



Research Paper

Mismatch between measures of “greenness”: Evidence from a field experiment on self-reports and observed behavior

Ann-Kathrin Blankenberg ^a, Martin Binder ^{b,c,*}, Israel Waichman ^{a,c}^a Bard College Berlin, Platanenstr. 24, 13156, Berlin, Germany^b Fakultät für Staats- und Sozialwissenschaften, Universität der Bundeswehr München, Werner-Heisenberg-Weg 39, 85577, Neubiberg, Germany^c Levy Economics Institute of Bard College, Blithewood, Bard College, 12504-5000, Annandale-on-Hudson, New York, USA

ARTICLE INFO

Keywords:

Pro-environmental preferences

Product choice

Donations

Prosocial preferences

ABSTRACT

We run a field experiment with $n = 599$ participants selected from the waiting room of a town hall of a medium-sized German town that demonstrates a clear mismatch between three measures of “greenness” frequently used by researchers: self-reports, donations to environmental charities, and product choice experiments. We find that an aggregate index of self-reports on 20 different past pro-environmental behaviors only weakly correlates ($r = .09^*$) with an observed pro-environmental product choice. Our preferred own money donation task also exhibits only weak correlation with both a self-report index ($r = .15^*$) and the product choice task ($r = .17^*$), but as our experimental variation shows, this is only the case if the donation paradigm is modified to allow sorting between prosocial and environmental charities. Conversely, we show that correlations disappear for the donation task if it is structured without a competing non-environmental prosocial option. Our results are unaffected by whether the respondent is a student or working professional and robust to different ways of aggregating individual self-report items. Our results support the concern that pro-environmental behaviors are difficult to measure consistently with the instruments that are to-date most commonly applied by researchers, but that there is some room to modify experimental designs to produce measures that yield more correlated outcomes across instruments.

1. Introduction

Preventing climate change and natural resource depletion is among the biggest challenges the world faces nowadays, not the least because of ecological tipping points that could lead to abrupt and irreversible consequences (e.g., [Delmotte et al., 2018](#); [Lenton et al., 2019](#)). Researchers seek to find and analyze institutions that would prevent or reduce climate change and lead to sustainable growth ([Kotchen, 2009](#); [Wagner, 2016](#); [Wolff, 2014](#); [Requate et al., 2019](#); [Fraser, 2023](#)). To successfully design such institutions and to generally gauge the support of the populace for climate change action, the accurate measurement of pro-environmental preferences and behavior is important.

To measure pro-environmental preferences, the empirical literature often uses surveys with self-reports on pro-environmental behavior ([Kormos and Gifford, 2014](#); [Brick et al., 2017](#); [Lange and Dewitte, 2021](#); [Bosshard et al., 2024](#)). The reason is convenience. It is relatively easy and cheap to administer a survey, but much more difficult to observe and measure actual pro-environmental behavior

* Corresponding author.

E-mail addresses: a.blankenberg@berlin.bard.edu (A.-K. Blankenberg), martin.binder@unibw.de (M. Binder), i.waichmann@berlin.bard.edu (I. Waichman).<https://doi.org/10.1016/j.jebo.2025.107347>

Received 30 August 2023; Received in revised form 7 November 2025; Accepted 18 November 2025

Available online 29 November 2025

0167-2681/© 2025 The Author(s).

Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Published by Elsevier B.V. This is an open access article under the CC BY license

from which one could infer pro-environmental preferences. Self-reports hence are popular in the field (Koller et al., 2023; Lange et al., 2023; Lange, 2023, p. 600). However, despite their popularity, self-reports of pro-environmental behavior only poorly track actual behavior (Kormos and Gifford, 2014; Lange and Dewitte, 2021; Koller et al., 2023; Nielsen et al., 2024), and in consequence, research involving them leads often to fundamentally different results than research focusing on actual behavior in incentivized settings (e.g., Galizzi and Navarro-Martinez, 2019; Klein and Hilbig, 2019; Lange et al., 2023). This is potentially due to social desirability bias and other reporting biases in self-reports (Koller et al., 2023).

These results reinforce a certain skepticism that economists have vis-à-vis “stated preference” measures, seeing them at the extreme as “cheap talk”. Should we then better measure actual pro-environmental behavior? In some obvious sense, behavior should track behavior better. But observing behavior is more costly and difficult to set up, and there is also evidence that behavior elicited in contrived lab situations does not predict the corresponding naturally occurring behavior well (Galizzi and Navarro-Martinez, 2019). But more fundamentally, one can ask whether observing actual (pro-environmental) behavior would indeed reveal a pro-environmental preference. In economics, observed behavior is often assumed (implicitly) to reveal a corresponding preference (compare the theory of revealed preference). For example, a donation of money to an environmental charity is a popular measure in experiments not to gauge the extent of pro-environmental donations, but rather to draw an inference about individuals’ pro-environmental preferences.

Our paper questions this inference. In the same way that a person may be observed to walk to work but does not do so due to a pro-environmental preference but to save money (compare Binder and Blankenberg, 2025), another person may be observed to donate money to an environmental charity but may do so for reasons that are not pro-environmental but prosocial. In both cases, their behavior may be pro-environmental, but the behavior may not reveal, or be indicative of, a pro-environmental preference.

The contribution of our paper is twofold: Firstly, in our field experiment, we demonstrate the mismatch between three measures of “greenness” (self-reports of pro-environmental behavior, donations to environmental charities and a variant of pro-environmental product choice) in a natural field setting. As hypothetical choice scenarios or using “house money” have been shown to bias such measures (e.g., Koller et al., 2023; Moser et al., 2013), our study uses an incentivized setting and naturally occurring behavior: if they wish to donate, participants must donate their own money and are not aware that the donation and product choice happen as part of an experimental study. We also use a general population sample with higher external validity than the standard student samples. Yet, our sample contains enough students in it to allow us to also make comparisons between students and non-students of the participant population. In our study, we find that an aggregate index of self-reports on 20 different past pro-environmental behaviors, an aggregated stated preference measure if you will, only weakly correlates ($r = .09^*$) with observed pro-environmental product choice (choosing a notepad made of recycled paper over a white paper notepad or refusing to choose a notepad for explicitly pro-environmental reasons).

Secondly, we show that observed donations better align with the aggregate index of pro-environmental behavior self-reports conditional on giving individuals the choice to explicitly donate for non-environmental causes as well. In our experimental setup, we offer two choice variations: in variation 1, if individuals wish to donate at all, a choice of either environmental or prosocial charity is offered (“charity choice condition”). In consequence, the correlation between donation to environmental charity and aggregate index of pro-environmental behavior self-reports in this group is larger than in the overall sample and statistically significant ($r = .15^*$). In addition, only in this condition, observed donations to an environmental charity are also significantly associated with observed product choice ($r = .17^*$). In variation 2, donations are possible only to an environmental charity (“no choice condition”). Here, observed donations are not significantly related to the aggregate index of pro-environmental behavior self-reports ($r = .01ns$) and neither to observed product choice ($r = -.04ns$). In essence, offering to prosocial individuals (who do not care specifically about the environment) the possibility to self-select into donating to a non-environmental charity will lead to a situation where observed donations to the environmental charity are more closely aligned with the other measures of “greenness”. This latter finding is the constructive part of our paper in that it implies that if one wants to use donations to environmental charities to reveal (or be indicative of) pro-environmental preferences, existing paradigms should be modified to allow for the donation to also go to prosocial but non-environmental causes.

Our experimental setup is thus designed to disentangle pro-environmental and prosocial preferences: while there is conceptual work separating both (e.g., Steg and De Groot, 2012; Stern, 2000), empirical work on their separability is scarce (but see De Groot and Steg, 2007, 2008; Klein et al., 2017, only the latter source shows this in a non-hypothetical setting). To bolster this argument, we also correlate our “greenness” measures with self-reports of prosociality (the social value orientation measure, see Murphy et al., 2011). In effect, observed donations here correlate significantly with the self-reports of prosociality (and observed product choice does not), whereas observed product choice correlates significantly with the aggregate index of pro-environmental behavior self-reports (but not the self-reports of prosociality). This correlation pattern and the results from the above experimental variation are consistent with the interpretation that behavior such as the above-mentioned pro-environmental product choice may more unambiguously reveal a pro-environmental preference. On the other hand, our results suggest that observed donations may not be good indicators for pro-environmental preferences, unless one modifies the donation task paradigm to include the option to also donate to prosocial charities.

The paper is structured as follows. In Section 2, we review the literature on which we have based our hypotheses. We then describe the design strategy and implementation of our field study in Section 3. The data is described in Section 4. Section 5 contains the results of our analysis and Section 6 discusses the results and limitations of our study and concludes.

2. Related literature and hypotheses

The term “pro-environmental behavior” (PEB) is used in ambiguous ways throughout the literature (see extensively [Lange, 2024](#)). From an impact-oriented point of view, “pro-environmental behavior” refers to any behavior that “provides relative benefits to the natural environment” compared to other behaviors ([Lange, 2024](#), p. 1). Intent-oriented definitions additionally also require a pro-environmental intent (or motivation), for instance defining behavior as pro-environmental “that consciously seeks to minimize the negative impact of one’s actions on the natural and built world” ([Kollmuss and Agyeman, 2002](#), p. 240). There are three approaches to measure pro-environmental behavior, with most studies focussing on one type of measure (for a review, see, e.g., [Lange et al., 2023](#); [Lange and Dewitte, 2019](#)): firstly, pro-environmental behavior can be observed in naturally occurring situations. Secondly, it can be observed in contrived or stylized lab or field tasks that are systematically arranged so as to let participants know about the experimental situation (e.g., mock recycling task, intergenerational public good games being played for small amounts of money, etc.). Thirdly, and due to its low cost most prominently, information on pro-environmental behavior can also be gained via self-reports, where individuals are asked to rate the extent to which they have (over some past time) committed the pro-environmental behavior in question.¹ These self-report items then are often aggregated into multi-item scales (e.g., [Binder and Blankenberg, 2017](#); [Brick et al., 2017](#); [Kaiser, 1998](#)), but what items are aggregated is often driven by data availability (for example in big household panel surveys, compare [Poortinga, 2022](#)) and few of these multi-items scales are validated ([Bossard et al., 2024](#); [Lange, 2024](#)).

But what is actually measured by these observed behaviors, behavior self-reports and their aggregates? Much of the ambiguity in the use of the term “pro-environmental behavior” comes from researchers referring with it to either the behavior or some underlying attitude or preference motivating the behavior ([Lange, 2024](#)). This way, both an observation or a self-report (or its aggregate) may be meant to capture either a property of the actual behavior, such as its frequency (“a behavioral property”, *ibid.*), or a psychological property of the person who commits the behavior (“a person property”, *ibid.*), such as that person’s pro-environmental preference (or some other pro-environmental attitude more generally).² In economics, observed behavior is often interpreted to reveal an individual’s preference (compare the theory of revealed preference), so that observed pro-environmental behavior would then be used to infer about an individual’s pro-environmental preferences. In analogy to this, self-reports of behavior are often considered to be “stated preference” measures, and this pertains specifically to multi-item scales composed of pro-environmental behavior self-reports ([Lange, 2024](#), pp. 4–5), which are typically considered to represent some underlying generalized “greenness” attitude/preference (e.g., [Binder and Blankenberg, 2017](#); [Brick et al., 2017](#); [Kaiser, 1998](#)).

Different measurements of observed behaviors, individual self-reports, and aggregates of self-report scales composed of pro-environmental behavior items will not only align better or worse with the different definitions of pro-environmental behavior, but also with whether one aims to capture behavioral properties or person properties with the observed behaviors or self-reported behavior items. This mismatch is at the heart of our paper, as our first contribution to the literature demonstrates this in a fully incentivized natural field setting. Consider self-reports of pro-environmental behavior. They are comparatively cheap and easy to administer and hence very popular in research ([Koller et al., 2023](#); [Lange et al., 2023](#); [Lange, 2023](#), p. 600). But such self-reports only poorly track actual behavior ([Kormos and Gifford, 2014](#); [Lange and Dewitte, 2019](#); [Koller et al., 2023](#); [Nielsen et al., 2024](#)), with correlations between self-reports and actual behavior being small to medium ([Kormos and Gifford, 2014](#); [Lange and Dewitte, 2021](#)) and “functionally small” in the sense that self-reports can only explain a small portion (21 %) of the variance of actual behavior, as shown in an influential meta-analysis ([Kormos and Gifford, 2014](#), p. 367). Given this, it is not surprising that research involving self-reports often leads to fundamentally different results than research focusing on actual behavior in incentivized settings (e.g., [Galizzi and Navarro-Martinez, 2019](#); [Klein and Hilbig, 2019](#); [Lange et al., 2023](#)).³ This leads us to hypothesize:

Hypothesis 1. A multi-item scale of pro-environmental behavior self-reports (as measure/indicator of pro-environmental preference) does not correlate strongly with observed pro-environmental behavior.

