can analyze moderating effects of other proxy characteristics (that were not included in the initial study) on this dataset by measuring new proxy characteristics post hoc and combining these with the initial data (as done in Study B; ResearchBox #1232). For convenience, we created a new ResearchBox (#2258) with dataset, and both Mplus and R syntax files so the reader can easily conduct the analyses themselves.

Once this has been done, follow-up studies can additionally (6) identify alternative proxy-level moderators, (7) measure these proxy-level moderators, and (8) analyze the effect of these moderators by merging the data from step 7 with those from step 4. The specific actions for steps 6, 7 and 8 are operationally identical to steps 2, 3 and 5, so we will not repeat them here. But importantly, the data in step 4 and step 7 do not need to be collected by the same research teams. This specific feature thus enables efficient and open scientific synergies, which may help to accelerate research to provide much-needed insights into the heterogeneity of pro-environmental choices.

REFERENCES

Graham, J. W., Taylor, B. J., Olchowski, A. E., & Cumsille, P. E. (2006). Planned missing data designs in psychological research. *Psychological Methods, 11*, 323.

Hox, J., & McNeish, D. (2020). Small samples in multilevel modeling. In R. van de Schoot & M Miočević (Eds). Small sample size solutions. *A Guide for applied Researchers and Practitioners*: Routledge.

Lembregts, C., & Cadario, R. (2024). Consumer-driven climate mitigation: Exploring barriers and solutions in studying higher mitigation potential behaviors. *International Journal of Research in Marketing, 41*, 513-528.

Moran, D., Wood, R., Hertwich, E., Mattson, K., Rodriguez, J. F. D., Schanes, K., & Barrett, J. (2020). Quantifying the potential for consumer-oriented policy to reduce European and foreign carbon emissions. *Climate Policy, 20*, S28–S38.

Millet, K., Du, G., Cabooter, E., & Weijters, B. (2022). The limited impact of positive cueing on pro- environmental choices. *Journal of Environmental Psychology, 79*, 101732.

Millet, K., & Weijters, B. (2023). The behavioral intervention positive cueing: altering self-perception, increasing green awareness, or undermining the signaling value of costly green behavior? *Journal of Environmental Psychology, 87*, 101979.

Appendix *A.S1*: Proxy pool used in empirical illustration

Proxy #	Ecological option	Non-ecological option	Identity signaling	Positive cueing effect
1	Buy only the food I need so I don't have leftovers	Buy some more food than I need (and throw away leftovers)	4.53	0.2
2	Drink oat-based milk substitute	Drink traditional cow milk	4.96	0.3
3	Drink tap water	Drink coffee or tea	3.64	0.5
4	Go to a local farmer to buy potatoes locally	Go to the supermarket to buy potatoes	4.91	0.1
5	Buy fruit / vegetables transported by ship	Buy fruit / vegetables transported by plane	3.52	0.0
6	Buy vegetables without packaging	Buy vegetables in convenient packaging	4.57	0.1
7	Only eat plant-based foods	Eat meat (in addition to plant-based foods)	5.43	0.0
8	Only buy vegetables in the right season (not in heated greenhouses)	Buy vegetables grown in heated greenhouses (e.g. tomatoes in winter)	4.36	0.8
9	Share leftovers via online food communities	Throw away leftovers	5.40	-1.4
10	Compost unavoidable food waste (e.g. potato peels)	Throw unavoidable food waste (e.g. potato peels) in the trash bin	4.95	-0.6
11	Buy food in recycled packaging	Buy food in plastic trays	5.02	-0.3
12	Recycle steel-cans	Throw steel-cans in the trash bin	4.91	-0.1
13	Use public transport to commute to work	Commute to work by car	4.20	0.6
14	Buy a smaller car	Buy a larger car or SUV	5.18	0.2
15	Buy a smaller, fuel-efficient car	Buy a larger, sportier car	5.35	-0.3
16	Buy a light-weight car that uses less fuel	Buy a heavier car	4.44	0.7
17	Buy a car with recycled material content	Buy a regular car with no recycled material content	4.50	0.2
18	Adopt a more fuel-efficient driving style when using the car	Adopt a sporty driving style when using the car	5.08	0.1
19	Join a car-sharing service	Have my own car	4.64	0.1
20	Car pool with other people	Drive my car alone	4.34	-0.2
21	Choose a location with job close to home	Live and work anywhere I like	4.32	-0.6
22	Shift my transport to public transport, cycling or walking	Keep on using a car whenever I want to	5.12	0.4
23	Keep a car for many years	Buy new cars frequently	4.81	0.9
24	Reduce my living space by 20%	Increase my living space by 20%	4.29	0.7
25	Use more teleworking	Use less teleworking	3.68	0.5

