Identifying bias in self-reported pro-environmental behavior

Koller, K.; Pankowska, P.K.; Brick, C.

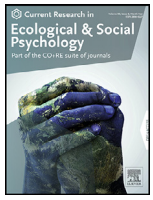
DOI
10.1016/j.cresp.2022.100087
Publication date
2023
Document Version
Final published version
Published in
Current research in ecological and social psychology

Current Research in Ecological and Social Psychology 4 (2023) 100087

Contents lists available at ScienceDirect

Current Research in Ecological and Social Psychology

journal homepage: www.elsevier.com/locate/cresp



Identifying bias in self-reported pro-environmental behavior☆,☆☆

Katharina Koller[a], Paulina K. Pankowska[b], Cameron Brick[c,*]

[a] Department of Work and Equal Opportunities, Centre for Social Innovation, Vienna, Austria
[b] Department of Sociology, Utrecht University, Utrecht, the Netherlands
[c] Department of Psychology, University of Amsterdam, Amsterdam, the Netherlands

ARTICLE INFO

Keywords:
Pro-environmental behavior
Measurement error
Question-behavior effect
Social identity
Social norms
Social desirability bias

ABSTRACT

Research on pro-environmental behavior (PEB) informs social policies and interventions, so the quality of PEB measurement is critical. Self-reported PEB measures in surveys often contain non-negligible measurement error that can bias estimates and lead to incorrect findings. Given the potential presence of error, we hypothesize that changes to the way self-reported PEB is measured might lead to systematic measurement errors that affect the validity of results. Study 1 (N = 951) showed that priming participants with related scales like environmentalist identity did not substantively change reported behavior (all $ds \leq 0.12$). To investigate the possibility of overreporting without priming, Study 2 (N = 385) measured littering prevention behavior using the Unmatched Count Technique. A standard questionnaire format led to much higher reported behavior compared to the more anonymous covert condition, $d = 0.53$, and this effect appeared driven by participants who reported a stronger environmentalist identity. These results may help to explain some of the observed error in PEB measures. We suggest that researchers could reduce measurement bias with indirect questioning techniques.

1. Introduction

Research on pro-environmental behavior (PEB) has strong potential to reduce environmental impacts by informing policy decisions and interventions, which can in turn promote more sustainable lifestyles and climate change mitigation (Shove, 2010). The success of these policies and interventions depends on the rigor of the underlying research and in particular the validity of the behavioral measures. Most existing research relies on self-report measures of pro-environmental behavior. However, the validity of these instruments is questionable. One sign of trouble is that self-reported behavior often shows weak relationships with objectively measured behavior, as investigated across domains such as weight, social media, and mobile use (Gorber et al., 2007; Kobayashi and Boase, 2012; Prince et al., 2008; Shields et al., 2008; Spielholz et al., 2001; Streff and Wagenaar, 1989; Verbeij et al., 2021).

In environmental psychology, a meta-analysis found that self-reported PEB only explained 21% of the variance in objective behavior (Kormos and Gifford, 2014). In that study, self-reports of behaviors such as energy usage or recycling were compared with objective measures (observations, peer ratings, or meter readings) of the same type of behavior in the same people. While the meta-analysis only included 15 studies, more recent experimental studies also found weak relationships between self-reported and objective PEB (Lange et al., 2018; Lange and Dewitte, 2021a, 2021b). Moreover, self-report instruments often show little predictive value: although many people report pro-environmental values, attitudes, and intentions in surveys, these beliefs are not well-reflected in their recycling, travelling, or consumer behavior (Chaplin and Wyton, 2014; Kroesen and Chorus, 2018; Redondo and Puelles, 2017; Sheeran and Webb, 2016). This suggests issues with the validity of self-reported measures and we evaluate them further in the current studies.

One of the key shortcomings of self-reported measures is their measurement error, which is present in all survey data. This substantially lowers validity and is likely one of the reasons why self-reports fail to accurately reflect objective behavior (Alwin, 2007; Pankowska et al., 2021). For continuous indicators, we use measurement error to indicate the difference between the true score of the indicator (or scale) and its observed score (Crocker and Algina, 1986). The presence of such error in surveys is often a result of inadequate questionnaire design, incorrect data collection procedures, as well as interviewer and respondent effects (Biemer, 2010; Pankowska et al., 2020; Saris and Gallhofer, 2007). The errors can be random, thus occurring by chance and without any specific patterns, or errors can be systematic and based on a specific pattern, e.g., when errors depend on interviewee/interviewer characteristics or variables are consistently under- or overestimated. Random errors

☆ **Authorship declaration:** Please see the contributories roles below. All authors approved the final manuscript for submission.
☆☆ *Note.* See https://www.elsevier.com/authors/policies-and-guidelines/credit-author-statement for the details and definitions of each role.
* Corresponding author.

https://doi.org/10.1016/j.cresp.2022.100087
Received 29 July 2022; Received in revised form 9 December 2022; Accepted 27 December 2022

increase variability and reduce measure reliability, but systematic errors shift the central tendency measure, leading to bias and lower validity (Bhattacherjee, 2012).

Because the validity threat is critical, we test for systematic error using different survey techniques measuring self-reported pro-environmental behavior. Systematic errors in self-report are often a result of deliberate or non-deliberate misreporting. Common sources of measurement error discussed in previous research are social desirability, memory/recall bias, participants' subjective interpretation of Likert response categories, or inaccurate self-perceptions of respondents' behavior and behavior frequency (Eisenhower et al., 1991; Kormos and Gifford, 2014; Lange and Dewitte, 2019). In our studies, we propose that the characteristics or design of self-report surveys can influence misreporting.

1.1. Order effects bias in PEB research

Many PEB studies rely on survey research that includes measures of environmental constructs. These constructs are often related to intentions, identity, or social norms and are used to predict self-reported behavior. When constructs are presented in a specific order, responding to the first construct might influence the responses to the next one. Such order effects can result in misreporting (Dillman et al., 2014; Perreault, 1975). Order effects can be caused by fatigue, boredom, or the way that information is processed (Egleston et al., 2011; Serenko and Bontis, 2013). Any sequence of measured constructs could result in order effects. However, order effects can also occur due to different cognitive processes sparked by the specific constructs and lead to systematic misreporting. Next, we outline evidence for specific order effects on self-reported measures which are related to measuring or manipulating intentions, social identities, or social norms. We discuss how these order effects may result in overreporting pro-environmental behavior.