The above mismatch between self-reports and observed behaviors has led some researchers (and especially economists) to prefer observing actual behaviors, such as observing travel modes ([Mayer and Geller, 1982](#); [Lange and Dewitte, 2019](#); [Kaiser et al., 2021](#)), sorting and counting waste and recycling behavior ([Knutsson et al., 2013](#)), as well as more stylized behaviors in lab situations, including turning off lights in the room after the experiment or choosing between new or recycled products ([Cornelissen et al., 2008](#)), disposing correctly of waste, or donating money to environmental charities (e.g., [Carrico et al., 2018](#); [Ibanez et al., 2017](#)). While there are some attempts at validation of lab tasks ([Lange, 2023](#), p. 610–17), many of these are still (ad hoc) measures that are easy to elicit.

¹ In some cases, instead of self-reported past behaviors, self-reported pro-environmental behavior *intention* is elicited, i.e. individuals are professing that they intend to execute certain pro-environmental actions in the future (on the problems regarding this, see e.g., [Klein and Hilbig, 2019](#); [Lange, 2023](#), p. 600).

² An attitude is a general cognitive judgement towards some object, whereas a preference in economics typically is seen as a binary relation between two objects, i.e. the individual prefers a over b. Opposed to an attitude, one could also hold that the preference contains a motivational component in addition to the cognitive component, i.e. the individual holding a preference strives to fulfill it. We note that usage between disciplines may vary, note for instance the huge but not complete overlap between theory of planned behavior and the economic preference-belief-constraint model of human behavior. For our purpose, potential finer distinctions between attitudes and preferences may not be necessarily relevant.

³ Reasons for these discrepancies have been attributed to a range of imperfections and biases in reporting, including lack of knowledge (e.g., when asked to estimate one’s ecological footprint, see [Bleys et al., 2018](#)), hypothetical bias (as no real costs are associated with the self-reports, see, e.g., [Moser et al., 2013](#)), own pro-environmental identity ([Binder and Blankenberg, 2017](#)), as well as biased responses to due social norms and social desirability (see more extensively [Koller et al., 2023](#)).

Especially donations to environmental charities have become popular both in laboratory and field experiment situations in economics (e.g., [Alpizar et al., 2008](#); [Arendt and Matthes, 2016](#); [Carrico et al., 2018](#); [Galizzi and Navarro-Martinez, 2019](#); [Ibanez et al., 2017](#)). Donations are particularly easy to implement in the lab as individuals get compensated for showing up or completing certain tasks, and hence have money at hand for a potential donation. A recent review of donations in an environmental context finds that only two of the studies surveyed require donations with own money (see [Lange, 2023](#), pp. 604–6) and this introduces a “house money” effect similar to “hypothetical bias” attached to self-reports ([Moser et al., 2013](#)). Playing with house money may be one reason why there is also mounting evidence that behavior elicited in contrived lab situations (such as social dilemma games) does not predict naturally-occurring behavior well. In their study on prosocial behavior, [Galizzi and Navarro-Martinez \(2019\)](#) find that both self-reports of altruism and lab behavior in dictator games did not predict actual altruistic behavior outside of the lab (which included donations to an environmental charity). This motivates our setup of a donation task, where individuals need to donate their own money in a fully incentivized and fully consequential design with real gains and losses.⁴

Our second contribution to the literature then pertains to observed donations and relates to the above ambiguity of using behavior to infer either behavioral or person properties. We ask whether observed donations will actually reveal a pro-environmental preference, or rather be indicative of non-environmentally focused prosociality (with the exception of [De Groot and Steg, 2008](#), this has been by and large ignored in the literature).⁵ While parts of the literature treat pro-environmental behavior (and preferences) as a form of prosociality (compare [Nolan and Schultz, 2015](#); [Turaga et al., 2010](#); [Venhoeven et al., 2013](#)),⁶ research in environmental psychology provides evidence that pro-environmental and prosocial preferences are distinct and separable ([Stern, 2000](#); [De Groot and Steg, 2007](#); [Steg and De Groot, 2012](#)): in a survey study, [De Groot and Steg \(2008\)](#) have shown that higher self-reported prosociality was linked to higher intention to donate to non-environmental causes, but not to pro-environmental causes, and stronger pro-environmental preferences were linked to higher pro-environmental behavior intentions, but to lower prosocial behavior intentions. This suggests that when pro-environmental and prosocial preferences are in conflict, using measures that may conflate both might be problematic (p. 343). Such conflict between prosocial and pro-environmental preferences has also been demonstrated in a lab experiment, where in-group cooperation (prosociality) was put into a trade-off against pro-environmental choices ([Klein et al., 2017](#)).

If pro-environmental preferences cannot be subsumed as prosocial preferences, the low correlation between a multi-item scale of pro-environmental behavior self-reports (as stated preference measure) and observed donations would be due to the fact that (some) donations to a pro-environmental cause may be motivated by prosocial but not pro-environmental preferences. In consequence, observing donation behavior may not reveal a pro-environmental preference (in contrast, pro-environmental product choice would not conflate prosocial and pro-environmental preferences). Since there are no studies in the literature using observed donations that offer a possibility to choose between pro-environmental and prosocial charities, we have designed an experiment that allows prosocial individuals to (alternatively) donate to prosocial causes in the first choice variation (“charity choice condition”), whereas the second choice variation (“no choice condition”) mirrors current practice in the literature and allows only donations to an environmental charity. For this setup, we hypothesize:

Hypothesis 2. Donations to pro-environmental causes correlate more strongly with a multi-item scale of pro-environmental behavior self-reports, and hence reveal more unambiguously pro-environmental preferences, when there is a possibility to choose between environmental and social charities (variation 1: “charity choice condition” vs. variation 2: “no choice condition”).

3. Field experiment design

In order to test the two hypotheses developed above, we have set up a field experiment that was conducted on account of, and with the cooperation of the municipal authority of the medium-sized German town of Goettingen. It consists of a short survey on pro-environmental behaviors and attitudes administered to willing individuals, who were appearing at city hall on the various errands that brought them there. As part of the experiment, we also observed two (potentially) pro-environmental behaviors from our participants. We did not announce the collection of those two behavioural measures, and we designed the experiment so that both behaviours would be perceived by our participants as naturally occurring behavioral opportunities rather than visibly arranged tasks in a study or lab situation.⁷

For the first observed pro-environmental behavior, unbeknownst to the participants, we have recorded their choice of a notepad and pen offered to take notes during the survey and as small thank-you gift (recycled/brown paper notepad vs. white notepad, in

⁴ That this is by no means standard has also been noted in a recent review in consumer research ([Klein and Hilbig, 2019](#)), where most studies on consumer choice were either hypothetical or only implemented real positive consequences (gains), whereas only six percent of studies had a fully consequential design, i.e. participants could both make real gains as well as losses. Considering that many decisions about pro-environmental behavior entail real costs to the decision-maker, this reinforces the need for studies where participants make donations (or other decisions) not with house money but their own money (for examples, see [Alpizar et al., 2008](#); [Moser et al., 2013](#)).

⁵ One could question whether a single behavior can plausibly reveal a general preference such as the ones in question here, or whether this at most can be understood to be an indicator for such a preference (compare [Lange, 2024](#)). We will discuss this question later on, but note for now that this inference, both from self-reports as well as from behaviors such as donations is often exactly what is made in the literature, explicitly or otherwise.

⁶ This is also the case for [Galizzi and Navarro-Martinez \(2019\)](#), as their study on *prosocial* behavior offered their participants one field situation, where donations could be made to an *environmental* charity.

⁷ Potential participants were informed about the survey and assured anonymity, but were not informed about the experimental branch of the study, i.e. variation of donation choices as well as recording of their product choice, as this would have defeated the study aim.

essence a pro-environmental product choice task; compare Cornelissen et al., 2008). The second observed behavior was a potential donation to a charity, which participants were asked to consider independent of taking the survey (more on this below).⁸

Experimental variation happened with regard to the donation behavior, where participants were given the opportunity to donate some of their *own money* to charity. We have varied the potential donation options, so that a choice of charity between environmental vs. non-environmental (but social) charity was offered in the “charity choice condition” (variation 1), whereas there was no choice of charity possible in the “no choice condition” (variation 2), but simply the option to donate to an environmental charity (of course, in both conditions, one could choose to not donate at all). To be precise, both options included a number of charities towards which the money would go, with the only choice being between either a prosocial or a pro-environmental bundle of charities/activities. Importantly, if money was donated, this money had to be the participants’ own money, i.e. there was no show-up fee given out first, from which to make the donation).⁹

To encourage participation for the survey, apart from the pen and notepad, all respondents participated in a lottery after finishing the survey, where they were handed gift vouchers worth between 5 and 25 Euros (everyone was guaranteed a minimum gift voucher value of 5 Euros). For participants, this compensation was expressly for the survey, as the experimental variation was not visible to them. No exclusion criteria were used in the recruitment of participants except for ability to understand German language during the recruitment process.

In more detail, our field experiment proceeded during five days in October/November 2019 in the following steps:¹⁰ Firstly, individuals came to city hall with any official business unknown to the experimenters and drew a numbered ticket for the waiting room queue, waiting to be called for their appointment. The average waiting time for the appointment after drawing the number was about 20–30 mins.

When drawing a ticket, a research assistant, wearing a simple t-shirt with a logo of city hall, invited them to participate in the survey, briefly mentioning that the survey is a part of the municipality’s activities and promising a small thank-you gift for participation. Research assistants also mentioned that independently of the survey, *voluntary* donations to the pro-environmental (and social) campaigns of the municipality would be collected and that participation in the survey would take up roughly 15 minutes of the participants’ time. After this, the potential participants were asked whether they would be willing to participate in the survey. If the individual agreed to participate in the survey, they were guided to a designated area in the waiting room. If no agreement was expressed, or the person approached did not speak German to a sufficient degree, the experiment ended for them.

In a second step, individuals willing to participate were welcomed in the designated area, which also contained informational posters on the environmental charities (or environmental and social charities) to provide further information to potential donors. The informational posters contained basic information on the concrete projects that the municipality supported in their campaigns and towards which a potential donation would then go. Participants were given their survey documents and then asked whether they would be willing to donate (“Besides the survey, would you like to make a small contribution to the climate campaign [or social activities] in Goettingen?”). This was noted by the experimenters after the participant moved on. Irrespective of whether individuals indicated a willingness to donate, they were handed an empty envelope in addition to the survey document and shown a voting cabin that would ensure anonymity while putting money into the envelope or not, which would then be left in a sealed urn. If individuals did not indicate willingness to donate, they were handed an envelope as well and asked to simply put it in empty (i.e., in the same fashion as when willing to donate). This was explained to ensure anonymity regarding the donation decision. For us, the procedure was meant to ensure avoiding social pressure on the donation decision, but also to allow matching donations to survey responses, as both envelope and survey document were marked with unobtrusive matching numbers. In addition, the following text to choose the contribution target was printed on the envelope: “Please tick: I give my contribution ... [] for the ‘climate protection activities’ ... [] for the ‘social activities’.” Both choices were printed on the envelopes for the “charity choice condition” (and option order randomized among envelopes), whereas in the “no choice condition” envelopes only had the ‘climate protection activities’ option to tick and no reference to a second option.

In a third step, after handing out survey document and envelope, participants were then offered a pen and notepad as thank-you gift for participation and to take notes during the survey (“Please also take a notepad and pen. You may keep both.”). Both types of notepads were displayed openly for participants to choose from. A non-negligible number of participants opted to not pick a notepad (or return their pen and notepad) citing expressly pro-environmental reasons, of which the research assistants took systematic note.