26	Use renewable building materials for a house (e.g. wood, straw)	Use commonly used materials for a house (brick, steel, concrete, etc.)	5.19	-0.8
27	Use construction material with recycled content	Use construction material without recycled content	4.51	-0.2
28	Use locally manufactured building materials	Use building materials that were not locally manufactured	4.25	-0.2
29	Initiate actions to thermally insulate my house / apartment	Do not initiate actions to thermally insulate my house / apartment	4.28	0.6
30	Have a green roof installed	Keep the roof as it is	4.93	0.0
31	Construct a zero-emission house	Construct a regular house	5.55	0.1
32	Buy an energy efficient refrigerator	Buy a larger but less energy efficient refrigerator	4.65	-0.9
33	Always completely turn off devices that are not in use	Put devices in standby mode when not in use	4.09	0.4
34	Reduce the room temperature by 2°C	Keep the room temperature as it is	3.79	0.2
35	Use cement/concrete more efficiently	Make regular use of cement/concrete	3.70	-0.3
36	Join a co-housing / shared space with other people / families	Have a place to live on my own (or with only my own family)	4.54	0.0
37	Refurbish my living space for greater efficiency in water and energy use	Keep my living space as it is	5.11	0.1
38	Refurbish a building with recycled materials	Refurbish a building with new materials	5.10	0.1
39	Recycle construction waste	Do not recycle construction waste	4.80	0.2
40	Share the use of a washing machine with other individuals/families	Have and use a washing machine for myself / my family	3.81	0.8
41	Repair a device when it breaks down	Replace a device when it breaks down	4.69	0.5
42	Postpone buying a new TV for one extra year	Buy a new TV when I feel like it	4.05	-0.1
43	Buy a computer with a smaller LCD monitor	Buy a computer with a larger LCD monitor	3.82	0.6
44	Buy energy efficient electronic devices used for communications	Buy regular electronic devices used for communications	4.50	0.1
45	Share, rent, or lease a computer	Buy my own computer	3.90	-0.3
46	Buy more long lasting & repairable equipment (e.g. Fairphone)	Buy the newest equipment on a regular base	5.19	-0.6
47	Replace and upgrade components (e.g. memory) of a computer	Buy a new computer	4.24	1.2
48	When my computer's hard disk breaks down, I take action to replace it	When my computer's hard disk breaks down, I buy a new computer	4.01	-0.5
49	Recycle ICT waste (e.g. old phones and computers)	Throw away ICT waste (e.g. old phones and computers)	4.81	-0.6
50	Take old devices back to a retailer	Throw away old devices	4.68	0.6
51	Buy textiles made of cotton, wool	Buy textiles made of silk instead of acrylic, polypropylene, PA6, polyester	4.50	-0.3
52	Purchase textiles that contain recycled fibers	Purchase textiles that do not contain recycled fibers	4.61	0.6
53	Use energy efficient washing machines, tumble dryers, and irons	Use any type of washing machines, tumble dryers, and irons	4.55	0.0
54	Reduce the number of washes through for my clothes	Keep the number of washes for my clothes as it is	4.12	0.0