1.1.1. Bias through measuring intentions

Reporting cognitions such as intentions that are related to a specific behavior increases later reports of having performed that behavior (Morwitz and Fitzsimons, 2004). This question-behavior effect occurs in health psychology as well as prosocial and consumer behavior (Spangenberg et al., 2003; Wilding et al., 2016; Wood et al., 2016). Usually, self-reports of performed behaviors will change in a normative direction, i.e., socially desirable behaviors will be overreported and socially undesirable behaviors underreported. For instance, if overweight adults responded to a questionnaire about intentions and attitudes regarding physical activity, the same adults reported more physical activity three months later than a control group who previously were asked about fruit and vegetable consumption (Godin et al., 2011). Likewise, young adults who answered questions about their alcohol consumption reported less problematic drinking behavior three months later compared to a control group who did not previously answer questions about alcohol consumption (McCambridge and Day, 2008).

1.1.2. Social identity bias

Responding to a social identity scale can increase social identity salience through priming, and increased identity salience can affect the subsequently measured behaviors. In fact, research on priming has intentionally used identity questionnaires to increase identity salience, purposefully affecting constructs measured afterwards (Diamond, 2020; Harrison et al., 2009; Howard and Borgella, 2018; Lalonde and Silverman, 1994; McGlone and Aronson, 2006; Shih et al., 1999; Transue, 2007). Answering identity questionnaires can affect later responses by putting concerns and interests relevant to the identity in the forefront (Klar, 2013). For instance, Diamond (2020) primed political identities by asking participants about their party affiliation and opinions or values regarding the party's policies. A priming effect can only occur if the identity being primed already "exists" within the participant and is relevant to them.

On the other hand, the fact that a specific identity is measured in a survey provides the participants with some information regarding the identity that is expected of them and that is relevant to the research (Brenner, 2014, 2017; Brenner and DeLamater, 2016). If this identity is also endorsed by societal norms and thus associated with social desirability, some participants might still want to express this identity even if it is not accurate. This is particularly the case for surveys because they are an easy and low-cost opportunity for expressing identities. As a result, surveys can easily lead to overreporting behaviors or attitudes consistent with a normative identity. This is consistent with identity playing an important role in socially desirable responding (Brenner, 2014, 2017; Brenner and DeLamater, 2016).

1.1.3. Social norms bias

Order effects can also arise because of how instructions are phrased and which information they reveal. Instructing participants to respond what they think other people would say, or phrasing a question about what behavior participants think they "should" do (compared to what they "would" do) can change reports of behaviors and attitudes (Findor et al., 2020; Kotzur et al., 2020; Ployhart and Ehrhart, 2003). This suggests that emphasizing other people's opinions and values in the instruction changes later reports. In addition, participants form beliefs about the study's purpose based on information presented in the survey, a phenomenon called demand characteristics (Corneille and Lush, 2022; McCambridge et al., 2012). Based on what participants think the study purpose is, they might change their later responses to try to please the researcher or conform with anticipated hypotheses (demand effects). Thus, even study's instructions can result in misreporting of later behavior, because presenting an instruction can increase the salience of social norms by emphasizing other people's and the researcher's opinion of the topic.

1.1.4. Evidence for order effects in self-reported PEB

The above findings show that measuring or manipulating specific constructs (i.e., intentions, identities, or norms) in a survey may cause order effects that bias later measures in a socially desirable direction. No study has systematically examined order effects in self-reported pro-environmental behavior (PEB). Yet, misreporting is a potential explanation for the gaps and inconsistencies between predictors of self-reported and objective PEB. There is further evidence suggesting that misreporting in PEB might be specifically due to order effects.

First, future intentions of performing pro-environmental behavior are often included in PEB surveys and used to predict PEB. While future intentions are often strongly associated with self-reported PEB, the associations of intentions with objective PEB such as meter readings or observations are weak or absent (Bamberg and Möser, 2007; Chao, 2012; Chao and Lam, 2011; Gkargkavouzi et al., 2019; Morren and Grinstein, 2016; Stutzman and Green, 1982). These findings suggest that self-reported PEB is misreported in questionnaire studies that also measure intentions to perform PEB.

Second, social identity is another common predictor of PEB that is often measured in survey studies. The dual conceptualizations of environmental*ist* identity and environmental self-identity positively predict self-reported PEB (Brick et al., 2017; Brick and Lai, 2018; Dono et al., 2010, p. 20; Lalot et al., 2019; Moser and Kleinhückelkotten, 2018; Schmitt et al., 2019; Van der Werff et al., 2013; Whitmarsh and O'Neill, 2010) but they do not always predict objective measures of PEB (Brick and Sherman, 2021; Moser and Kleinhückelkotten, 2018; van der Werff and Steg, 2016). These mixed findings show that more work is needed to verify when identities predict PEB and under what conditions. Identity is one of the most important and consistent predictors of self-reported PEB and some research also suggests indirect effects of identity on objective PEB (Moser and Kleinhückelkotten, 2018; Nielsen et al., 2022). Given the relevance of identity in PEB research and responding, we included how much someone sees themselves as an environmentalist to predict PEB in both studies.

Finally, social norms interventions are considered a promising avenue for increasing pro-environmental behavior. Several studies examined norms in the lab and in the field but found mixed effects on self-reported and objective PEB (Cialdini et al., 1991; Farrow et al., 2017; Gifford, 2014). In field experiments, social norm interventions are often disguised or embedded within the environment, for instance by using posters and flyers (e.g., Schultz et al., 2008). We suggest that study instructions may emphasize social norms similar to the interventions used in field studies, and therefore lead to demand effects. If social norms in the survey situation are salient due to the instructions, this could result in misreporting of subsequently measured PEB.

We hypothesize that measuring self-reported intentions to perform pro-environmental behavior, measuring environmentalist identity, and presenting survey instructions highlighting norms of other people or the researcher before a PEB measure will cause overreporting of pro-environmental behavior.

H1–3. Relative to a control task, individuals will self-report more past PEB after completing a future behavioral intention measure (*H1*), an environmentalist identity measure (*H2*), or receiving an instruction emphasizing social norms (*H3*).

1.2. Social desirability

The above literature suggests a potential mechanism for overreporting could be social desirability. We hypothesized that presenting intentions, identity, or norms to participants reminds them that pro-environmental behavior is valued and desired by society, and therefore participants may overreport PEB. However, we also assume that the importance of acting pro-environmentally is already embedded in public discourse (European Parliament et al., 2021; Milfont et al., 2021; Tyson et al., 2022). If that were true, the self-reported PEB measures might be subject to misreporting due to socially desirable responding even without order or priming effects (i.e., overreporting even in the control condition).