In step four, participants had all materials and were asked to first drop off the donation envelope and then fill in the survey in the regular waiting area. They were informed that the survey itself would take roughly ten minutes to fill in and that after handing it back, they would receive a thank-you gift voucher based on a small lottery, where the minimum win was 5 Euros (and some vouchers were up to 25 Euros). If participants had questions regarding procedure or donation campaigns, research assistants answered them during these steps.

⁸ Experimental procedures were approved by the municipal authority of Goettingen and all donation money was collected in a tamper-proof collection box. Donations were then counted and handled by two employees of city hall and were used for their designated charitable purposes (counting was supervised by one of the authors). Money to finance the vouchers was provided by the municipal authority.

⁹ Given the different opening hours and days of the various services provided at city hall, we have rotated the two variations in treatments in the following way: in week one we offered the “no choice condition” on Monday and Tuesday. In week two, on Monday and Tuesday, we offered the “charity choice condition”. On the final day five, we varied conditions, offering each for half a day, in order to achieve balance in observations between both variations. Balance tests in Online Appendix 2 suggest that this randomization procedure worked.

¹⁰ Precise wordings for the experimental and survey task are provided alongside other replication materials in the Online Appendix to be found at <https://osf.io/zcdqw/>.

In step five, after putting money in the envelope (or leaving it empty) in private, and checking on it a box to which campaign the donation should go in the case of the “charity choice condition”, the envelope was put into a tamper-proof donation box by the participants, who then proceeded to a separate waiting area where they filled in the questionnaire while waiting to be called for their appointment with the city.

Lastly, in step six, when done, participants returned the questionnaire. After turning in the survey document, participants drew a number and were handed their promised gift voucher. Some individuals chose to donate the gift voucher, which we have recorded as well, but will not use in the main analysis, as it defeats the aim of making donations fully consequential, i.e. having individuals not play with house money. After this step, the experiment ended for our participants.

Each participant executed all tasks (choice of notepad, opportunity to donate, survey, although a number of surveys were partly incomplete, see data description below). During design, we have aimed at an overall number of 600 participants in accordance with experience on typical sample sizes for economic experiments and to be able to detect small correlations.¹¹ Refusal rates were quite low (15 %),¹² which we attribute mainly to two factors: participants did not express any surprise at the municipality eliciting a survey, as it appeared to be a normal activity amongst many initiatives the town conducted or supported. This supports our contention of having created naturally-occurring possibilities for individuals to act in pro-environmental ways. Secondly, the waiting area had a shortage of chairs to wait for the appointments (not by our design), and individuals saw the additional chairs in the designated survey area when drawing their waiting room queue number. We suspect that this may have incentivized also the less prosocial individuals in participating in our experiment. In any case, the additional voucher lottery after the survey led to many participants expressing surprise in being “compensated” for their participation in the survey.

4. Data set and descriptive statistics

A total of $n = 599$ individuals participated in our field experiment.¹³ We have collected information on our participants both through the survey, as well as via observed behaviors regarding donation decision and product choice (noted by the research assistants conducting the study). Table 1 shows descriptive statistics of our variables of interest for the full sample and separately for both donation choice variations. We discuss both behavioral variables and survey variables in turn.¹⁴

4.1. Observed behavior

Our main behavioral outcome variables are pro-environmental product choice and donation decision (and amount). Regarding pro-environmental product choice, all participants were offered the choice between an environmentally-friendly (brown paper/recycled) or regular (white paper/non-recycled) notepad (plus pen) as thank-you gift for participating in the survey, and ostensibly to take notes while filling in the questionnaire. 42 % of the participants chose the environmentally-friendly notepad, which we have coded in the form of a dummy variable for pro-environmental behavior (“Behavior: Recycled notepad chosen (y/n)”). An unanticipated result observed in the experiment is that a non-negligible number of participants either returned the notepad after filling in the survey or refused both pen and notepad citing in both cases explicitly environmental considerations. As these behaviors can be clearly seen as a form of pro-environmental choice, and have been systematically recorded by the research assistants, we have post-hoc decided to create a more comprehensive pro-environmental behavior dummy variable (“Behavior: Pro-environmental choice (y/n)”), that equals one in case of recycled notepad chosen, or notepad returned or notepad refused for explicitly pro-environmental reasons (the variable is coded as zero for refused or returned notepads without a clear pro-environmental rationale provided to the research assistants). 53 % of the participants acted pro-environmentally according to this broader definition.

As regards donation decision, all respondents were informed about the possibility to donate some money, either to an environmental charity (choice variation 2: “no choice condition”, $n = 299$), or they had a choice between environmental or social charity (choice variation 1: “charity choice condition”, $n = 300$). We have created a dummy variable that captures whether a donation was made (“Behavior: Donated money (y/n)”; 28 % of all participants donated money) as well as a continuous variable capturing the “Behavior: Amount of money donated”. Some individuals did not choose between charities or split their donation to both charities, both of which

¹¹ Regrettably, we did not pre-register the study and its design beforehand. Sensitivity power analysis informs us that $n = 600$ is sufficient to detect correlations bigger than $r > .11$ with 80 % power and an alpha level of 5 %. This n is close to the typical “small effect” effect sizes label of $r = 0.1$ given by Cohen (1988), but this too is only an approximation, as what should be considered a small or medium or large effect size surely varies by research field (Lakens, 2022). If we look at the meta-analysis by Kormos and Gifford (2014), an average correlation of $r = .46$ between single-item self-reports and the corresponding observed behavior was found, which the authors themselves considered as untypically large. Lange and Dewitte (2021) report “small to medium” ($r = .10 - .29$) correlations between self-reports and a behavioral lab measure (p. 5). Our study would be adequately powered to detect those. Given the minuscule correlations between prosocial behavior in the field and corresponding self-reports reported in Galizzi and Navarro-Martinez (2019), which ranged from $r = .02$ to $r = .09$, our study would have had to include $n = 19,617$ participants, which simply has not been feasible. However, if true effects are really much smaller than what our study is powered to detect, they might well be too small to matter substantively.

¹² We have no information on refusal rates for the first day of the experiment, but have no reason to suspect the refusal rate computed over the remaining days would differ, had we collected this information also for the first day.

¹³ One participant ran off with the questionnaire after making a donation.

¹⁴ With the exception of the income variable, we cannot reject null hypotheses of equal distributions of our variables across the two choice variations, suggesting that the randomization described above did work. Balance tests should not be over-interpreted as their informational value seems to be misunderstood (Deaton and Cartwright, 2018, section 1.1-1.3.). Online Appendix 2 contains the summary statistics for the balance tests.

Table 1

Summary statistics for the full sample and split by the two choice variations.

	(1) Full sample			(2) Variation 1: Charity choice condition			(3) Variation 2: No choice condition		
	mean	sd	count	mean	sd	count	mean	sd	count
Behavioral variables:									
Behavior: Pro-environmental choice (y/n)	0.53		595	0.55		298	0.51		297
Behavior: Recycled notepad chosen (y/n)	0.42		593	0.42		297	0.43		296
Behavior: Donated money (y/n)	0.28		599	0.27		300	0.29		299
Behavior: Amount of money donated (no vouchers)	1.26		599	1.19		300	1.33		299
Donation by type of charity (y/n)									
no donation (%)	0.72		599	0.73		300	0.71		299
environmental charity (%)	0.22		599	0.15		300	0.29		299
social charity (%)	0.04		599	0.07		300	0.00		299
no choice made (%)	0.02		599	0.05		300	0.00		299
Self-reports:									
Self-report: Aggregate index of PEB items	0.00		545	-0.04		270	0.04		275
Social Value Orientation Scale (y/n)									
Prosocial (%)	0.82		532	0.81		268	0.82		264
Individualistic (%)	0.15		532	0.16		268	0.14		264
Altruistic (%)	0.02		532	0.02		268	0.01		264
Competitive (%)	0.02		532	0.01		268	0.03		264
Simplified Social Value Orientation Scale (y/n)									
Individualistic or Competitive (%)	0.17		532	0.17		268	0.17		264
Prosocial or Altruistic (%)	0.83		532	0.83		268	0.83		264
Socio-demographic variables:									
Age (in years)	33.60		562	32.79		284	34.44		278
Gender									
Male (%)	0.46		586	0.48		296	0.43		290
Female (%)	0.54		586	0.51		296	0.57		290
Diverse (%)	0.00		586	0.01		296	0.00		290
Education level									
Other (%)	0.20		577	0.20		288	0.19		289
German Abitur (%)	0.42		577	0.40		288	0.44		289
Completed University (%)	0.38		577	0.40		288	0.37		289
Citizenship									
German (%)	0.90		580	0.91		293	0.89		287
Non-German (EU) (%)	0.03		580	0.03		293	0.03		287
Non-German (Non-EU) (%)	0.07		580	0.06		293	0.07		287
Income									
Less than 1000 Euros (%)	0.36		567	0.41		279	0.31		288
1000–1999 Euros (%)	0.22		567	0.20		279	0.24		288
2000–2999 Euros (%)	0.16		567	0.17		279	0.14		288
3000–4999 Euros (%)	0.16		567	0.13		279	0.18		288
5000–6999 Euros (%)	0.07		567	0.05		279	0.09		288
7000 Euros or more (%)	0.04		567	0.04		279	0.03		288
Employment									
Fulltime work (%)	0.27		555	0.26		281	0.28		274
Parttime work (%)	0.12		555	0.11		281	0.12		274
Precarious employment (%)	0.01		555	0.01		281	0.01		274
School/student (%)	0.41		555	0.43		281	0.38		274
Homemaker/maternal leave/caretaker (%)	0.03		555	0.02		281	0.03		274
Unemployed (%)	0.05		555	0.05		281	0.04		274
Retired (%)	0.08		555	0.08		281	0.09		274
Apprentice/voluntary work gap year (%)	0.03		555	0.03		281	0.04		274
Marriage status									
Married/in relationship (%)	0.43		587	0.38		294	0.49		293
Married but separated (%)	0.01		587	0.01		294	0.01		293
Single (%)	0.49		587	0.54		294	0.44		293
Divorced (%)	0.06		587	0.06		294	0.05		293
Widowed (%)	0.01		587	0.02		294	0.01		293
Observations	599			300			299		

went against the instructions provided and cannot be used for our analysis below. This is listed as “no choice made” in Table 1. A small number of participants opted to spontaneously donate their gift voucher ($n = 11$), but since this can be seen as “playing with house money”, we have chosen not to count this as a donation of money (results would not change substantively when including this and are provided in Online Appendix 3). Overall, participants donated on average 1.26 Euros ($SD = 2.95$ Euros, minimum donation of those who donated was 11 cents and maximum donation was 20 Euros).

4.2. Self-report variables

In the questionnaire, our participants reported on their past pro-environmental behaviors, their social value orientation as well as on a number of socio-demographic characteristics.

For our self-report measure of stated preference, the aggregate index of pro-environmental behavior self-reports, we have included 20 different questions, where respondents were asked to rate the frequency with which they executed a number of pro-environmental behaviors in the last year. The question read “The following is about your personal behavior. Please indicate whether (yes/no) and if so, how often you have done the named behavior in the last 12 months. With the categories ‘rarely’ to ‘always’, please grade your answer more precisely according to your personal behavior.” We have elicited responses to different areas of pro-environmental behavior, including resource conservation, pro-environmental donations, boycotts, car sharing, use of public transport and so on (for the complete list, see Table 2, next section). With this, the frequency with which respondents perform a certain behavior was measured on a 5-point rating scale (from never (0) to rarely (1), sometimes (2), often (3), and always (4)). The literature often aggregates responses to a number of such questions in order to make inferences about an underlying pro-environmental attitude or preference (on this, see Lange (2024), and our discussion in Section 6). To do so, we have aggregated the 20 items via principal component analysis into an aggregate index of self-reports of pro-environmental behavior items (which we call “Aggregate index of pro-environmental behavior self-reports”). $n = 545$ individuals responded to all 20 questions on pro-environmental behavior and the index is centered around zero with a standard deviation of 2.15.