55	Plan the load of washing machines for optimal energy efficiency	Use my washing machine whenever and however it suits me	4.44	0.5
56	Always hang clothes to dry without using a tumble dryer	Dry my clothes faster in a tumble dryer	4.24	0.8
57	Wash clothes at lower temperatures	Wash clothes at regular temperatures	3.91	-0.3
58	Buy more durable garments	Buy new, fashionable clothes on a regular basis	5.03	0.2
59	Extend the life of clothing and textiles through repairs	Buy new, fashionable clothes on a regular basis	5.30	-0.4
60	Reuse textiles for a different purpose	Throw away used textiles	4.78	0.1
61	Donate clothes to charity	Throw away used clothes	5.08	0.0
62	Drop off unwanted clothes into recycling bins at stores	Throw away used clothes	5.06	-0.2
63	Buy furniture produced with low CO2 emissions (Ecolabel)	Buy any furniture I like	4.61	-0.9
64	Buy naturally treated timber furniture	Buy any furniture I like	5.03	0.5
65	Buy furniture that contains a high amount of recycled materials	Buy any furniture I like	4.89	0.7
66	Buy furniture made within your region	Buy furniture made outside your region	4.23	0.0
67	Buy products that are packaged in recyclable materials	Buy new products with any type of packaging	4.82	-0.5
68	Buy products where replacement of furniture parts is provided	Buy any furniture I like	4.11	-0.1
69	Reuse/redesign furniture components	Throw away old furniture	5.01	0.5
70	Bring old furniture to a second-hand store	Throw away old furniture	5.11	-0.4
71	Donate furniture to charity	Throw away old furniture	4.91	0.0
72	Only print what is needed	Print anything that is convenient	4.57	0.0
73	Avoid mailshot by placing an anti-advertisement sticker	Throw uninteresting mail away	4.08	0.5
74	Buy recycled writing paper	Buy whiter, nicer non-recycled writing paper	4.60	-0.4
75	Buy recycled toilet paper	Buy non-recycled toilet paper	4.56	-0.5
76	Print paper two-sided	Print paper one-sided	3.92	0.4
77	Read books and papers electronically	Read printed versions of books and newspapers	4.07	0.0
78	Reduce the number of plastic items	Keep the number of plastic items as it is	4.85	-0.1
79	Drink tap water	Drink bottled water	3.68	0.9
80	Use my own reusable bag when shopping	Get a plastic shopping bag	5.09	-0.1
81	More efficient use of chemicals	Less efficient use of chemicals	4.16	0.2

Appendix *A.S2*: Illustrative AI-generated list of proxies

Illustrative list of proxies formulated by https://chat.openai.com/, accessed on Sept 27, 2023 in response to the question: "Come up with a series of approximately 80 binary choices that pitch a non-sustainable option vs. an environmentally sustainable option. The choices should be everyday choices that consumers typically face on a regular basis, e.g., drink water from the tap vs. bottled water, go to work by car vs. by bicycle."