Yet, previous research found that social desirability did not meaningfully affect self-reported PEB (Kormos and Gifford, 2014; Milfont, 2009; Pfattheicher et al., 2016; Vesely and Klöckner, 2020; Vilar et al., 2020). Previous studies mostly conceptualized social desirability as a personality trait of conformity and seeking social approval and measured it using a questionnaire. However, standard social desirability questionnaires tend to show poor reliability and construct validity (Barger, 2002; Beretvas et al., 2002). An alternative approach to reduce socially desirable responding would be indirect questioning techniques that allow participants to mask their exact responses, such as Nominative Questions, Randomized Response, and the Unmatched Count Technique (UCT). In nominative questions, participants are asked to report a peer's behavior instead of their own. However, this technique often fails to elicit honest responses (St. John et al., 2010). Randomized response allows participants to only answer truthfully if a randomizer device yields a specific number. However, this technique is considered unreliable as it often yields low validity, low compliance, and is considered difficult and/or untrustworthy (Coutts and Jann, 2011; Holbrook and Krosnick, 2010; John et al., 2018; Lensvelt-Mulders et al., 2005).

In the Unmatched Count Technique (UCT), participants are presented with several statements. To increase anonymity, participants only indicate how many total statements apply to them and do not respond to the statements separately (Dalton et al., 1994). Overall, participants experience the UCT as more trustworthy, easier to understand, and more anonymous than randomized response (Coutts and Jann, 2011). These results indicate that the UCT provides an overall valid estimate of socially desirable behavior and that comparing UCT responses with those obtained using a standard PEB measure can help to identify social desirability effects.

We used the UCT to investigate misreporting in self-reported littering prevention behavior. If participants reporting more anonymously report less PEB, this would be consistent with the possibility of social desirability in the standard questionnaire. We chose littering prevention behavior as it is commonly seen as problematic, unacceptable, and consequential for the environment across different worldviews and ideologies (Doherty et al., 2022; Freije et al., 2019; Ibrahim et al., 2021; Kallgren et al., 2000). For Study 2, we formulated the following hypothesis:

H4. Individuals who respond to a standard PEB measure will report more littering prevention behavior than individuals who respond covertly through the UCT.

1.3. Overview of studies

The validity of self-report measures is surprisingly low (e.g., Kormos and Gifford, 2014). These inconsistencies might be caused by systematic measurement error in the self-reported instruments. We report two studies testing for the effects of systematic error on PEB self-reports. Study 1 investigated the effects of order bias on overreporting PEB by experimentally manipulating the question-behavior effect, social identity salience, and social norms and compared these conditions with participants who only completed a control task. Because pro-environmental behavior might already be embedded in social values and does not need to be primed through order effects, Study 2 further investigated misreporting by changing the questionnaire design. Using the Unmatched Count Technique (UCT), we obtained both covert and standard measures of littering prevention behavior to reveal overreporting. In a pilot study before Study 2, we investigated the psychometric properties of the Littering Prevention Behavior Scale (Ojedokun, 2016), which we adapted for use with the UCT. All studies were pre-registered, approved by the university's Ethics Review Board, and used informed consent. The pre-registrations for Study 1 and 2 are at https://osf.io/z975b. All materials, analysis scripts, and anonymized datasets are at https://osf.io/pfw4d. The Supplemental Material available at https://osf.io/2fas3 provides the links to each of the materials.

2. Study 1

Study 1 investigated whether changing questionnaire characteristics of an online survey could change reports of pro-environmental behavior. We tested three potential sources of bias: the question-behavior effect from measuring behavioral intentions; social identity salience from measuring identity; and normative influence from questionnaire instructions. The experimental manipulations could be described as order effect conditions, since they were based on common features of PEB studies that were not intended as manipulations in previous work. Any change in PEB compared to the control condition could help explain inconsistencies in previous research and might provide insights into the misreporting of PEB.

2.1. Study 1 Methods

2.1.1. Sample characteristics

Sample size planning was based on a power analysis using G*Power (Erdfelder et al., 2009; see Supplement B for details). We recruited a sample based on a one-tailed *t*-test of manipulations leading to increases in reported PEB and assumed similar group sizes with $\alpha = 0.05$, power = 0.80, and effect size $d = 0.24$. This effect size was based on a meta-analytic estimate of the question-behavior effect, as reliable estimates regarding the other manipulations were not available (see Supplement B for further discussion of effect sizes). The decision was also limited by available funds. We oversampled by 10% to account for attrition. Thus, Study 1 aimed for 946 online participants.

979 participants were recruited on Prolific in April 2021 in exchange for financial compensation. The only criterion was residency in the United States. Four participants responded to all items except for the sociodemographic questions and were retained in the data set. 27 participants terminated the survey prematurely and were excluded. One