To measure individuals’ prosocial orientation, the questionnaire also included the “social value orientation” scale (SVO, see Murphy et al., 2011), where six hypothetical money allocation choices allow to characterize individuals as “individualistic” (concern for own outcomes), “competitive” (concern for the difference between own and others’ outcomes), “prosocial” (concern for joint outcomes) and “altruistic” (concern for others’ outcomes even at one’s own detriment). For more information on the computation of the social value orientation measure, the questions and their validity, see Murphy et al. (2011). In our sample, 82 % of respondents fell into the “prosocial” category, 15 % in the “individualistic category” and 2 % each in the remaining categories. It has to be noted that the SVO questionnaire part had the highest non-response and only $n = 532$ individuals completed the social value orientation questions (Gärting et al., 2003, report a similar observation in that many respondents seem to have refused answering their SVO question). In addition, a number of participants remarked to research assistants about their difficulties in understanding the social value orientation task and scale. For our analysis below, we use a simplified version of the SVO scale, where we merge altruistic and prosocial categories on the one hand, and competitive and individualistic categories on the other hand, as altruistic and individualistic are only indicated by 2 % of the respondents ($n = 9$ each; see Online Appendix 4 for similar results when focussing only on the individualistic and prosocial categories).

Finally, we asked participants to provide information on their socio-demographic background, including gender, age, nationality, education, marriage and employment status, as well as net (post-transfer) household income. Using a heterogeneous sample better representative of the general population allows us to see whether our results will vary with regard to different socio-demographic variables (e.g., gender, age, income, and education). We have conducted this experiment in the field to achieve high generalizability in inference. In this regard, previous experiments provide mixed evidence as to whether students behave different than broader samples including also non-students (e.g., see Andersen et al., 2010; Cappelen et al., 2015; Diederich et al., 2016; Falk et al., 2013; Gächter et al., 2004). Our data will allow us to see whether there are significant differences between students and non-students with regard to the different measures of pro-environmental behavior. Comparing our sample socio-demographic information with census and administrative data (Bundeswahlleiter, 2019), we can see that in some respects our data is representative of the overall population of Göttingen (a university town): regarding nationality, our sample percentage of 10 % non-German respondents is close to the official count of 8.1 % in 2017. Of our sample, 4.9 % were unemployed, a percentage also close to the official 5.7 % rate (end of January 2019). On the other hand, we over-sample both the young (67.6 % of our sample is aged 34 years or younger, compared to 37.8 % as per census data for 2017) and the highly educated (80 % in our sample have either Abitur or a university degree, as compared to 43.7 % in the census data referenced).

5. Results

5.1. Correlation between self-reports and observed behavior

To test our hypothesis about the correlation between a multi-item scale of self-reports of pro-environmental behavior and observed pro-environmental behavior, we can look both at the aggregated index of self-reports as well as the 20 individual self-report items in isolation (Table 2, these are computed over the full sample). Incidence of individual self-reported pro-environmental behaviors varies in our sample: the most common pro-environmental behavior is “turning off the tap water while brushing teeth”, with a mean of 3.72 on our rating scale from zero to four, i.e. on average, individuals nearly “always” do this. This self-report item also has the lowest variation with a standard deviation of .71. On the opposite end, the self-reported behavior with the lowest mean across our

Table 2

Aggregate index of self-reports and individual self-reports of pro-environmental behavior. 20 self-reported behaviors sorted by percentage of individuals reporting to have executed the behavior “sometimes” or more often in the previous year (on a 5-point rating scale from never (0) to rarely (1), sometimes (2), often (3), and always (4)). For each individual self-reported behavior, mean, standard deviation and count are provided as well as the correlation with observed product choice (GREEN) and observed donations (DONATION).

Aggregate index of self-reports and individual self-reports of pro-environmental behavior Description	Mean	SD	% Sometimes +	Corr(GREEN)	Corr(DONATION)	Count
Self-report: Aggregate index of PEB items	0.00	2.15		0.09*	0.07	545
(12) Turned off the tap water while brushing teeth	3.72	0.71	93.9 %	-0.03	0.04	592
(11) Turned off lights in rooms that were not used	3.52	0.78	92.4 %	0.05	0.03	592
(3) Brought own shopping bags	3.41	0.95	86.6 %	0.06	0.07 +	596
(14) Walked/cycled shorter distances (up to 4-5km)	3.16	1.02	79.6 %	0.05	0.00	593
(13) Put on warmer clothing against the cold instead of using the radiator	2.90	0.93	74.7 %	0.07 +	0.05	592
(16) Used public transport (e.g. bus/train) instead of the car	2.75	1.16	66.7 %	0.01	-0.04	588
(1) Purchased regional products	2.51	0.77	56.5 %	0.07 +	0.07 +	593
(18) Purchased bio-products (food and drinks)	2.43	1.03	55.9 %	0.02	0.02	594
(20) Had broken items repaired (if possible)	2.54	1.00	54.0 %	0.03	0.00	589
(17) Forewent air travel	2.18	1.43	45.5 %	0.02	0.02	585
(2) Not eaten meat as the principal meal	2.19	1.14	41.0 %	0.03	0.03	595
(15) Car-pooled with others	1.98	1.30	39.9 %	0.06	-0.02	577
(10) Completely switched off appliances over night (no standby)	1.88	1.40	38.3 %	0.06	-0.02	592
(4) Did not buy a product with unnecessary packaging	2.05	1.05	37.7 %	0.03	0.06	594
(9) Boycotted firms/products that were not environmentally friendly	1.64	1.23	28.5 %	0.10*	0.08 +	585
(19) Bought clothing labelled as environmentally friendly	1.69	1.19	27.8 %	0.05	0.04	590
(6) Signed a petition regarding environmental protection	0.98	1.15	12.0 %	0.06	0.00	593
(5) Donated money to an environmental charity	0.87	1.12	10.2 %	0.06	0.15***	596
(8) Actively participated in an initiative for environmental protection	0.71	1.08	9.5 %	0.01	-0.01	592
(7) Participated in rallies for environmental protection	0.75	1.04	7.3 %	0.00	0.03	593

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

sample is “participating in environmental initiatives” (mean 0.71, sd 1.08). “Avoiding unnecessary flights” is the pro-environmental behavior with the highest variation in self-reported frequency (sd 2.18, with a mean for this item of 1.43). The distribution for this individual behavior self-report is rather uniform.

Mean self-reports may be distorted by extreme values so that it is also instructive to look at the proportion of individuals who execute the behavior (extensive vs. intensive margin). We have chosen a cut-off value of two, i.e. we look at the percentage of individuals, who stated they execute the behavior at least “sometimes”. Table 2 orders the 20 pro-environmental behavior items by this proportion, and we can see that there are five behaviors, where roughly three out of four individuals and more have reported executing them at least sometimes. The most common behavior is not letting tap water run while brushing teeth (93.9%), and other common behaviors are “turning off lights in unused rooms” (92.4%), “bringing own shopping bags” (86.6%), “cycling/walking for short distances” (79.6%), and “dressing warmer instead of heating” (74.7%). On the other end of the continuum are behaviors commonly not executed (roughly one in ten or less individuals), such as “signing environmental petitions” (12.0%), “donating to environmental groups” (10.2%), and “participating in environmental initiatives” (9.5%) and “rallies” (7.3%). While these differences in incidence may be due to cost, we cannot know whether other reasons play a role as well, such as social norms or lack of opportunity.

In a correlation table in Online Appendix 5, we also show pairwise correlations of these individual self-report items with each other. Correlations of the individual self-report items are all positive, but vary wildly in effect size, from close to nil to as large as $r = .56$. For comparison, Lange and Dewitte (2021), p. 5, report average correlations amongst their battery of self-report items on pro-environmental behavior in the range of $r = .18$ to $r = .36$. But we note that the largest correlations in our analysis are produced from closely-related items such as participating in environmental initiatives and rallies ($r = .56^{***}$) or buying bio-products and environmentally-friendly clothes ($r = .50^{***}$), whereas more distinct items produce correlations that are smaller than this.

Let us now turn to the relationship between self-reports and observed behavior, the topic of our first hypothesis. To begin with, Table 3 contains pairwise (Pearson) correlations between the two observed behaviors (pro-environmental product choice and donation) and the aggregate self-report index. These correlations are computed over the full sample, i.e. across both choice variations.¹⁵ With regard to our expectation that the aggregate index of self-reports of pro-environmental behavior items (as stated preference measure) is at best moderately correlated with observed pro-environmental behavior, we find indeed positive correlations between our aggregate index of self-reports and the two observed behaviors across the full sample (see Table 3): self-reported pro-environmental behavior (“Self-report: Aggregate index of PEB items”) is positively correlated with acting pro-environmentally overall (“Behavior: Pro-environmental choice (y/n)”, which includes choosing a pro-environmental notepad, refusing either notepad, returning the notepad stating pro-environmental motivation; $r = .09^*$). It is only (statistically) weakly correlated with choosing the pro-environmental notepad alone ($r = .07^+$, not shown in table). While the correlation between observed pro-environmental product

¹⁵ Accounting for the binary nature of some of our variables, Pearson correlations are sometimes avoided in favor of other correlation measures, specifically, tetrachoric correlations should be preferred to the Pearson correlation coefficient when calculated over two binary variables. This leads to similar association patterns, but in general somewhat larger correlation coefficients, as shown in Online Appendix 6.

Table 3

Pearson correlations of self-reports and observed behaviours over the full sample (i.e., no experimental variation, data pooled over both choice variations). “Behavior: Pro-environmental choice (y/n)” includes the observed behavior: either not taking or returning a notepad or taking a recycled-paper notepad. “Donated money (y/n)” is the proportion of people donating any positive amount of money (excluding the donation of their participation lottery vouchers). Self-report measures are “Self-report: Aggregate index of PEB items”, which is the aggregated index of all self-report items on pro-environmental behavior last year, whereas “Self-report: Prosocial (SVO, y/n)” is the proportion of participants who are characterized by the SVO as prosocial (this is not part of the self-report index).

	Self-report: Aggregate Index	Behavior: Pro-env. Choice	Behavior: Donated	Self-report: Prosocial (SVO)
Self-report: Aggregate index of PEB items	1.00			
Behavior: Pro-environmental choice (y/n)	0.09* (0.037)	1.00		
Behavior: Donated money (y/n)	0.07 (0.120)	0.04 (0.284)	1.00	
Self-report: Prosocial (SVO, y/n)	0.18*** (0.000)	0.07 (0.120)	0.09* (0.047)	1.00
Observations	599			

P-values in parentheses (+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$)

Table 4

OLS regression analysis with robust standard errors to account for unequal variance between groups. Dependent variable is the aggregate index of pro-environmental behavior self-report items. Independent variables are the two observed behaviors, i.e. pro-environmental product choice (left) and donation decision (right).

	Aggregate index of PEB self-reports	
	(1) Pro-environmental product choice	(2) Donated money
Observed behavior (y/n)	0.39* (2.09)	0.32 (1.59)
Observations	541	545
F-statistic	4.39	2.54
Adjusted R-squared	0.01	0.00

t statistics in parentheses (+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$)

choice behavior and the aggregate self-report index is statistically significant, the observed behavior is not significantly correlated with many of the individual self-report items (see Table 2), and these individual correlations are typically even smaller. The only significant correlations are with the self-report item on boycotting companies for pro-environmental reasons ($r = .10^*$), as well as two weakly significant correlations with dressing warmer instead of heating and purchasing regional products (each $r = .07^+$).

For observed donations, the evidence is weaker as the aggregate self-report index does not correlate significantly with the actual donation decision (“Behavior: Donated money (y/n)”; $r = .07$, *ns*). Please note that this variable should not be overinterpreted for the full sample, as donation options were different between the two choice variations, and given our hypothesis that the choice of donation options influences how well the behavior variable tracks the self-report, the aggregate over treatments should probably be considered uninformative.¹⁶ The observed donation decision variable is also not correlated with many individual pro-environmental behavior self-report items (see Table 2), with the exception of the corresponding self-report on having donated to an environmental group ($r = .15^{***}$). Only three of the studies analyzed in Kormos and Gifford (2014) have yielded such a low correlation between self-report item and corresponding observed behavior, and this correlation is below the lower threshold of the confidence interval for their meta-analysis (95 % CI [.28 – .60]). The only other significant correlations in our study (but only weakly so at the 10 % level) are between donation decision and bringing own shopping bags ($r = .07^+$), purchasing regional products ($r = .07^+$), and boycotting companies ($r = .08^+$).¹⁷

We can look at the relationship between self-reports and observed behaviour also via OLS regression analysis. These regressions also provide the baseline case for our later analysis of the experimental variation. In Table 4, we thus provide bivariate regressions

¹⁶ The same would be true about the correlation between aggregate self-report index and the amount donated, which exhibits a positive correlation over the full sample (“Behavior: Amount of money donated”, $r = .10^*$, not shown in table).