1. Drive a gas-guzzling SUV vs. a fuel-efficient hybrid car.

2. Take a long, hot shower vs. a short, efficient shower.

3. Purchase single-use plastic water bottles vs. use a reusable water bottle.

4. Toss recyclables in the trash vs. recycle them properly.

5. Use disposable diapers vs. cloth diapers.

6. Keep the lights on when not in the room vs. turn them off.

7. Buy fast fashion clothing vs. invest in quality, sustainable fashion.

8. Eat beef daily vs. incorporate plant-based meals into your diet.

9. Purchase a new smartphone every year vs. keep your current one longer.

10. Fly frequently for short trips vs. use alternative transportation.

11. Leave the thermostat on high vs. adjust it for energy savings.

12. Use a gas-powered lawnmower vs. an electric or manual one.

13. Toss food waste in the garbage vs. composting.

14. Buy products with excessive packaging vs. minimal or eco-friendly packaging.

15. Use disposable plastic cutlery vs. bring your reusable utensils.

16. Keep appliances plugged in when not in use vs. unplug them.

17. Buy single-serving coffee pods vs. use a reusable coffee filter.

18. Drive alone to work vs. carpool or use public transportation.

19. Use paper towels for everything vs. reusable cloth towels.

20. Purchase items with excessive plastic wrapping vs. choose plastic-free options.

21. Use chemical-based cleaning products vs. eco-friendly alternatives.

22. Print hard copies of documents vs. digital storage and sharing.

23. Dispose of old electronics in the trash vs. recycling e-waste properly.

24. Order takeout with single-use containers vs. dine in or bring your containers.

25. Maintain a large, resource-intensive lawn vs. xeriscaping with native plants.

26. Use conventional incandescent light bulbs vs. energy-efficient LED bulbs.

27. Leave the tap running while brushing teeth vs. turning it off.

28. Buy heavily processed foods with excessive packaging vs. whole, unpackaged foods.

29. Use a gas-powered generator vs. invest in solar panels or a battery backup.

30. Upgrade to the latest gadgets and devices vs. make the most of what you have.

31. Dispose of hazardous materials in the trash vs. proper disposal at a facility.

32. Wash clothes with hot water vs. cold water.

33. Use disposable plates and utensils at parties vs. reusable ones.

34. Drive alone during rush hour vs. flexible work hours or telecommuting.

35. Purchase products with microplastics vs. those without.

36. Keep old, energy-inefficient appliances vs. replacing them with efficient ones.

37. Use plastic straws vs. reusable or biodegradable alternatives.

38. Buy non-organic produce vs. organic and locally sourced options.

39. Leave your computer on 24/7 vs. shut it down when not in use.

40. Use a gas-powered snowblower vs. shoveling snow manually.

41. Buy individually wrapped snacks vs. bulk snacks in reusable containers.

42. Keep a single-use coffee pod machine vs. a traditional coffee maker.

43. Use traditional, chemical-laden pesticides vs. natural pest control methods.

44. Drive to the gym vs. exercise outdoors or at home.

45. Purchase bottled juices vs. making fresh juices at home.

46. Use disposable razors vs. safety razors with replaceable blades.

47. Buy products tested on animals vs. cruelty-free alternatives.

48. Use conventional, petroleum-based lip balm vs. natural alternatives.

49. Print documents single-sided vs. double-sided.

50. Choose fast food with excessive packaging vs. dine-in restaurants.

51. Use non-rechargeable batteries vs. rechargeable ones.

52. Buy cosmetics with harmful chemicals vs. natural and eco-friendly cosmetics.

53. Toss old clothes in the trash vs. donating or recycling textiles.

54. Use plastic shopping bags vs. reusable shopping bags.

55. Buy bottled cleaning products vs. making homemade cleaners.

56. Purchase heavily packaged toys vs. minimal packaging or secondhand toys.

57. Eat out every meal vs. cooking at home with local ingredients.

58. Use disposable cameras vs. digital photography.

59. Buy products with excessive advertising vs. making informed choices.

60. Purchase imported goods with a high carbon footprint vs. local alternatives.

61. Keep windows open while the heating/cooling system is running vs. seal drafts.

62. Drive aggressively vs. drive conservatively to save fuel.

63. Use plastic-based wrapping paper vs. reusable gift wrap or cloth.

64. Buy single-serving snacks vs. bulk snacks in reusable containers.

65. Toss old furniture vs. repairing or repurposing it.

66. Use chemical air fresheners vs. open windows for ventilation.

67. Buy fast food with excessive disposable utensils vs. bring your own.

68. Wash dishes with a running tap vs. filling a basin with water.

69. Purchase rare or endangered species products vs. supporting conservation.

70. Use disposable contact lenses vs. daily reusable lenses.

71. Order imported seafood vs. sustainably sourced local seafood.

72. Leave a car engine idling vs. turning it off when parked.

73. Use disposable party decorations vs. reusable decorations.

74. Print travel itineraries and tickets vs. digital versions on your phone.

75. Buy bottled fruit juices vs. making your own fresh juice.

76. Use disposable wipes vs. reusable washcloths.

77. Purchase products with excessive single-use plastics vs. plastic-free options.

78. Keep the refrigerator door open while deciding what to eat vs. closing it promptly.

79. Use non-biodegradable glitter vs. biodegradable alternatives.

80. Keep the car windows down at high speeds vs. using air conditioning efficiently.

Appendix *A.S3*: Mplus syntax for analyzing multiproxy experiments.

MODEL 1 input: random main effect (i.e., no moderator included)

```
TITLE: Mplus syntax for multiproxy experiment;
DATA: file = multiproxydata.dat;
VARIABLE:
names = id stimulus x y;
!id is the individual identity marker
!stimulus is the stimulus identity marker
!x is a dummy coded variable for the experimental manipulation (0 =
control, 1 = experimental condition)
!y is the dependent variable
cluster = stimulus;
idvar = id;
usevar =  x y;
within = x;
ANALYSIS: type = twolevel  random;
OUTPUT: cinterval;
DEFINE: center y (GRANDMEAN); !mean center the dependent variable
MODEL:
%within%
s | y on x; !specifies a random slope for the experimental effect
%between%
y on z; !specifies the direct cross-level effect of the moderator
y; !residual variance in the DV
s; !variance in the experimental effect across stimuli
[s]; !mean cross-level interaction labeled 'a'
```

MODEL 2 input: cross-level interaction (i.e., moderator included)

```
TITLE: Mplus syntax for multiproxy experiment with cross-level
moderation;
DATA: file = multiproxydata.dat;
VARIABLE:
names = id stimulus x y z;
!id is the individual identity marker
!stimulus is the stimulus identity marker
!x is a dummy coded variable for the experimental manipulation (0 =
control, 1 = experimental condition)
!y is the dependent variable
!z is the stimulus-level moderator
cluster = stimulus;
idvar = id;
usevar =  x y z;
between = z;
within = x;
ANALYSIS: type = twolevel  random;
OUTPUT: cinterval;
DEFINE:
center y (GRANDMEAN); !mean center the dependent variable
center z (GRANDMEAN); !mean center the moderator variable
MODEL:
%within%
s | y on x; !specifies a random slope for the experimental effect
```

```
%between%
s on z(b); !specifies the cross-level interaction and labels it 'b'
y on z; !specifies the direct cross-level effect of the moderator
s with y; !allows the slope to covary with y
y; !residual variance in the DV
s; !residual variance in the slope
[s](a); !mean cross-level interaction labeled 'a'
!Last lines optionally plot experimental effect on moderator
MODEL CONSTRAINT:
PLOT(crosslvl); !
LOOP(z, -2, 2, 0.1);
crosslvl = a+b*z;
PLOT: TYPE=PLOT2;
```