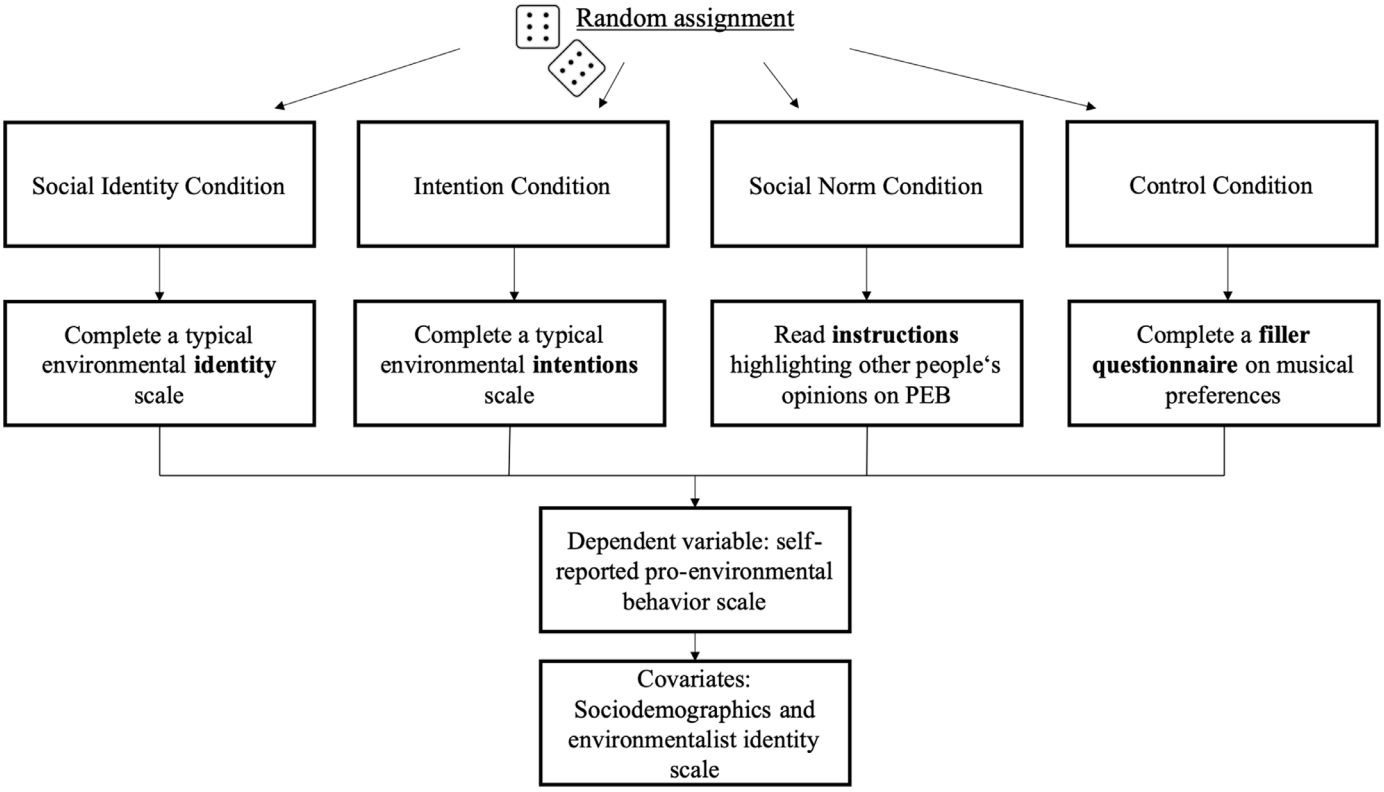


Fig. 1. Study 1 Experimental design.

Table 1
Sample descriptive statistics for all studies.

	Study 1	Pilot Study	Study 2
N	951	91	385
Male	49.5%	45%	50.1%
Female	48.9%	54%	47.3%
Non-binary/Third gender	1.2%	1%	1.3%
Prefer not to say	0.4%	0	1.3%
Age in years, *M* (*SD*)	35.6 (13.0)	34.63 (11.8)	38.3 (12.6)
Age range	18 - 79	18 - 68	18 - 77
Less than a high school degree	1.1%	1.1%	0.5%
High school degree	36.9%	39.6%	28.3%
Bachelor's degree	47.2%	44%	45.5%
Master's degree	13.3%	15.4%	23.6%
Doctorate	2.6%	0	2.1%
Employed full-time	51.4%	48.4%	59%
Employed part-time	14.8%	24.2%	11.9%
Unemployed	14.6%	13.2%	13%
Student	12%	12.1%	6.2%
Retired	4.9%	1.1%	6.2%
Unable to work	2.3%	1.1%	3.6%
Supports Democrats (1 – 3)	57.7%	50%	51.4%
Politically Independent (4)	22.1%	30%	21.6%
Supports Republicans (5 – 7)	20.2%	20%	27%
Social status (1–10), *M* (*SD*)	5.26 (1.7)	5.12 (1.5)	5.48 (1.7)

participant failed both attention check items and was excluded, as preregistered. Thus, the final sample was $N = 951$ (see Table 1). Study 1's sample was fairly balanced with regards to gender and age, but university-educated, full-time employed, and Democratic-leaning participants were overrepresented.

2.1.2. Procedure

Participants in Study 1 were randomly assigned to one of three treatments or the control condition (see Fig. 1). Participants in the Social Identity Condition completed an environmentalist identity measure to increase their identity salience. Participants in the Intention Condition indicated their future intentions to perform pro-environmental behavior (PEB) to investigate the question-behavior effect. Participants in the Social Norm Condition read instructions that emphasized the injunctive norms (what people should do) around PEB. Participants in the control condition responded to a filler questionnaire. Afterwards, all participants self-reported their PEB. Then, participants in the Norm, Intention, and Control condition responded to the environmentalist identity measure (which was also used in the Identity condition). Finally, all participants responded to sociodemographic items and were debriefed.

2.1.3. Randomization

After random assignment, 239 participants were in the Identity condition, 236 in the Intention condition, 234 in the Norm condition, and 242 in the Control condition. Exploratory chi-square tests showed no association between condition and gender, $\chi^2(9) = 6.95$, $p = .642$, nor employment status, $\chi^2(15) = 14.23$, $p = .508$. Exploratory Kruskal-Wallis tests showed no significant differences between condition and employment ($\chi^2 = 0.597$, $p = .899$, $df = 3$), condition and political identity ($\chi^2 = 2.61$, $p = .456$, $df = 3$), nor condition and social status ($\chi^2 = 0.78$, $p = .855$, $df = 3$). This suggests that the random assignment was successful. Descriptive statistics are in Supplement E.

2.1.4. Experimental manipulations

The full text of the experimental manipulations and items is available on OSF.

Social identity condition. Participants first responded to four items representing an established environmentalist identity scale (Brick and Lai, 2018 e.g., "I see myself as an environmentalist.") and to three items adapted from previous research on identity priming that used questionnaires to increase identity salience (Brenner and DeLamater, 2016; Diamond, 2020; Transue, 2007). The additional three items were: "Different decisions have different impacts on our environment. When you

make decisions, how important is it to you to protect the environment?", "How important is it to you, personally, to make the world a greener world?", both from 1 (*not at all important*) to 5 (*extremely important*), and: "How close do you feel to other people that want to protect the environment?", from 1 (*not at all close*) to 5 (*extremely close*).

Intention condition. These participants reported their future PEB intentions referring to the following six months. The 21 items were introduced by a phrase commonly used to measure intentions in research on the question-behavior effect: "Ask yourself: will you do … ?" (see Godin et al., 2012; Spangenberg et al., 2003). Participants then rated each item on a 7-point Likert scale from "*extremely unlikely*" to "*extremely likely*". The items used in the manipulation were rephrased from the Recurring pro-Environmental Behavior Scale (REBS; Brick et al., 2017). For instance, the item "How often do you carry a reusable water bottle?" was changed to "Now, ask yourself … Will I carry a reusable water bottle?". This set-up was based on many previous studies on PEB, which first measured behavioral items phrased to capture future intentions, and then asked participants how often they had performed the same behaviors in the past (Barr, 2007, p. 200; Carfora et al., 2017; Chao, 2012; Harland et al., 1999; Howell et al., 2015; Niaura, 2013; Whitmarsh and O'Neill, 2010).

Social norm condition. This condition aimed to manipulate bias through a survey instruction stating the goal of the study. The phrasing was based on previous manipulations or interventions of social norms highlighting injunctive norms of PEB (Schultz et al., 2008). We used injunctive norms because they allow for wording similar to how environmental groups or labs might present survey instructions. Participants learned that a fictitious research institute investigated the popularity of different ecological behaviors, as many people are concerned about the environment. This narrative was supported by a logo representing the institute and a picture of a tree.

Control condition. These participants responded to a filler measure without any connection to the environment or environmental social identity: the Short Test of Musical Preferences – Revised (Rentfrow and Gosling, 2003). Respondents rated 23 music genres on a scale from 1 (*dislike strongly*) to 7 (*like strongly*).