¹⁷ Also note that observed donations correlate comparatively well with a donation intention self-report item ($r = .31^{***}$), which we have collected to assess whether ensuring anonymity of decision would decrease donations. Given that most respondents said they would not know whether they wanted to donate ($n = 376$), we cannot report that there was a face-to-face overstatement of donation preference. On the contrary, in the end, more people donated money than those who said they would they donate ($n = 167$ vs. $n = 20$). Nevertheless, it is interesting to see that donation intention, while better correlated with observed donation, potentially to being temporally quite close, is not significantly correlated with the individual self-report item of having donated in the past year ($r = .06$, *ns*).

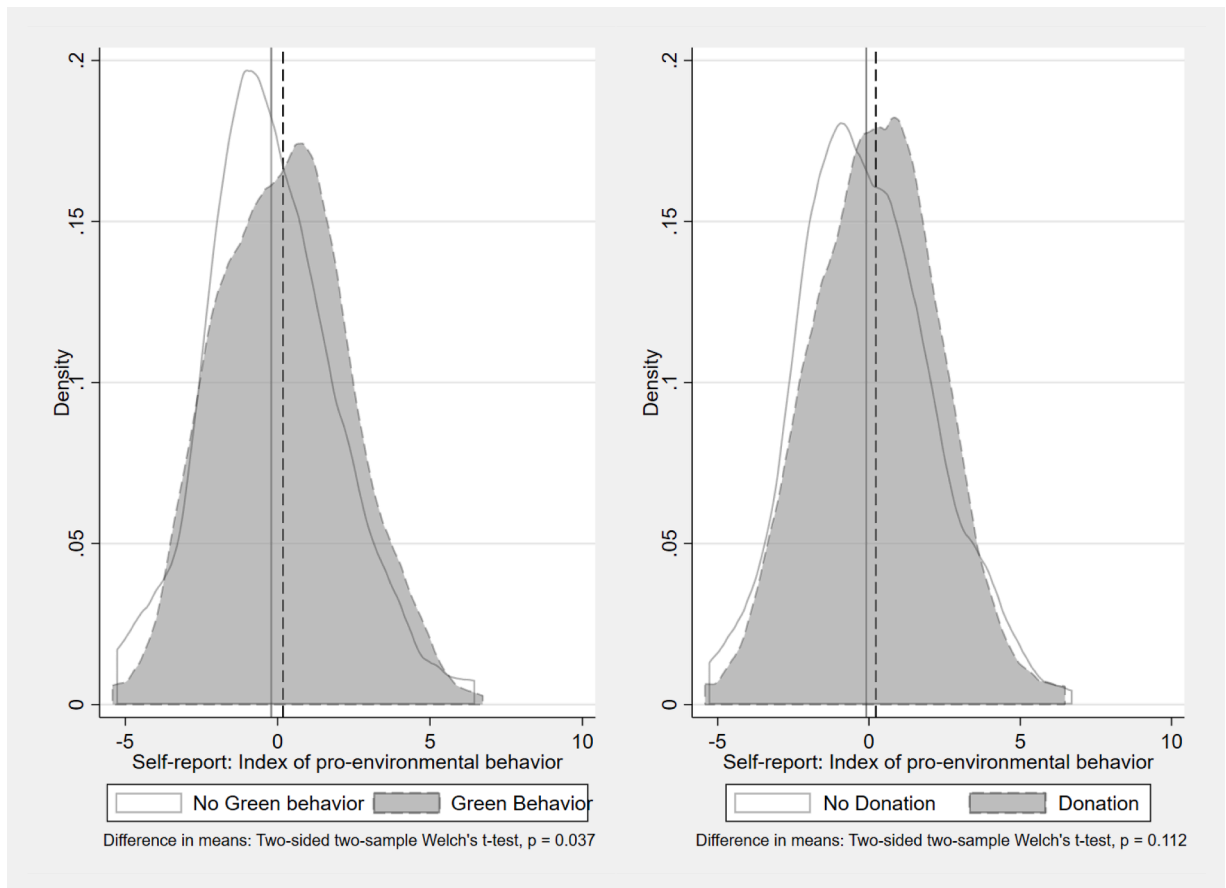


Fig. 1. Kernel density plots of the aggregate index of self-reports across groups showing observed pro-environmental behavior (yes/no, left) and deciding to donate (yes/no, right). Two-sample two-sided Welch's *t*-test rejects hypothesis of equality of means for pro-environmental behavior comparison at the 5% level of significance, but not for donation decision. Vertical lines show group means. Those individuals who acted pro-environmentally scored significantly higher on the self-report index than those who did not act pro-environmentally, whereas the difference is statistically indistinguishable between donors and non-donors.

(with robust standard errors), regressing our self-report index on either pro-environmental product choice (left panel) or donation decision (right panel). Of note here is that the coefficient for pro-environmental product choice is positive and statistically significant ($b = .39^*$), which amounts to 18 % of a standard deviation of the self-report index variable, a small effect size.¹⁸

The coefficients from the bivariate regression models are equivalent to a difference in means *t*-test and hence can be visualized via the kernel density plots of the dependent variable for the two different groups (showing pro-environmental behavior/donation or not, see Fig. 1). Looking at the left panel of the plots of our self-report index variable (see Fig. 1), the aggregate index of self-reports of pro-environmental behavior items is statistically significantly higher for those who choose to act pro-environmentally in terms of product choice (two-sample two-sided Welch's *t*-test). The mean level of the index of self-reports is .1782 for those who acted pro-environmentally and −.2095 for those who did not (i.e. difference of .3877 from the regression above). For the donation variable, we do not find a significant difference in the aggregate self-report index (right panel, Fig. 1, but note again that donation choice varied between participants).

All in all, this provides evidence in favor of **Hypothesis 1**: a correlation between a multi-item scale of pro-environmental behavior self-reports and actual behavior exists, but is small in effect size.

¹⁸ We do not claim causality for this analysis, but are merely interested in a different quantification of the association between self-report and observed behavior. We have chosen the self-report index here as the dependent variable since the survey is temporally after the observed behaviors, and using a continuous variable as dependent variable is easier to interpret than the odds ratios of a logit model, when using a binary dependent variable.

Table 5

Pearson correlations of self-reports and observed behaviors for “charity choice condition” (left) and “no choice condition” (right). Observed donation variable here captures whether the individual donated to an environmental charity in variation 1 (charity choice). In variation 2 there was no choice but to donate to an environmental charity if individuals chose to donate.

	“Charity choice condition”			“No choice condition”		
	Self-report: Aggregate Index	Behavior: Pro-env. Choice	Behavior: Donated Green	Self-report: Aggregate Index	Behavior: Pro-env. Choice	Behavior: Donated Green
Behavior: Pro-environmental choice (y/n)	0.12* (0.049)	1.00		0.06 (0.316)	1.00	
Behavior: Donated to green charity (y/n)	0.15* (0.015)	0.17** (0.003)	1.00	0.01 (0.842)	-0.04 (0.487)	1.00
Self-report: Prosocial (SVO, y/n)	0.17** (0.007)	0.09 (0.138)	0.15* (0.018)	0.18** (0.004)	0.04 (0.476)	0.11 + (0.080)
Observations	300			299		

P-values in parentheses (+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$)

5.2. Experimental variation of donation choice

As discussed in Section 2, donations might not be well suited to reveal a pro-environmental preference, if individuals donate to environmental charities due to prosocial preferences as well. In such a case, we would not expect observed donations to correlate well with our aggregate index of self-reports of pro-environmental behavior items. While the overall correlation between observed donations and self-report index in the full sample ($r = .07, ns$) is difficult to interpret due to choice of charities varying, we find evidence for our hypothesis when comparing correlations between both choice variations (see Table 5).

Recall that, whereas all donations in the “no choice condition” must be to environmental charities, in the “charity choice condition”, participants had a choice between local environmental or local social charities and could thus select into their preferred donation category based on their underlying preferences. In the “no choice condition”, we find a small and statistically insignificant association between self-report index and observed decision to donate to environmental charities ($r = .0121, p = .8421$, Table 5, right correlation panel). By contrast, in the “charity choice condition”, a choice of type of charity leads to a considerably stronger correlation between self-report index and observed donation decision for the environmental charity ($r = .1515, p = .0152$).¹⁹

Our experimental variation has also affected the alignment between both observed behaviors: the correlation between them is stronger and statistically significant ($r = .17^{**}$) when given the choice between charities. When no choice is allowed, similar to the overall sample, this correlation is small and statistically not distinguishable from zero ($r = -.04, ns$). The difference between the two groups’ correlation coefficients itself is statistically significant ($p < .01$).

For this analysis, too, we want to provide a regression analysis formulation of our findings (see Table 6), which allows for a better interpretation of effect sizes and will also allow us to increase the precision of the estimates of our experimental variation by including pre-treatment variables (Gelman et al., 2020, ch. 19). To model the effect of our experimental variation, we here interact the observed behavior variables with a treatment indicator variable to test whether the association between dependent self-report measure and our two different observed pro-environmental behaviors changes through the experimental variation. In Table 6, we first present the interacted model for pro-environmental product choice without covariates and then with covariates (columns 1 and 2), and then repeat this for observed donations (columns 3 and 4). We can see that the interaction terms are statistically indistinguishable from zero in the case of pro-environmental product choice, i.e. we cannot find evidence that the experimental variation increased the correlation between observed behavior and self-report index (and we did not expect it to do so). On the other hand, interaction terms are positive and statistically significant for the donation variable (at the 10 % level without pre-treatment variables and at the 5 % level when including pre-treatment variables). For the full model (column 4), the experimental variation increased the association of observed donations with the self-report index, so that the self-report index was higher for those who donated to an environmental charity by $b = .94^*$, which is 44 % of its standard deviation.

Here, too, we can visualize the difference in mean levels of the aggregate index of self-reported pro-environmental behavior items across groups and choice variations via kernel density plots (see Fig. 2), and provide the associated Welch’s t-tests. We find no significant difference in comparing observed donors vs. non-donors in the “no choice condition” (left panel, $p = .8372$), but the mean aggregate index of self-reports is higher for those who are observed to donate to the environmental charity in the “charity choice condition” (right panel, $p = .0079$, and the difference in means is itself significantly different at the 10 % level, but this is nothing but a restatement of the interaction term in the regression model (3) without pre-treatment variables).

¹⁹ A similar picture emerges when we correlate self-report of pro-environmental behavior and observed donation amount, with a non-significant relationship in the “no choice condition” ($r = .0691, p = .2535$) and a significant relationship between self-report and observed donation amount in the “charity choice condition” ($r = .1280, p = .0355$; the difference in correlation coefficients here, however, is not statistically distinguishable despite the difference in effect sizes).

Table 6

OLS regression analysis with robust standard errors to account for unequal variance between groups. Dependent variable is the aggregate index of pro-environmental behavior self-reports. Independent variables are the two observed behaviors, i.e. pro-environmental product choice (columns 1 and 2) and donation decision (columns 3 and 4). Interaction with treatment indicator represents effect of experimental variation (i.e. the “charity choice condition” is represented by the treatment indicator variable coefficients so that the interaction term captures the effect that a choice of charity has on the observed behavior variable coefficients). Model columns 1 and 3 contain simple interacted models, columns 2 and 4 include pre-treatment variables to improve precision of estimates.

	Aggregate index of PEB self-reports			
	(1) Pro-environmental product choice	(2) + Pre-treatment variables	(3) Donated money	(4) + Pre-treatment variables
Observed behavior (y/n)	0.26 (1.00)	0.33 (1.24)	0.06 (0.21)	-0.07 (-0.24)
Treatment (y/n)	-0.20 (-0.73)	0.14 (0.53)	-0.18 (-0.83)	-0.07 (-0.32)
Interaction term: Treatment X Behavior	0.26 (0.69)	-0.07 (-0.20)	0.82+ (1.93)	0.94* (2.13)
Pre-treatment variables (age, gender, education level, income, nationality)	No	Yes	No	Yes
Observations	541	476	531	467
F-statistic	1.67	4.73	2.54	4.75
Adjusted R-squared	0.00	0.10	0.01	0.10

P-values in parentheses (+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$)

In sum, comparing correlation patterns between both choice variations, we find that without offering a choice of charities, there is no statistically significant correlation between the self-report index and observed donations, whereas both are better aligned when allowing for a choice of charities. Here, self-report index and observed donations are significantly correlated (at a strength comparable to the correlation between the individual self-report donation item and observed donations). In addition, in the “charity choice condition”, both observed behaviors are significantly correlated, again something we did not find for the “no choice condition” (nor the full sample). Given that the interaction term for the regression analysis is statistically significant, we have evidence that the experimental variation had an effect in increasing the correlation between self-report index and observed donations (pro-environmental product choice exhibited a significant correlation with the self-report index in both variations). This we count as evidence in favor of our [Hypothesis 2](#).