MODEL 2 selected and annotated output: cross-level interaction (i.e., moderator included)

MODEL RESULTS

		Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
	Within Level				
	Residual Variances				
$Var(\varepsilon_{ij})$	Y	1.159	0.066	17.440	0.000
	Between Level				
	S ON				
γ_{11}	Z	-0.275	0.099	-2.782	0.005
	Y ON				
γ_{01}	Z	0.655	0.184	3.554	0.000
	S WITH				
	Y	-0.022	0.046	-0.478	0.633
	Intercepts				
γ_{00}	Y	-0.030	0.093	-0.321	0.748
γ_{10}	S	0.063	0.049	1.307	0.191
	Residual Variances				
$Var(\zeta_{0j})$	Y	0.598	0.089	6.730	0.000
$Var(\zeta_{1j})$	S	0.006	0.048	0.132	0.895

Note: Mplus output for model 2 is shown in the text box on the right hand side, with annotations linking to the multilevel model equations in the main text shown on the left hand side. As a key parameter estimate, γ_{11} (or 's on z') represents the cross-level interaction effect between the proxy-level moderator and the experimental effect.

Appendix *A.S4*: R syntax for analyzing multiproxy experiments.

```
# Load required packages
library(lme4)
library(ggplot2)

# Read the data (use setwd() to set working directory)
setwd("C:/…")
df <- read.table("multiproxydata.dat", header = F)
colnames(df)<-c("id","stimulus","x","y","z")

# Centering variables
df$y_c <- scale(df$y, center = TRUE, scale = FALSE)
df$z_c <- scale(df$z, center = TRUE, scale = FALSE)

# Fit multilevel model1
model1 <- lmer(y_c ~ x +  (1 + x| stimulus) , data = df)

# Display model summary
summary(model1)

# Fit multilevel model2
model2 <- lmer(y_c ~ x +  (1 + x| stimulus)  + z_c + x:z_c, data = df)

# Display model summary
summary(model2)

# Plot the interaction effect
z_values <- seq(-2, 2, 0.1)
crosslvl <- predict(model2, newdata = data.frame(x = 1, z_c = z_values),
re.form = ~0)

# Plot the interaction
plot_data <- data.frame(z = z_values, crosslvl = crosslvl)
ggplot(plot_data, aes(x = z, y = crosslvl)) +
  geom_line() +
  xlab("Moderator (z)") +
  ylab("Interaction Effect") +
  ggtitle("Interaction Effect Plot")
```

MODEL 2 selected and annotated output: cross-level interaction (i.e., moderator included)

	`Random effects:`
	` Groups   Name        Variance  Std.Dev. Corr`
$Var(\zeta_{0j})$ →	` stimulus (Intercept) 0.6133087 0.78314`
$Var(\zeta_{1j})$ →	`          x           0.0006328 0.02516  -1.00`
$Var(\varepsilon_{ij})$ →	` Residual             1.1613553 1.07766`
	`Number of obs: 1615, groups:  stimulus, 81`
	`Fixed effects:`
	`            Estimate Std. Error t value`
γ_{00} →	`(Intercept) -0.03145    0.09505  -0.331`
γ_{10} →	`x            0.06400    0.05383   1.189`
γ_{01} →	`z_c          0.65493    0.19510   3.357`
γ_{11} →	`x:z_c       -0.27451    0.11096  -2.474`

Note: R output for model 2 is shown in the text box on the right hand side, with annotations linking to the multilevel model equations in the main text shown on the left hand side. As a key parameter estimate, γ_{11} (or x:z_c) represents the cross-level interaction effect between the proxy-level moderator and the experimental effect.