2.1.5. Measures

Pro-environmental behavior. The main outcome was the Recurring Pro-environmental Behavior Scale (REBS; Brick et al., 2017), which has 21 behaviors that are rated from 1 (*never*) to 5 (*always*). Scale analysis indicated good model fit, but we excluded the item about air travel because it was the only item with an unexpected negative factor loading after reversed items were recoded (see factor loadings in Supplement E). In retrospect, frequency of air travel might not have been a good measure of environmentalism during the COVID-19 pandemic. The model was estimated again without this item, and again indicated good fit and high reliability. A composite was created from the mean of the 20 items after reversals with Cronbach's $\alpha = 0.82$.

Manipulation reflections. Participants indicated their thoughts and deliberations when answering the PEB measure in three ad-hoc manipulation reflections items (one for each experimental condition) on a 5-point Likert scale from 1 (*not at all*) to 5 (*a great deal*): When you were thinking about your past behavior, to what extent did you feel like an environmentalist?; When you were thinking about your past behavior, to what extent were you thinking about what you would like it to be in the future?; When you were thinking about your past behavior, to what extent were you thinking about what your family, friends, colleagues, or other people important to you, think about these behaviors? All participants responded to these manipulation reflections directly after the PEB measure.

Environmentalist identity. After responding to the PEB measure, participants in the Intention, Social Norm, and Control conditions responded to the four-item environmentalist identity scale by Brick and Lai (2018). This scale measures how much a person sees themselves and wants to be seen as an environmentalist, with items rated from 1 (*strongly disagree*) to 5 (*strongly agree*). Participants in the Identity condition already responded to this measure as part of their experimental manipulation in the beginning of the survey. Because those authors reported no indication of validity beyond Cronbach's α, we investigated the factorial validity of the original 2018 data before using the scale in the present study (see Supplement C). CFA indicated moderate fit, and this was considered acceptable. Item scores were averaged into a mean composite, Cronbach's $\alpha = 0.93$. Additionally, we tested whether the factorial structure of the environmentalist identity items was equivalent across conditions (Social identity condition vs. Social Norm and Intention condition) by comparing two CFA models in a Likelihood-ratio test. The Likelihood-ratio test indicated no differences between the CFA models.

Sociodemographic variables. Study 1 also contained sociodemographic questions about age, gender, education, employment status, political party identification, and subjective social status (see Table 1 for demographics). Higher values indicate more formal education. For political party identification, lower values indicate a preference for Democrats and high values indicate a preference for Republicans. Higher values for social status indicate higher self-assessment of one's social status (being at the 'top' of society). For the ANCOVA, nominal variables were dummy-coded. Specifically, gender was dummy coded using male as a reference category with 0. Employment status was dummy coded using participants who are unemployed or unable to work as a reference category (coded with 0). Please refer to Supplement A for the rationale behind dummy coding.

2.2. Analysis plan

The main ANCOVA on pro-environmental behavior was preregistered, which is available at the OSF repository. We first conducted confirmatory factor analysis and reliability analysis of the PEB measure (see Supplement E), and then analyzed the manipulation reflections. After checking the statistical assumptions, we fit an ANCOVA model comparing the four conditions on pro-environmental behavior, with sociodemographic variables as covariates, followed by post hoc Tukey tests. ANCOVA was performed using the *aov* function in the R package *stats* (R Core Team, 2021). ANCOVAs yield marginal means by averaging over the covariates, and we obtained these using the R package *emmeans* (Lenth et al., 2021). Post hoc tests were performed with the *glht* function in the R package *multcomp* (Hothorn et al., 2021). Deviations from the preregistration are reported in Supplement A, like removing two covariates because they violated the statistical assumptions of ANCOVA (see further explanation below).

2.3. Study 1 Results

2.3.1. Manipulation reflections

Participants in all conditions responded to all the ad-hoc manipulation reflections. The purpose of these reflections was to measure the extent to which participants were aware of cognitive processes that might drive the order effects on behavior we investigated. These items should not be considered typical manipulation checks, but rather awareness of each manipulation factor. Differences between the conditions were tested using Kruskal-Wallis and pairwise Wilcoxon tests (see Supplement E for descriptive statistics and hypothesis tests). Participants in the identity condition scored the highest on the identity manipulation reflection, as expected, M (SD) = 2.34 (1.04). However, participants in the identity condition also showed the highest score in the intention manipulation reflection, M (SD) = 2.92 (1.21) and participants in the control condition showed the highest scores in the social norm manipulation reflection, M (SD) = 2.22 (1.11). These results suggest that participants were not consciously influenced by the social norm (i.e., what important people around them think) and intentions (i.e., how they would like to act in the future) manipulations when reporting their PEB. As biased responding is not necessarily conscious or intentional, we proceeded with our

Table 2
ANCOVA Predicting pro-environmental behavior.

Variable	Sum of Squares	*df*	Mean Square	*F*	*p*	Partial η^2
Overall model	26.7	11	2.43	8.90	< .001	
Condition	3.77	3	1.26	4.56	0.004	.014
Age	.424	1	.424	1.54	0.215	.002
Gender (female)	.163	1	.163	.593	0.442	.001
Education	4.47	1	4.47	16.2	< .001	.017
Unable to work	.12	1	.12	.434	0.510	.000
Student	1.42	1	1.42	5.13	0.024	.005
Part-time employment	.893	1	.893	3.24	0.072	.003
Full-time employment	2.23	1	2.23	8.09	0.005	.009
Political identity	13.2	1	13.2	48.0	< .001	.049
Residuals	258	935	.28			

Note. Predictors $p \leq 0.05$ are in bold. Environmentalist identity and social status were excluded from the model because they violated assumptions of ANCOVA. Political identity was coded from 1 (*strong Democrat*) to 7 (*strong Republican*).

Table 3
Estimated marginal means of PEB for each condition including standard errors and 95% CIs.

Condition			95% Confidence Interval		
	M	*SE*	Lower	Upper	*N*
Control	2.92	0.08	2.75	3.08	242
Identity	3.07	0.08	2.90	3.23	239
Intention	2.92	0.08	2.76	3.09	236
Norm	3.02	0.08	2.85	3.18	234

Note. These marginal means resulted from the ANCOVA model. Raw means by condition are reported in Supplement E.

analysis and discuss the manipulation reflection results in the Discussion section.

2.3.2. Experimental conditions and pro-environmental behavior

The primary confirmatory analysis was preregistered as ANCOVA testing differences in PEB between experimental conditions, with environmentalist identity and sociodemographic variables (age, gender, education, employment, social status, and political identity) as covariates predicting self-reported PEB. Sociodemographic variables were dummy-coded if they were not ordinal. The statistical assumptions were met for homogeneity of variances and normality. However, we then excluded environmentalist identity and social status as predictors because environmentalist identity violated the assumption of independence of covariates and experimental effect and social status violated the assumption of homogeneity of regression slopes. Including the two variables could have induced spurious relationships (Field et al., 2012, pp. 464–467). Further explanation is reported in Supplement E.