Lastly, it is also instructive to look at the correlation between self-reports of *prosociality* and observed donations and product choice (full sample). Here, we find that self-reports of prosociality are correlated positively with the decision to donate ($r = .09^*$) and the amount donated ($r = .08^+$). The mean amount donated is significantly higher for prosocials than for individualists ([Fig. 3](#), right panel). And while self-reports of prosociality are correlated significantly with self-reports of pro-environmental behavior ($r = .18^{***}$, see [Table 3](#), and [Fig. 3](#), left panel),²⁰ self-reports of prosociality are not, however, significantly correlated with our *pro-environmental* behavior variable: [Fig. 4](#), left panel, shows that the percentage of donors is significantly higher in the group of prosocial individuals than in the group of individualistic persons, whereas the difference in proportions of actual pro-environmental behavior (“Behavior: Pro-environmental choice (y/n)”) is not statistically significant between individualistic and prosocial individuals (right panel).

These correlations, too, support our second hypothesis and indicate that observed pro-environmental product choice and donations are differently aligned with self-reports of prosociality and our aggregate index of pro-environmental behavior self-reports.

5.3. Students vs. broader population sample

To analyze the extent to which results from student samples overlap with more general samples, we now compare our overall results with our large student sub-sample.²¹ Comparing students vs. non-students, we find that the aggregate index of self-reported pro-environmental behavior items among students in our sample is significantly higher than for non-students ($d = .4601$, $t(df = 489.353) = 2.4535$, $p - \text{value} = 0.0145$), but the proportion of observed donors is less for students than for non-students ($d = -0.0963$, $z = -2.4764$, $p - \text{value} = 0.0133$). The proportion of students who are observed to act pro-environmentally with regard to product choice is not statistically distinguishable from non-students (overall pro-environmental behavior proportion difference: $d = 0.0142$, $z = 0.3272$, $p - \text{value} = 0.7435$). When it comes to the correlations between these self-reports and observed behaviors, we

²⁰ While we have $n = 532$ answers to the prosociality question, overlap of social value orientation responses with the observed pro-environmental behavior is smaller, viz. $n = 490$.

²¹ The questionnaire answer category lumps together school children and university students. University starting age is 17 or 18 years in Germany and six of 222 respondents in the student category are less than 17 years old. We therefore do not count those six individuals as university students, but results are not sensitive to this choice.

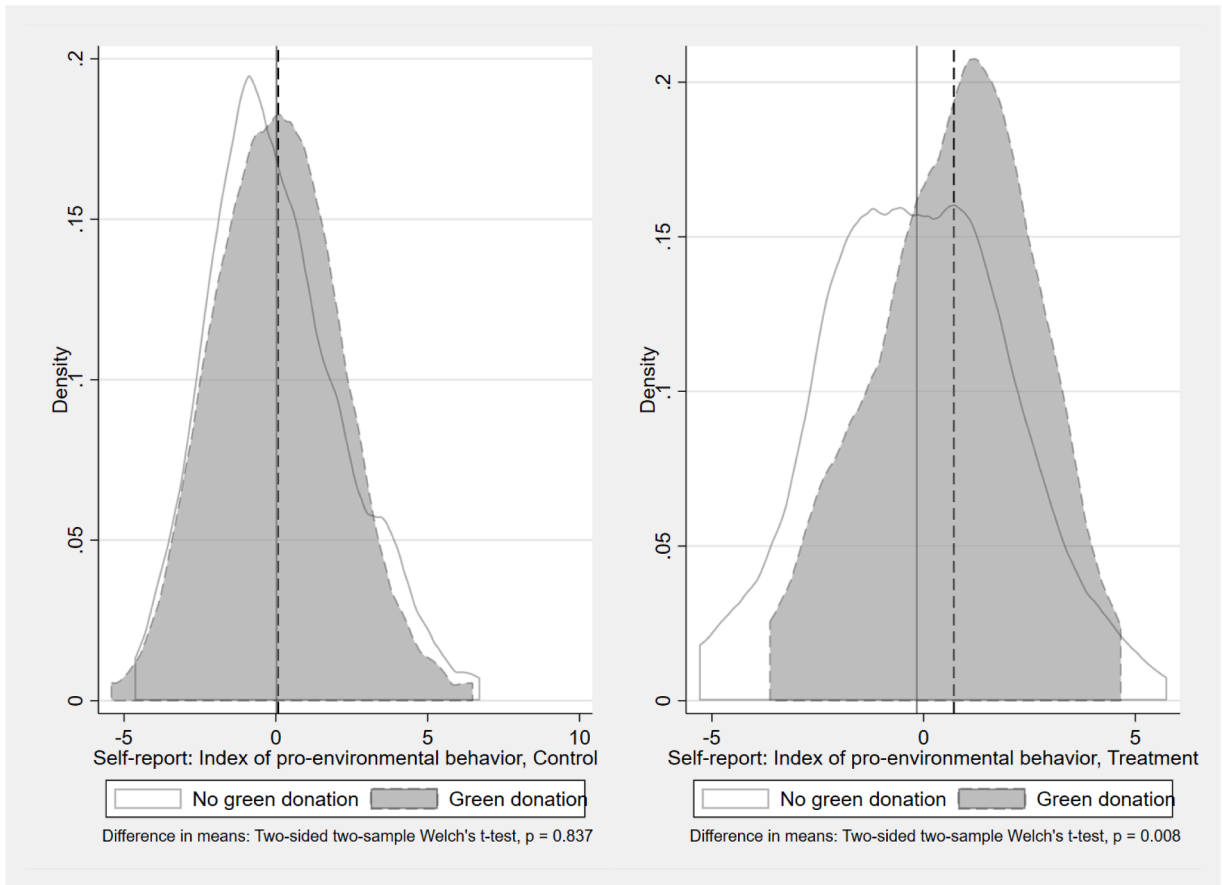


Fig. 2. Kernel density plots: Aggregate index of self-reported pro-environmental behavior items, comparing those who are observed to donate to a pro-environmental charity within “no choice” and “charity choice” conditions. Welch’s t-tests reject equality of means for the “charity choice condition” ($p = .0079$, right panel) but not for the “no choice condition” ($p = .8372$, left panel). Mean index of self-reports is significantly higher for those being observed to donate across the treatments (pro-environmental donation, comparison left with right panel, $p = .0841$).



Fig. 3. Kernel density plots: Aggregate index of self-reports of pro-environmental behavior items (left) and amount donated (right), comparing group of prosocial respondents with group of individualistic respondents (over full sample). Welch’s t-test rejects equality of means for both comparisons. Vertical lines show group means. Prosocial individuals both donate more and exhibit higher levels of the aggregate self-report index.

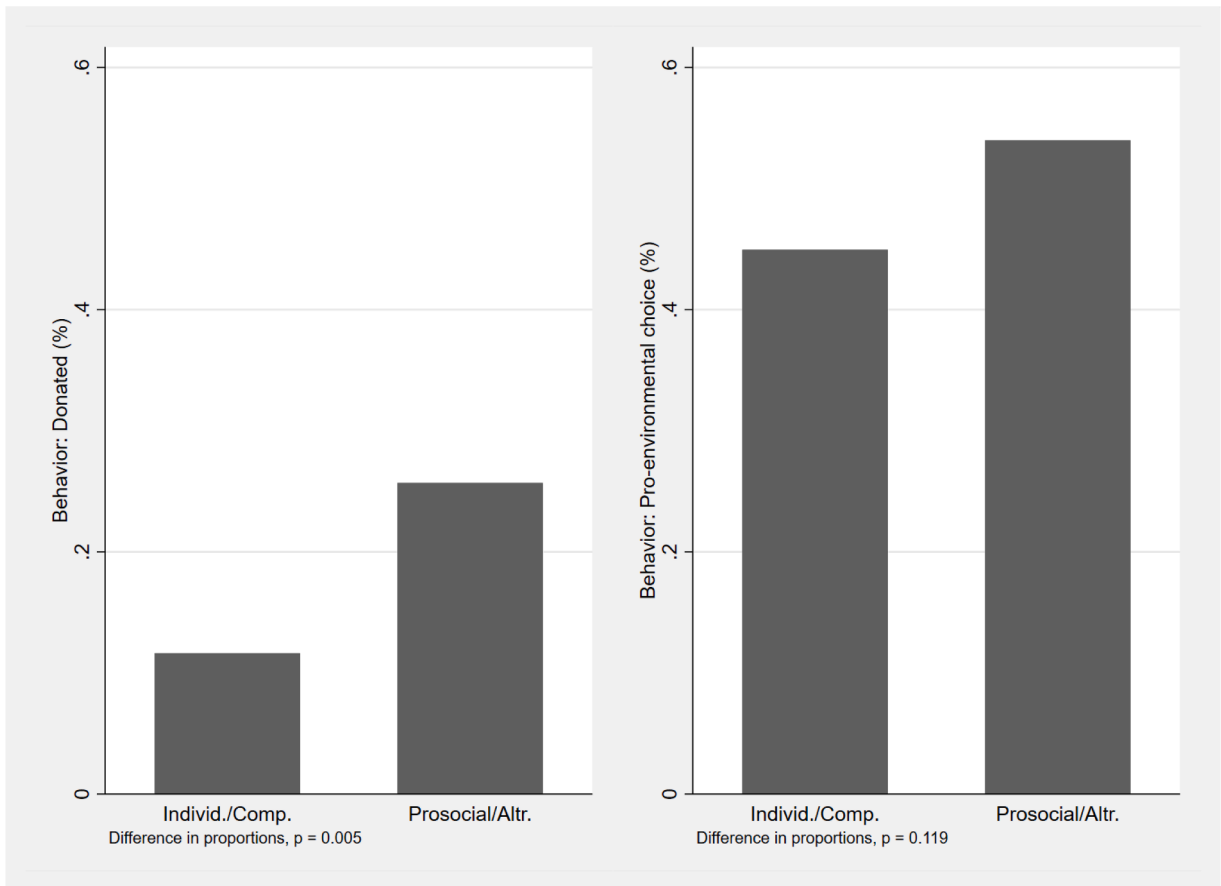


Fig. 4. Comparing prosocial to individualistic respondents (full sample): Percentage of subgroup who donated (left, “Behavior: Donated money (y/n)”) and percentage who acted pro-environmentally (right, “Behavior: Pro-environmental choice (y/n)”). Percentage of donors is significantly higher in prosocial individuals, whereas percentage of respondents acting pro-environmentally is not significantly different between prosocial and individualistic respondents.

find that the correlation between the aggregate index of self-reported pro-environmental behavior items and the observed product choice variable is slightly lower for students than for non-students ($r = 0.0624$, $p - value = .3741$) and the same holds true for observed donations ($r = 0.0706$, $p - value = 0.3147$), but in both cases, we cannot reject the null hypothesis that this is due to chance (at conventional levels of statistical significance). In sum, this means that we have no evidence for large systematic differences in the correlations of our main variables when comparing students to non-students.

5.4. Sensitivity analysis

While our index of self-report items on pro-environmental behavior contains 20 items on diverse pro-environmental behaviors in order to capture an underlying pro-environmental preference, the index construction (like in much of the literature) is ad hoc and one can ask whether our results may be driven by this selection of items. We have thus conducted a sensitivity analysis, creating a data set that contains all 1,048,575 combinations of indices we can compute, and checking whether correlations between the two observed behaviors and these self-report indices differ based on what items are used for index construction (a full report on this analysis is given in Online Appendix 7).