Appendix B

Some considerations regarding power analyses and sample size

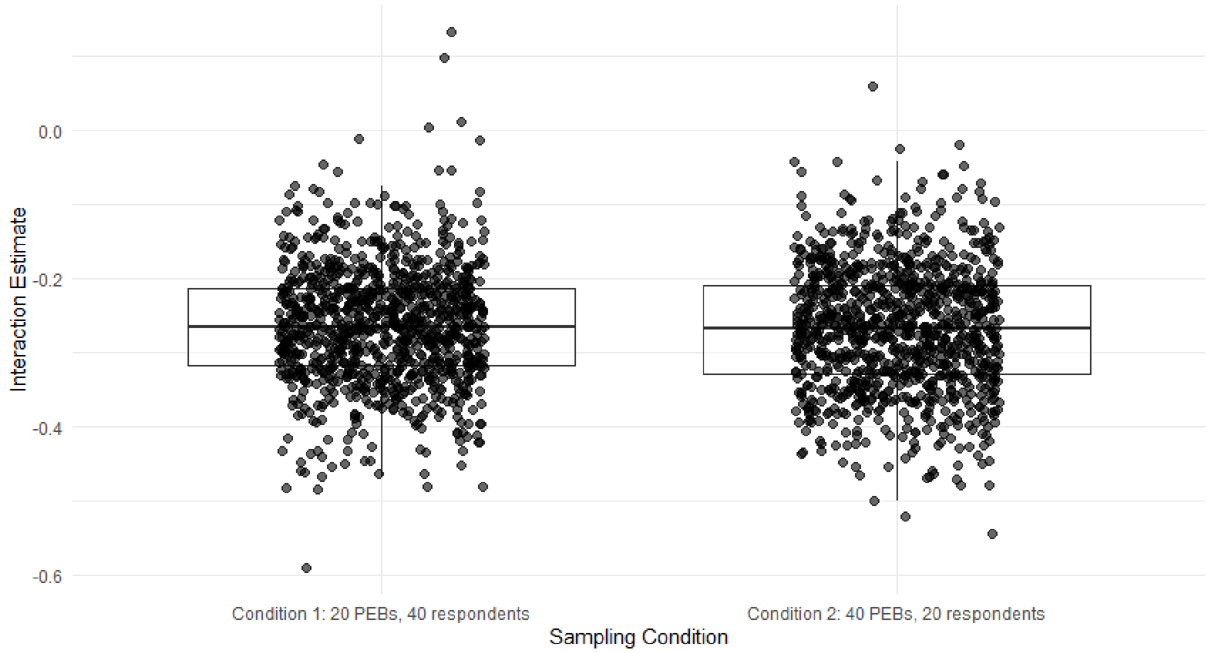
It is not fully clear how many participants need to be sampled for multi-proxy studies to have sufficient statistical power. For now, we can formulate some relevant considerations (but further research will be needed to clarify this matter). When first encountering the multi-proxy approach, experimentalists may intuitively assume that the sample size required for a multi-proxy experiment is equal to the sample size required for a traditional single-proxy experiment multiplied by the number of proxies. Fortunately, this is not the case.

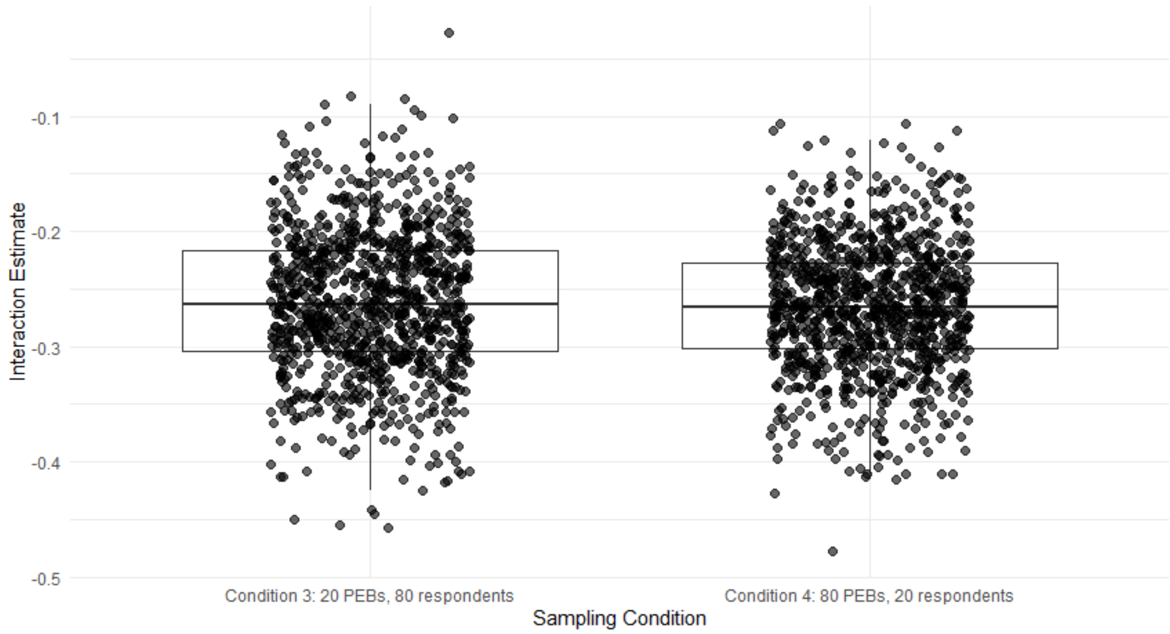
Traditional experiments often implicitly make the assumption that the effect of interest is the same regardless of which proxy is being used (by formulating conclusions at the level of the construct, not its specific operationalization). If this assumption were correct, the sample size required for running an experiment in the novel multi-proxy paradigm would be identical to that required for a traditional experiment. For instance, imagine that choosing tap water (over bottled water) would be the proxy for PEB in a traditional experiment that requires a subject sample of N = 100 to reach 80% statistical power for finding an effect with alpha = .05. If other proxies were perfectly interchangeable with choosing tap water as proxies for PEB, a multi-proxy experiment with, say, 50 different stimuli, would also require a subject sample of N = 100 (in total) to reach 80% statistical power for finding an effect with alpha = .05. The reason is that under the assumption of perfect interchangeability of proxies, the different proxies would not introduce additional variance.