The ANCOVA model revealed there were differences in pro-environmental behavior between experimental conditions controlling for sociodemographic variables with Type II sums of squares, as the model did not include an interaction effect, $F(11) = 8.90$, $p < .001$. The experimental conditions differed in self-reported behavior with a small effect, $\eta^2_p = .01$. Additionally, higher formal education, being a student, full-time employment, and supporting the Democratic party predicted more behavior (see Table 2; see Table S5 in the Supplement for the covariate means).

One-tailed post-hoc Tukey tests revealed that participants in the Identity condition reported more PEB than participants in the Control condition, $t(935) = 3.09$, $p = .006$, $d = 0.12$. There was no difference between participants in the Intention and Control conditions, $t(935) = 0.09$, $p = .94$, nor between participants in the Norm and Control conditions, $t(935) = 2.03$, $p = .10$. The marginal means and 95% CIs of PEB adjusted for the covariates in the final ANCOVA model in each condition are depicted in Table 3 (see Figure S6 in Supplement E for the 95% CI of the mean differences).

2.4. Study 1 Discussion

Study 1 investigated whether measuring environment-related constructs results in misreporting of pro-environmental behavior (PEB) via order effects or priming. We hypothesized that the order effects could bias self-reported PEB. We did not find substantial effects to support this. Compared to a control condition, measuring intentions, or presenting norms did not result in overreporting of subsequently measured PEB, whereas measuring environmentalist identity resulted in a small effect of $d = 0.12$. One potential explanation for these findings is that the social desirability and relevance of acting pro-environmentally was already prevalent in the public discourse and participants' minds. Therefore, the control group would already be highly aware of the importance of PEB and be subject to social desirability bias and would not differ from the experimental groups.

In Study 2, we used an experimental approach (Unmatched Count Technique) to further investigate overreporting in pro-environmental behavior. We also account for the possible role of environmentalist identity, as previous studies have pointed towards the relevance of identity in social desirability (Brenner, 2014, 2017; Brenner and DeLamater, 2016). Moreover, environmentalist identity emerged as a source of bias in Study 1, albeit with a small effect.

3. Pilot study

A preregistered pilot study (https://osf.io/xnbq6; code, materials, and data available at the OSF repository) tested a revised measure of the Littering Prevention Behavior Scale (Ojedokun, 2016) used in Study 2. In addition to reducing the number of items, it included a dichotomous response format instead of a Likert scale. The original scale consisted of 41 items. Item selection for the shorter version was not based on statistical criteria but qualitative interpretation, as the original data was not available (see Supplement E for the selection process). The analysis consisted of evaluating factorial validity and reliability of the new, short scale. Sample size considerations for the pilot were not based on power analysis. 93 United States participants recruited from Prolific in March 2021 responded to 10 items of the Littering Prevention Behavior Scale (see Table 1 for demographics). The only criterion for participants' eligibility was residency in the United States. The pilot study's sample was fairly balanced with regards to gender and age, but university-educated, full-time employed, and Democratic-leaning participants were overrepresented. Two participants only completed the short scale, but not the sociodemographic items. They were retained and used in the factor analysis.

We fit a model with 10 items loading on one factor using a Robust Diagonal Weighted Least Squares Estimator. This CFA model indicated good fit, but four items showed relatively low factor loadings and non-significant *p*-values. We estimated a second model without the three

items with the lowest loadings. This second model also indicated good fit. Comparing the two models with the Likelihood-ratio test favored the second model. Therefore, those seven items were selected for Study 2.

4. Study 2

Study 2 used the Unmatched Count Technique (UCT) to investigate overreporting littering prevention behavior. Littering prevention behaviors are easy to do frequently and immediately reduce pollution and beautify the environment (Freije et al., 2019; Ibrahim et al., 2021; Kallgren et al., 2000). The UCT is an indirect questioning technique in which participants only reveal the number of statements that apply to them, but not which items (Coutts and Jann, 2011). Through this technique, participants' responses are more convincingly anonymized than in typical (online) surveys, and therefore socially desirable responses may be reduced. The UCT provides an experimental test of response bias and avoids the main pitfalls of using individual difference measures of social desirability, such as confounding third variables or common method bias (Jakobsen and Jensen, 2015; Schwarz et al., 2008).

4.1. Study 2 Methods

4.1.1. Sample

Sample size was based on a power analysis in G*Power (Erdfelder et al., 2009; see Supplement B). We recruited for a one-tailed *t*-test with power = 0.80, $\alpha = 0.05$, and $d = 0.26$. This effect size was extracted from a meta-analysis by Vesely and Klöckner (2020). The sample size was also determined by available funding. We oversampled by 10% to account for attrition. Thus, Study 2 aimed for 398 participants residing in the United States.

419 respondents were recruited and paid on Prolific in April 2021. The only criteria were adulthood and residency in the United States. 16 participants failed the attention check and were excluded as preregistered. 15 participants were excluded for exiting the survey before the critical UCT procedure. Three respondents answered all items except the sociodemographic questions at the end of the survey and were retained. The final sample was $N = 388$, with 190 randomly assigned to the standard condition and 198 to the covert condition (see Table 1 for demographics). Study 2's sample was fairly balanced in gender and age, but university-educated, employed, and Democrat participants were overrepresented.

4.1.2. Unmatched count technique procedure

Participants may feel more anonymous and be more willing to answer honestly if their responses to each individual item cannot be traced. Instead of indicating for each item whether it applies to them or not, in the UCT, participants indicate only how many of all the presented items apply. In this questioning setting, participants do not need to share that they *read the newspaper, play football, watch soap operas*, but *never recycle*; instead, they only share that three of the behaviors are true for them. Due to this design, the UCT does not allow conclusions about whether individual participants perform the specific behavior of interest (in this case, recycling). The UCT technique only provides one estimate of behavior frequency for the entire sample.

The UCT design typically consists of two randomly assigned experimental groups: covert and standard (see Table 4 for the study design; Dalton et al., 1994). The covert condition provides a list or set of items, and participants indicate how many items apply instead of answering each item individually. The item set is a combination of a) the item assumed to be sensitive and affected by social desirability, which in this study is littering prevention behavior, and b) general decoy" items about a range of different topics from daily life (e.g., *I have a dog*; *I have a dishwasher in my kitchen*; see Table 5). The standard condition receives a set of only the decoy items. Because the random assignment to condition should result in similar answers to the decoy items, subtracting these decoy estimates between conditions yields the prevalence of the sensitive item in an analysis method called the difference-in-means approach (Dalton et al., 1994). The UCT has been applied to obtain more accurate responses for voting, atheism, breastfeeding, sexual risk-taking, and alcohol consumption (Comşa and Postelnicu, 2013; Gervais and Najle, 2018; LaBrie and Earleywine, 2000; Lippitt et al., 2014).