Randomly splitting our initial data set into a training and test data set, we do not find big overlap in the top ten highest performing aggregate indices (in terms of correlation with observed behaviors). There is also not much overlap between the individual self-report items that go into the high correlation indices between training and test data set. For the full sample, individual self-report items 1 (regional products), 9 (firm boycotts), 15 (car-pool), and to a lesser extent 5 (donation), 11 (lights off), and 13 (warmer clothing) all figure in the highest correlation indices. This lack of fit between training and test data is less pronounced for the self-report indices and observed donations. One item that prominently figures here is the item on last year donations (no. 5, no surprise), but aside from this, there is only some overlap for items 12 (no tap water), and 13 (warmer clothing). In the full sample, aside from the individual donation item, items 1 (regional products), 3 (own bags), 4 (unnecessary packaging), 12 (no tap water), 13 (warmer clothing) play a somewhat common role for the highest correlation indices.

A second test compares the distribution of correlations between all conceivable self-report indices and observed donations along our experimental variation. Along this experimental variation, apart from the donation item, the items that produce the highest correlations between self-report index and observed green donations in the “charity choice condition” are different from those in the “no choice condition”. For the former, individual items 2 (no meat), 3 (own bags), 7 (rallies), 9 (firm boycotts), 14 (walk/cycle), and 15 (car-pool) figure in all ten indices, and items 4 (unnecessary packaging), and 5 (donation) in most, whereas the highest correlations in the no choice condition seem mostly be driven by the single-item on donation (no. 5), and with less clear contributions of any other items.

What is more striking here, however, is the minimal overlap in correlation distributions: where the “charity choice condition” leads to an average correlation of the indices with observed (green) donations of $r = .14$, the average correlation is close to zero ($r = .01$) in the “no choice condition”. That the experimental variation increases the correlation and hence decreases the mismatch between observed donations and self-report index is something that happens irrespective of how we construct the self-report index, as it seems to happen over nearly *all* potential self-report indices we could compute with our data.

In sum, this shows that with regard to the correlation between product choice and self-report index in our sample, choosing different items may yield higher (but also lower) correlations, but the comprehensive 20-item index we chose produces a correlation ($r = .09$) close to the average over all indices ($r = .08$). While our random sample split suggests that leaving out items for the construction of the self-report index may increase the correlation with observed product choice (up to $r = .22$ for this random split), it does not do so consistently. And across the full sample, the highest correlation yielded is only $r = .13$.

With regard to observed donations, a clearer case might be made that the single-item on donations may produce a more cost-effective self-report than aggregating multiple items (the maximum index correlation in the full sample is $r = .15$, the average is $r = .06$ vs. individual item correlation $r = .15$, which produces the second highest correlation over all indices). However, in this case, most is gained by modifying the donation paradigm to allow self-sorting, which then increases the correlation by an additional third. This also seems preferable if we recall that the individual self-report donation item does not correlate significantly with the observed product choice variable so that the aggregate self-report index is the better choice if one is interested in an underlying “green” preference.

6. Discussion and conclusion

Our field experiment has elicited three common measures of “greenness” (viz. self-reports, donation to environmental charities and pro-environmental product choice) in a natural setting with a general population sample, and using a fully consequential design with regard to observed donations (i.e. any donation had to be own money, and could not be money earned in the experiment). With our study, we have made two contributions to the literature: Firstly, we have demonstrated that there is considerable mismatch between these measures of “greenness”. We found that an aggregated index of self-reports on 20 different past pro-environmental behaviors, a stated preference measure, if you will, only weakly correlates ($r = .09^*$) with observed pro-environmental product choice. The aggregate index of self-reports of pro-environmental behavior items does not significantly correlate with observed donations in the full sample ($r = .07ns$) and when donations can only go to an environmental charity ($r = .01ns$).

Our second contribution finds that this latter relationship is more complicated, as we have shown that observed donations will better align with our aggregate self-report index conditional on giving individuals the choice to explicitly donate for non-environmental causes as well. When we offered participants a choice of charity to donate to (either pro-environmental or prosocial causes), the correlation between donation to environmental charity and self-report index in this group became larger than in the overall sample and statistically significant ($r = .15^*$). A similar pattern can be observed with regard to the correlation between observed donations and observed product choice. Only in the “charity choice condition” do both observed behaviors align well with each other. With this, our experiment provides some tentative evidence that prosocial preferences and pro-environmental preferences may be conflated when observing individuals donate in a typical experiment that doesn’t allow the choice between prosocial and specifically pro-environmental causes.

In addition, our study did not find evidence for a systematic difference in the correspondence of our aggregate index of self-reports and observed (pro-environmental) behaviors between students and non-students. This question is important for policy-makers trying to use research on pro-environmental behavior and preferences to inform their pro-environmental policies, as many results in the literature on prosociality and pro-environmental behavior draw on student samples and doubts have been raised about whether such results would generalize to a more diverse population.²²

Our study has limitations. For one, effect sizes are small and our correlation patterns between the different observed behaviors and self-reports are not that strongly distinguishable from a point of view of statistical significance; for instance, the key treatment interaction term is only significant at 10 % level absent pre-treatment variables and at the 5 % level once we include those variables. Despite the fact that our core finding chimes with supporting evidence in De Groot and Steg (2008) and Lange and Dewitte (2021), it would be desirable to gather more data points regarding our key hypotheses. To put this in perspective: compared to the literature, our correlations between self-reports and observed behaviors are smaller than the average $r = .46$ reported in the meta-analysis by Kormos and Gifford (2014). When we compare the donation self-report item with the observed behaviour ($r = .15$), our correlation is still lower than the lower end of the confidence interval reported in this meta-analysis (95 % CI [.28 – .60]). Our correlations between observed donations and observed product choice ($r = .04, ns$ in the overall sample and $r = .17^{**}$ in the “charity choice condition”)

²² To be clear: we do find some differences in absolute values regarding some of the individual variables, but we do not have evidence that the low correlation between self-reports and observed behavior found in the full sample would be different if focussing only on the student sub-sample.

are smaller than the correlation in [Lange and Dewitte \(2021\)](#) between an observed donation using house money and their artificial pro-environmental lab task ($r = .28$, *ibid.*, p. 4). Yet, the findings in [Lange and Dewitte \(2021\)](#) support our study's results regarding the general picture that prosocial donations and pro-environmental donations exhibit a distinguishable correlation pattern with other measures of "greenness". Our effect sizes are also roughly comparable to the effect sizes regarding prosociality in [Galizzi and Navarro-Martinez \(2019\)](#).

Secondly, participating in a survey is an inherently prosocial activity, and one may wonder whether this may limit our experiment's external validity. Given our setup at city hall, we would speculate that enough of the less prosocial individuals did participate in the survey, as the study has appeared to be an official enquiry from the administration and as chairs to wait on for an appointment were scarce unless one participated in our survey. Low refusal rates suggest that we did not only include prosocial individuals in the experiment. In the same vein, we have to note that donating to government initiatives is not the norm in Germany, but the framing of our donation task was such as to make clear that individuals donating here would support third sector initiatives (in addition to support to those initiatives given by the municipality).

Thirdly, both observed behaviors in our study are notably different (as per design aim), but a design with real losses is more strongly implemented with regard to observed donations, as it entails a visible cost to the participant in the outlay of own money. In the case of choosing a recycled notepad over a white notepad, the cost to the individual is an opportunity cost, which also varies between individuals, who chose the notepad and those who chose not to avail themselves to one. In addition to this variance in cost, opportunity costs are often much less well understood and accounted for by non-economists, as they do not involve a layout of money. In this way, some of the correlation patterns we have uncovered may be driven by this difference and it might be instructive to also provide a more explicitly costly product choice task (such as for instance having to buy the notepad).

Fourthly, using a design that requires own money in order to make a donation might also have driven our results compared to existing literature, but this should actually be understood to be a strength of our study. In a systematic review of donation tasks in the environmental literature, [Lange \(2023\)](#) notes that only two out of 22 studies use a design, where individuals have to donate their own money (which wasn't earned in the experiment or given as show-up fee) so that stronger correlations between observed behavior and self-reports might be expected when a "hypothetical bias" or "house money" effect is present. For instance, in one study, correlations between an environmental lab task and donations from the money earned in the lab task have been stronger than ours ([Lange and Dewitte, 2021](#)), but similar to our study, a smaller correlation was found between lab task and donations to a non-environmental charity. Our correlation between the self-report index and observed donation might have been larger, if we had given out a fee to donate, and correlations are indeed minimally larger (on the order of $+0.02$), when also counting donations of lottery vouchers in our study for the respective observed behavior. But our experiment has shown that even with own money, a correlation between pro-environmental donations and self-reports can be observed, as soon as the donation paradigm is modified to allow self-sorting of prosocial and pro-environmental individuals. For our main finding, whether or not house money effects exist thus seems of secondary importance. We would thus not see as a weakness that observed donations do not correlate well with self-reports in our setup, if these results were driven by the absence of a "house money" effect, as this would underscore the need to move away from such designs (compare [Moser et al., 2013](#)). After all, pro-environmental behavior is typically not financed with money provided by someone else but one's own money. Experiments better reflective of real-world conditions should then be preferred to inform our understanding of these relationships, even if those show weaker correlations than the usual lab experiments (compare also [Galizzi and Navarro-Martinez, 2019](#); [Klein and Hilbig, 2019](#)).

Given this, what are the take-aways from our study and how should we interpret the mismatch between those commonly used measures of "greenness"? Firstly, our study underscores the need to be more careful in selecting measures depending on whether we wish to learn something about individuals' pro-environmental behavior as opposed to their pro-environmental preferences. Unfortunately, the term "measures of pro-environmental behavior" is often used to refer to both, even though, for instance aggregate self-report scales of pro-environmental behavior items tend to be used to refer to preferences more than to behavior. Secondly, when the aim is to infer from a donation to an environmental charity to an underlying pro-environmental preference, researchers should modify their donation paradigm to offer the possibility for prosocial individuals to self-select into non-environmental charities. We will discuss both take-aways in turn.

As regards the first implication of our study, we note that both self-reports and observed behaviors of pro-environmental behavior are often interpreted to indicate (or reveal) pro-environmental preferences (or a pro-environmental attitude more generally). While common in the literature, this is not necessarily unproblematic. Economists here eschew stated preference measures for observed behaviors, but most often, only one or two behaviors are observed, such as a product choice or a donation to a pro-environmental charity. Even though these behaviors are observed and potentially costly (and hence not hypothetical as opposed to stated preference measures), observing an individual choose -for instance- a recycled notepad allows only to reveal a very specific preference about preferring the recycled notepad to the non-recycled one, but the inference to a more general pro-environmental preference seems a stretch. To infer about a broader preference, a wider variety of behaviors and some aggregation might be desirable or else inferring from a product choice to a pro-environmental preference should only be seen as a weak and *indicative* relationship as opposed to precisely recovering such a preference. In this sense, each individual behavior might not directly reveal such a preference but merely indicate it ([Lange, 2024](#), p. 6). This is plausible also given the variation in motivations why people execute each individual behavior (e.g., taking public transport might be due to financial motivation, or due to a pro-environmental preference, see [Binder and Blankenberg, 2025](#)).

A similar problem poses itself for self-reports, where no individual item can easily be claimed to well capture a broader pro-environmental preference. Here, however, it is less costly to collect multiple individual items, which may then be aggregated by averaging, or by principal component analysis (or factor analysis) to capture some underlying shared variance of multiple items

(on this practice, see [Lange, 2024](#)). While this may have more credibility for an attempt to capture a broader pro-environmental preference than a single item, one needs to bite the bullet of “hypothetical bias”. In addition, the ad hoc nature of deciding which items to include and their varying incidence in the population (see our [Table 2](#) and [Binder and Blankenberg \(2025\)](#)) still seems problematic, despite this being often done (on criticisms, see [Kaiser, 1998](#); [Lange, 2024](#)). Nevertheless, being broader than just one behavior (or behavior self-report), one might argue that such a stated preference measure may better serve as an indicator of an underlying attitude or preference, even if it does not perfectly measure the underlying preference. For economists, this may be disappointing, as observed behaviors are often preferred to the “cheap talk” in surveys. But when it becomes prohibitively expensive to observe multiple behaviors in order to be able to infer from this to a more general pro-environmental preference, aggregated self-report indices may also have their place in research despite them poorly tracking actual behavior.