Things turn more complicated, however, if this assumption of proxy interchangeability does not hold. Here we provide some initial rules of thumb. As the multiproxy experiment paradigm is new, there is limited experience to draw from for informing sample size decisions. We therefore ran a preliminary, small-scale power analysis informed by parameter estimates from study 2 of Millet et al. (2022). Specifically, we used the parameter estimates (also reported in Appendix A.S3 and A.S4) as population values to generate simulated population data from which we can then sample using alternative sampling strategies; specifically, we generate data for N = 16000 individuals nested in 400 PEB proxies. We then estimated the cross-level interaction effect (where a PEB attribute moderates the experimental effect at the individual level) in four sampling conditions: (1) 20 PEBS with 40 individuals each (N = 800), (2) 40 PEBs with 20 individuals each (N = 800), (3) 20 PEBs with 80 individuals each (N = 1600), (4) 80 PEBs with 20 individuals each (N = 1600). For each condition, we took 1000 samples, estimated the interaction parameter, and checked how many p-values were lower than .05. In the two conditions with N = 800, power was .84 for condition 1 (i.e., 20 PEBS with 40 individuals each; N = 800), and .91 for condition 2 (40 PEBs with 20 individuals each; N = 800). In the two conditions with N = 1600, power was >= 95% (i.e. for nearly all samples the interaction effect turned out to be significant), specifically it was .96 in condition 3 (i.e., 20 PEBS with 80 respondents each) and .997 in condition 4 (i.e., 80 PEBS with 20 respondents each). Comparing condition 2 to condition 1 and condition 4 to condition 3 for now suggests that it is probably slightly more useful to try and sample many proxies (rather than try and maximize the number of respondents per proxy). Figure 1 reports the interaction parameter estimates by sampling condition, illustrating the substantive variation in estimates across samples, albeit mostly within an acceptable range (and with mostly statistically significant results, given the power estimates).

Figure 1.

Interaction estimate by sampling condition





In conclusion, our preliminary power analysis provides useful insights into the sample size requirements for multi-proxy experiments. Most importantly, these experiments do not simply multiply the sample size requirements of traditional single-proxy studies. The simulation results suggest that, for an adequately powered multi-proxy experiment, increasing either the number of proxies or the number of respondents per proxy can improve power. Even with a moderately large total sample size (N = 800), power reached acceptable levels (over .80). Increasing the sample size to N = 1600 resulted in very high power (>= 96%) for the effects found in the empirical study we used as an illustration. We need further research to establish more precise guidelines for sample size decisions in multi-proxy experiments.