Table 4
Study 2 Research design with the unmatched count technique (UCT).

	Standard Condition	Covert Condition
Random assignment	X	X
Covert responding to decoy items	X	X
Covert responding to littering prevention items		X
Direct responding to littering prevention items	X	
Direct responding to environmentalist identity	X	X
Direct responding to sociodemographic items	X	X

Note. Direct responding is the normal questionnaire format. A comparison between direct responding and covert responding is shown in Table 5.

Table 5
Example unmatched count technique item set.

Below you see [6/7] statements. Please indicate how many apply to you personally. Don't tell me exactly which of those are true and which are not. Only indicate at the bottom how many of those are true for you.

Standard Condition, Set 1	*Covert Condition, Set 1*
	I usually encourage friends and family to pick up their litter.
My birthday is on an even day.	My birthday is on an even day.
My job allows me to work from home more than once per week.	My job allows me to work from home more than once per week.
I have attended a professional soccer match.	I have attended a professional soccer match.
I have played scrabble.	I have played scrabble.
I exercise regularly.	I exercise regularly.
I don't smoke cigarettes.	I don't smoke cigarettes.
Number of statements that apply to you: __	

Note. The condition names in italics were not shown to participants. The litter item (emphasized here in bold) is the covert measure in this set. All littering prevention items were completed in both conditions. Participants responded to five sets of items, only one of which is shown here.

In this study, each participant received five item sets, all in the same order. For participants in the covert condition, each set consisted of six decoy items and one or two littering prevention items. Participants in the standard condition received five sets with only the same six decoy items. Using several sets of items allows calculating a composite score instead of using the answers to a single item as outcome variable. Every participant received a sequence of five item sets and indicated for each set how many items applied to them. When finished with their five sets, participants in the standard condition also reported their littering prevention behaviors in the traditional way, by answering each item individually. This enabled us to compare covert and standard estimates of the same behavior. Finally, all participants responded to comprehension checks and reported their environmentalist identity and sociodemographic information (see Table 1).

4.1.3. Randomization

190 participants were randomly assigned to the standard condition and 198 to the covert condition. The success of random assignment was assessed by comparing descriptive statistics across conditions (see Supplement E). Age, education, employment, and social status were comparable across conditions, but the descriptive statistics of gender and political identity differed slightly across conditions. We tested whether these differences were statistically significant with exploratory chi-square and

Wilcoxon's rank sum tests, which showed no statistically significant association between condition and gender, $\chi^2(3) = 1.60$, $p = .659$, nor political identity, $W = 19{,}638$, $p = .298$. Thus, we concluded that the experimental groups were similar and random assignment was successful.

4.1.4. Measures

The complete questionnaire is uploaded on OSF.

Littering prevention behavior. Study 2 included the 7-item Littering Prevention Behavior Scale short scale as the main outcome variable, which was developed and validated in the pilot study. The original 41-item scale was developed by Ojedokun (2016). In the covert condition, the 7 littering prevention items were spread evenly across the five item sets (that also included 6 decoy items each) and participants only indicated how many items in each set applied to them. The score was obtained by summing the number of items that applied to each participant across sets. This score was used in the main analysis and could range between 0 and 37 (five sets consisting of decoy items and the littering prevention items). As the analysis also requires the score for the standard condition (that only includes the decoy items), we summed the number of items across sets per participant for this condition as well and obtained a score ranging from 0 to 30 (five sets consisting of decoy items only). In the standard condition, all the 7 littering prevention items were used for standard responding and answered with "yes" (coded with 1) or "no" (coded with 0). Responses to each item were summed to obtain a score for the standard estimate, with a possible range between 0 and 7.

Decoy items. We used 30 decoy items asking about hobbies, interests, pets, or habits, all taken from previous studies applying the UCT (Gervais and Najle, 2018; Holbrook and Krosnick, 2010; Kalinin, 2016; LaBrie and Earleywine, 2000; Lippitt et al., 2014). Each of the five item sets included six decoy items.

Environmentalist identity. Environmentalist identity was measured with the same four-item scale by Brick and Lai (2018) used in Study 1, rated from 1 (*strongly disagree*) to 5 (*strongly agree*).

Comprehension check. The interpretation depends on participants following the directions correctly and trusting that their UCT answers were anonymous. All participants reported the extent to which they understood the instructions and to which they believed the technique guaranteed the anonymity of their answers from 1 (*strongly disagree*) to 5 (*strongly agree*).

Sociodemographic variables. Study 2 measured the same sociodemographic variables as study 1: age, gender, education, employment status, political party identification, and subjective social status (see Section 2.1.5 Measures). For the regressions, nominal variables were dummy-coded. Specifically, gender was dummy coded using male as reference category with 0. Despite being preregistered, employment status was removed from the model because the dummy codes resulted in errors. Please refer to Supplement A for more information on dummy coding.

4.2. Analysis plan

The key test compares the covert and standard mean estimates of littering prevention behavior to identify any social desirability effect or bias. The means were estimated based on two separate regression models and adjusted for the influence of covariates. We used the R-package *list* to fit a regression for obtaining the covert estimate, which was specifically developed for analyzing UCT designs (Blair et al., 2020; Blair and Imai, 2012). This regression model uses the total number of behaviors that apply to each participant as outcome variable while controlling for the number of decoy items and participants' condition (only decoy items vs. decoy items and the littering prevention items). Using the regression model results, we can estimate the covert mean of littering prevention for the covert condition as a whole, by considering the responses to decoy items in the standard condition and averaging over covariates (see Blair and Imai, 2012 for details). To obtain the standard estimate, we fit a linear model predicting the traditionally answered littering prevention behavior using the *glm* function. We obtained the estimated mean using the *emmeans* package (Lenth et al., 2021) by averaging over all the predictors. We then compared the estimated means in a two-sample *t*-test. The analysis plan was preregistered, and minor deviations are reported in Supplement A.

4.3. Study 2 Results

Confirmatory factor analysis (CFA) of the Littering Prevention Behavior short scale again indicated good model fit. The reliability coefficients were also good (Cronbach's $\alpha = 0.82$). The CFA details and descriptive statistics across sociodemographic groups are shown in Supplement E.

4.3.1. Comprehension checks

Participants in both conditions thought to a great extent that they understood the instructions: standard condition M (SD) = 4.86 (0.36), covert condition M (SD) = 4.87 (0.50). Participants also thought the anonymity of their answers was highly assured: standard condition M (SD) = 4.42 (0.83), covert condition M (SD) = 4.34 (0.94), out of 5.