As donations have come to be a popular measure in experimental economics when it comes to pro-environmental behavior, our second take-away is that we do recommend studies should frame the donation task in a way that individuals may self-select into both prosocial and pro-environmental donations, if researchers wish to use observed donations to infer on an underlying pro-environmental attitude or preference. As our experiment has shown, some prosocial individuals donate to pro-environmental causes when no donation option to prosocial causes is present, and this weakens the correlation between pro-environmental self-reports and observed donations. This poses no problem when the study aim is merely assessing donation behavior properties such as frequency, but it is a problem, when a pro-environmental preference is supposed to be inferred from this measurement. Without the choice of charity to self-select into, we fear a donation measure will not allow us to claim that those individuals donating exhibit pro-environmental preferences. Given that pro-environmental behavior can be executed for different reasons, inferring a pro-environmental preference from behavior is not always unambiguously possible, as we have demonstrated in this article.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at [10.1016/j.jebo.2025.107347](https://doi.org/10.1016/j.jebo.2025.107347).

Data availability

Data uploaded on OSF: <https://osf.io/zcdqvw/>

Funding sources

This research project was partly funded by the City of Göttingen, Department for Sustainable Urban Development.

Declaration of competing interest

No conflicts of interest to report.

Acknowledgement

Two anonymous referees, an anonymous associate editor, and the handling editor Daniela Puzzello offered advice from which this manuscript has benefitted immensely. To them our heartfelt gratitude. Any errors and omissions left are due to the authors. Our sincere thanks goes also to Dinah Epperlein and Janes Grewer at the Rathaus Göttingen for their support before and during the execution of the field experiment. We also wish to thank Laura Wedemeyer, Helene Schüle, Hanna Malin Sasse, Felix Pesch, Anton Ellinger, Gregor Christiansmeyer, Hendrik Bernhardt and Henrik Steindam for research assistance. Replication materials and the Online Appendix can be found here: <https://osf.io/zcdqvw/>.

References

- Alpizar, F., Carlsson, F., Johansson-Stenman, O., 2008. Anonymity, reciprocity, and conformity: evidence from voluntary contributions to a national park in costa rica. *J. Public Econ.* 92 (5), 1047–1060.
- Andersen, S., Harrison, G.W., Lau, M.I., Rutström, E.E., 2010. Preference heterogeneity in experiments: comparing the field and laboratory. *J. Econ. Behav. Organ.* 73 (2), 209–224.
- Arendt, F., Matthes, J., 2016. Nature documentaries, connectedness to nature, and pro-environmental behavior. *Environ. Commun.* 10 (4), 453–472.
- Binder, M., Blankenberg, A.-K., 2017. Green lifestyles and subjective well-being: more about self-image than actual behavior. *J. Econ. Behav. Organ.* 137, 304–323.
- Binder, M., Blankenberg, A.-K., 2025. Motives for Pro-environmental Behavior. <https://doi.org/10.2139/ssrn.5529038>
- Bleys, B., Defloor, B., Ootegem, L.V., Verhofstadt, E., 2018. The environmental impact of individual behavior: self-assessment versus the ecological footprint. *Environ. Behav.* 50 (2), 187–212.
- Bosshard, A., Berger, S., Lange, F., Sosa, A., Kankaanpää, E., Fellegi, E., Dydula, J., Pulicelli, M., Aliyeva, O., Brick, C., 2024. Limited overlap among behavioral tasks, pro-environmental propensity, and carbon footprint. *J. Environ. Psychol.* 97, 102297.
- Brick, C., Sherman, D.K., Kim, H.S., 2017. “Green to be seen” and “brown to keep down”: visibility moderates the effect of identity on pro-environmental behavior. *J. Environ. Psychol.* 51, 226–238.
- Bundeswahlleiter, 2019. Strukturdaten Goettingen. <https://www.bundeswahlleiter.de/europawahlen/2019/strukturdaten/bund-99/land-3/kreis-3159.html>.
- Cappelen, A.W., Nygaard, K., rensen, E. Ø.S., Tungodden, B., 2015. Social preferences in the lab: a comparison of students and a representative population. *Scand. J. Econ.* 117 (4), 1306–1326.
- Carrico, A.R., Raimi, K.T., Truelove, H.B., Eby, B., 2018. Putting your money where your mouth is: an experimental test of pro-environmental spillover from reducing meat consumption to monetary donations. *Environ. Behav.* 50 (7), 723–748.

- Cohen, J., 1988. *Statistical Power Analysis for the Behavioral Sciences*. Routledge 2nd ed.
- Cornelissen, G., Pandelaere, M., Warlop, L., Dewitte, S., 2008. Positive cueing: promoting sustainable consumer behavior by cueing common environmental behaviors as environmental. *Int. J. Res. Market.* 25 (1), 46–55.
- Deaton, A., Cartwright, N., 2018. Understanding and misunderstanding randomized controlled trials. *Soc. Sci. Med.* 210, 2–21.
- Delmotte, V. I., Zhai, P., Poertner, H.-O., Roberts, D., Skeja, J., Shukla, P., Pirani, A., Moufouma-Okia, W., Pean, C., Pidcock, R., Connors, S., Matthews, J., Chen, Y., Zhou, X., Gomis, M., Lonnoy, E., Maycock, T., Tignor, M., Waterfield, T., 2018. Summary for policymakers. In: *Global Warming of 1.5°C*. An IPCC Special Report on the Impacts of Global Warming of 1.5°C above Pre-industrial Levels and Related Global Greenhouse Gas Emission Pathways. In the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty. World Meteorological Organization.
- Diederich, J., Goeschl, T., Waichman, I., 2016. Group size and the (in) efficiency of pure public good provision. *Eur. Econ. Rev.* 85, 272–287.
- Falk, A., Meier, S., Zehnder, C., 2013. Do lab experiments misrepresent social preferences? the case of self-selected student samples. *J. Eur. Econ. Assoc.* 11 (4), 839–852.
- Fraser, A., 2023. Success, failure, and information: how households respond to energy conservation goals. *J. Assoc. Environ. Resour. Econ.* 10 (1), 95–120.
- Gächter, S., Herrmann, B., Thöni, C., 2004. Trust, voluntary cooperation, and socio-economic background: survey and experimental evidence. *J. Econ. Behav. Organ.* 55 (4), 505–531.
- Galizzi, M.M., Navarro-Martinez, D., 2019. On the external validity of social preference games: a systematic lab-field study. *Manage. Sci.* 65 (3), 955–1453.
- Gärting, T., Fujii, S., Gärting, A., Jakobsson, C., 2003. Moderating effects of social value orientation on determinants of proenvironmental behavior intention. *J. Environ. Psychol.* 23 (1), 1–9.
- Gelman, A., Hill, J., Vehtari, A., 2020. *Regression and Other Stories*. Cambridge University Press.
- De Groot, J. I., Steg, L., 2007. Value orientations and environmental beliefs in five countries: validity of an instrument to measure egoistic, altruistic and biospheric value orientations. *J. Cross Cult. Psychol.* 38 (3), 318–332.
- De Groot, J. I. M., Steg, L., 2008. Value orientations to explain beliefs related to environmental significant behavior: how to measure egoistic, altruistic, and biospheric value orientations. *Environ. Behav.* 40 (3), 330–354.
- Ibanez, L., Moureau, N., Roussel, S., 2017. How do incidental emotions impact pro-environmental behavior? evidence from the dictator game. *J. Behav. Exper. Econ.* 66, 150–155.
- Kaiser, F.G., 1998. A general measure of ecological behavior 1. *J. Appl. Soc. Psychol.* 28 (5), 395–422.
- Kaiser, F.G., Kibbe, A., Hentschke, L., 2021. Offsetting behavioral costs with personal attitudes: a slightly more complex view of the attitude-behavior relation. *Pers. Individ. Dif.* 183, 111158.
- Klein, S.A., Hilbig, B.E., 2019. On the lack of real consequences in consumer choice research. *Exper. Psychol.* 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. *J. Environ. Psychol.* 53, 40–49.
- Knutsson, M., Martinsson, P., Wollbrant, C., 2013. Do people avoid opportunities to donate?: a natural field experiment on recycling and charitable giving. *J. Econ. Behav. Organ.* 93, 71–77.
- Koller, K., Pankowska, P.K., Brick, C., 2023. Identifying bias in self-reported pro-environmental behavior. *Curr. Res. Ecol. Social Psychol.* 4, 100087.
- Kollmuss, A., Agyeman, J., 2002. Mind the gap: why do people act environmentally and what are the barriers to pro-environmental behavior? *Environ. Educ. Res.* 8 (3), 239–260.
- Kormos, C., Gifford, R., 2014. The validity of self-report measures of proenvironmental behavior: a meta-analytic review. *J. Environ. Psychol.* 40, 359–371.
- Kotchen, M.J., 2009. Voluntary provision of public goods for bads: a theory of environmental offsets. *Econ. J.* 119 (537), 883–899.
- Lakens, D., 2022. Sample size justification. *Collabra Psychol.* 8 (1).
- Lange, F., 2023. Behavioral paradigms for studying pro-environmental behavior: a systematic review. *Behav. Res. Methods* 55 (2), 600–622.
- Lange, F., 2024. What is measured in pro-environmental behavior research? *J. Environ. Psychol.* 98, 102381.
- Lange, F., Berger, S., Byrka, K., Brügger, A., Henn, L., Sparks, A.C., Nielsen, K.S., Urban, J., 2023. Beyond self-reports: a call for more behavior in environmental psychology. *J. Environ. Psychol.* 86, 101965.
- Lange, F., Dewitte, S., 2019. Measuring pro-environmental behavior: review and recommendations. *J. Environ. Psychol.* 63, 92–100.
- Lange, F., Dewitte, S., 2021. Test-retest reliability and construct validity of the pro-environmental behavior task. *J. Environ. Psychol.* 73, 101550.
- Lenton, T.M., Rockström, J., Gaffney, O., Rahmstorf, S., Richardson, K., Steffen, W., Schellnhuber, H.J., 2019. Climate tipping points-too risky to bet against. *Nature* 575, 592–595.
- Mayer, J., Geller, E.S., 1982. Motivating energy efficient travel: a community-based intervention for encouraging biking. *J. Environ. Syst.* 12 (2), 99–111.
- Moser, R., Raffaelli, R., Notaro, S., 2013. Testing hypothetical bias with a real choice experiment using respondents' own money. *Eur. Rev. Agric. Econ.* 41 (1), 25–46.
- Murphy, R.O., Ackermann, K.A., Handgraaf, M. J.J., 2011. Measuring social value orientation. *Judgm. Decis. Mak.* 6 (8), 771–781.
- Nielsen, K.S., Cologna, V., Bauer, J.M., Berger, S., Brick, C., Dietz, T., Hahnel, U. J.J., Henn, L., Lange, F., Stern, P.C., Wolske, K.S., 2024. Realizing the full potential of behavioural science for climate change mitigation. *Nat. Clim. Chang.* 14 (4), 322–330.
- Nolan, J.M., Schultz, P., 2015. Prosocial behavior and environmental action. In: Schroeder, D., Graziano, W. (Eds.), *The Oxford Handbook of Prosocial Behavior*. Oxford University Press, pp. 626–652.
- Poortinga, W., 2022. Review of Environmental Attitudes and Behaviour Questions in the Understanding Society Survey. Mimeo.
- Requate, T., Camacho-Cuena, E., Siang, C.K., Waichman, I., 2019. Tell the truth or not? the montero mechanism for emissions control at work. *J. Environ. Econ. Manage.* 95, 133–152.
- Steg, L., De Groot, J. I. M., 2012. Environmental values. In: Clayton, S.D. (Ed.), *The Oxford Handbook of Environmental and Conservation Psychology*. Oxford University Press, pp. 81–92.
- Stern, P., 2000. Toward a coherent theory of environmentally significant behaviour. *J. Social Issues* 56 (3), 407–424.
- Turaga, R. M.R., Howarth, R.B., Borsuk, M.E., 2010. Pro-environmental behavior: rational choice meets moral motivation. *Ann. N. Y. Acad. Sci.* 1185 (1), 211–224.
- Venhoeven, L.A., Bolderdijk, J.W., Steg, L., 2013. Explaining the paradox: how pro-environmental behaviour can both thwart and foster well-being. *Sustainability* 5 (4), 1372–1386.
- Wagner, K., 2016. Environmental preferences and consumer behavior. *Econ. Lett.* 149, 1–4.
- Wolff, H., 2014. Keep your clunker in the suburb: low-emission zones and adoption of green vehicles. *Econ. J.* 124 (578), 481–F512.