Appendix *B.S1*: R syntax power analysis

```
### Power analysis simulation study for the multi-proxy approach #####
# Load required libraries
library(lme4)
library(lmerTest)
library(ggplot2)

############### GENERATE DATA ################################
# Set parameters based on Appendix Letter JEVP
g_00 <- -.030   # Intercept
g_01 <- .655  # Main effect of z
g_10 <- .063   # Mean slope
g_11 <- -.275   # Interaction effect of z
sd_zeta_0 <- sqrt(.598) #standard deviation of zeta_0, residual of
intercept
sd_zeta_1 <- sqrt(.006) #standard deviation of zeta_1, residual of slope
sd_epsilon <- sqrt(1.159) #standard deviation of epsilon, residual of y

# Create simulated dataframe with identifiers and IVs
# create dataframe simdata with 400 PEBs, 400 respondents per PEB,
experimental dummy x, y_control, y_treatment
simdata <- data.frame(
                      PEB = rep(1:400, each = 400),
                      id = (1:160000),
                      id_rep = rep(1:400, times = 400),
                      z = rep(rnorm(400, 0, 1), each = 400),
                      zeta_0 = rep(rnorm(400, 0, sd_zeta_0), each = 400),
                      zeta_1 = rep(rnorm(400, 0, sd_zeta_1), each = 400),
                      x = rep(0:1, 160000),
                      epsilon = rnorm(160000, 0, sd_epsilon),
                      y = rep(NA,16000)
                      )
# Compute y variable
simdata$y <- g_00 + g_01 * simdata$z +
  g_10 * simdata$x + g_11 * simdata$z * simdata$x +
  simdata$zeta_0 + simdata$zeta_1 * simdata$x +
  simdata$epsilon

############# SIMULATION STUDY ###############################

# Define function to run simulation for a specific condition
run_power_simulation <- function(n_proxies, n_respondents_per_proxy,
n_simulations = 1000) {
  interaction_estimates <- numeric(n_simulations)  # to store interaction
estimates
  p_values <- numeric(n_simulations)  # to store p-values

  for (i in 1:n_simulations) {
    # Randomly sample PEBs and respondents per PEB
    sampled_pebs <- sample(unique(simdata$PEB), n_proxies, replace = FALSE)
    sampled_data <- simdata[simdata$PEB %in% sampled_pebs, ]

    # Sample the respondents per PEB
    sampled_data <- do.call(rbind, lapply(sampled_pebs, function(peb) {
      sample_respondents <- sample(sampled_data$id_rep[sampled_data$PEB ==
peb], n_respondents_per_proxy)
      sampled_data[sampled_data$PEB == peb & sampled_data$id_rep %in%
sample_respondents, ]
    }))
```

```
    # Fit the multilevel model
    model <- lmer(y ~ x * z + (1 + x | PEB), data = sampled_data)

    # Store the interaction effect and p-value
    interaction_estimates[i] <- fixef(model)['x:z']
    p_values[i] <- summary(model)$coefficients['x:z', 'Pr(>|t|)']
    # Access p-values using lmerTest
  }

  # Calculate the proportion of significant results (p-value < 0.05)
  power <- mean(p_values < 0.05)

  return(list(power = power, estimates = interaction_estimates, p_values =
p_values))
}

# Condition 1
n_proxies_cond1 <- 20
n_respondents_per_proxy_cond1 <- 40
results_cond1 <- run_power_simulation(n_proxies_cond1,
n_respondents_per_proxy_cond1)

# Condition 2
n_proxies_cond2 <- 40
n_respondents_per_proxy_cond2 <- 20
results_cond2 <- run_power_simulation(n_proxies_cond2,
n_respondents_per_proxy_cond2)

# Compare the power of both conditions
cat("Power for Condition 1 (20 PEBs, 40 respondents):",
results_cond1$power, "\n")
cat("Power for Condition 2 (40 PEBs, 20 respondents):",
results_cond2$power, "\n")

################## Box plots with estimates #################
interaction_estimates_cond1 <- results_cond1$estimates
interaction_estimates_cond2 <- results_cond2$estimates

# Combine data for plotting
estimates <- data.frame(
  Estimate = c(interaction_estimates_cond1, interaction_estimates_cond2),
  Condition = factor(rep(c('20 PEBs, 40 respondents', '40 PEBs, 20
respondents'), each = 100))
)

# Boxplot of interaction estimates
# Combined boxplot and scatter plot with jitter
ggplot(estimates, aes(x = Condition, y = Estimate)) +
  geom_boxplot(outlier.shape = NA, alpha = 0.5) +
# Boxplot without outliers
  geom_jitter(width = 0.2, height = 0, alpha = 0.6, size = 2) +
# Add jittered dots
  theme_minimal() +
  labs(title = "Interaction Estimates by Sampling Condition",
       y = "Interaction Estimate", x = "Sampling Condition")
```