4.3.2. Self-reported littering prevention by condition and drivers of responding

The main analysis consisted of four steps in an uncommon and relatively novel procedure. We describe each step below and in the analysis code uploaded to OSF.

First, we calculated the outcome for the covert regression model. The answers to each item set are saved as a separate variable, so there are ten behavior variables in total: one for each of the five covert item sets and five standard item sets. Each of the ten variables contained the number of behaviors applying to each participant in this set. We summed the five covert item sets to get a score of behaviors for the covert condition. We also summed the five standard item sets. These sums were combined into one composite variable indicating the total number of behaviors that applied to each participant across all sets and conditions. This variable was used as the outcome variable in the UCT regression analysis to estimate the covert mean. We also investigated the robustness of this composite variable by repeating the main analysis for each set separately, with consistent results (see Supplement E). The outcome variable in the regression analysis predicting the standard estimate was the sum score of traditionally answered littering prevention behavior.

In the second step, we tested for each item set whether respondents' answers to decoy items were affected by the inclusion of the socially desirable items. For the UCT to work, it is critical that responses are independent of each other and of the treatment condition (the assumption of no design effects; Blair and Imai, 2012). The assumption was satisfied in all item sets except for Set 3, p-values above 0.05 (applying Bonferroni correction as required by the R package *list;* Blair and Imai, 2012). Set 3 contained two littering prevention items instead of one, but the R function to test the UCT assumption was developed for only one socially desirable item. Thus, the function appeared to consider the second socially desirable item as a violation of the UCT assumption, even though this is not the case. Therefore, we concluded that the assumption was not violated.

In the third step, we fit two regression models to obtain two mean estimates: the covert model and the standard model. The regression model predicting the covert estimate of littering prevention behavior is depicted in Table 6. The covert mean estimate based on the regression model amounted to $M = 1.34$, $SE = 0.38$, 95% CI [0.597; 2.077]. Thus, participants in the covert condition reported 1.3 littering prevention behaviors on average. Littering prevention behavior was not significantly predicted by any of the covariates. This suggests that none of the predictors drove anonymous responses of pro-environmental behavior.

Table 6
Regression model of littering prevention behavior in the covert condition.

	b	*SE*	Lower 95% CI	Upper 95% CI
Intercept	0.29	2.32		
Environmentalist Identity	0.30	0.37	−0.43	1.03
Age	0.01	0.03	−0.05	0.07
Gender	−0.33	0.77	−1.84	1.19
Education	0.17	0.54	−0.90	1.24
Political Identity	0.17	0.20	−0.22	0.56
Social Status	−0.21	0.25	−0.70	0.28

Note. None of the predictors were significant; all p values > 0.05 based on the 95% CI. CIs were calculated with an online tool (Soper, 2021).

Table 7
Regression model of littering prevention behavior in the standard condition.

	b	*SE*	Lower 95% CI	Upper 95% CI
(Intercept)	−1.26	0.91	−3.05	0.53
Environmentalist Identity	**1.13**	**0.15**	0.84	1.43
Age	−0.01	0.01	−0.04	0.01
Gender	−0.14	0.29	−0.70	0.43
Education	0.20	0.21	−0.22	0.62
Political identity	0.06	0.08	−0.10	0.22
Social status	0.12	0.10	−0.07	0.31

Note. The one significant predictor is in bold.

The second regression model predicted the directly answered littering prevention behavior, resulting in a standard mean estimate of $M = 3.51$, $SE = 0.14$, 95% *CI* [3.23; 3.79]. Thus, participants in the standard condition reported 3.5 littering prevention behaviors on average. The model results are depicted in Table 7 and allow for the analysis of the drivers of misreporting behavior. People who identified more as environmentalists reported more littering prevention behaviors ($b = 1.13$, $\beta = 0.51$, $p < .001$, 95% CI [0.84; 1.43]) and all other predictors were nonsignificant. The interpretation of the models' regression coefficients was not part of the preregistered confirmatory analysis.

The fourth and final step consisted of testing H4. A one-tailed *t*-test established whether the conditions differed based on estimated means, standard deviations, and group sizes. We calculated the standard deviation of the covert estimate with $SE * N = 5.30$. We found a significant and large effect, indicating that participants reported much more littering prevention behavior in the standard condition than in the covert condition, $t(394) = 5.29$, $p < .001$, $d = 0.53$.

These analysis steps offer several advantages over most UCT studies. In the typical difference-in-means approach, the proportion of subjects involved in a given socially desirable behavior is a function of subtracting the mean of affirmative responses in one group from the mean of affirmative responses in the other group (Dalton et al., 1994). However, this does not allow to control for potential confounders that might influence socially desirable responding, analyze potential predictors of socially desirable responding, nor test the assumption of design effects, all of which were possible in the current analysis (Blair and Imai, 2012). Moreover, the regression approach reduces the standard error of the obtained mean estimate. This choice was pre-registered. To compare this regression approach with the typical differences-in-means approach, we also obtained the covert and standard estimate using the difference-in-means approach (see Supplement E). The results were consistent with the regression approach reported in the main analysis.

4.4. Study 2 Discussion

Study 2 revealed substantial overreporting in the standard condition. This is potentially consistent with a social desirability bias on self-reported pro-environmental behavior. Contrary to previous findings, exploratory analyses revealed that the participants who considered themselves environmentalists did *not* report more littering prevention behaviors in the covert condition than non-environmentalists. However, environmentalists did report more behaviors in the standard condition, which suggests that identity may drive socially desirable responding of PEB. This pattern is consistent with the interpretation of Study 1 that all participants were already aware of the importance of PEB, which could explain why priming did not change self-reports.

5. General discussion

Self-reported scales of pro-environmental behavior (PEB) are prone to misreporting, which might be due to systematic measurement error. We tested whether responding to common predictor scales or an instruction (Study 1) and answering a standard questionnaire compared to providing more anonymous responses (Study 2) could cause overreporting of PEB.

Study 1 did not find substantial effects but filling out an identity questionnaire led to slightly overreporting PEB. This is also consistent with research outside of environmental psychology: several priming studies have used single identity questions or established identity questionnaires to increase identity salience (Diamond, 2020; Howard and Borgella, 2018; Transue, 2007). Therefore, including an identity scale in a survey may prime identity and affect self-reported behavior whether or not that priming is intended.

There are several potential explanations for the null effects found for the intentions (H1) and social norms (H3) conditions. First, participants are likely to already be well-aware of the norms and values associated with PEB and therefore self-reported behavior measures may already contain social desirability bias (i.e., also in the control condition). Furthermore, previous research has established the question-behavior effect resulting from measuring intentions as robust. However, these meta-analyses did not include PEB and considered both observations and self-reports (Wilding et al., 2016; Wood et al., 2016). Thus, the question-behavior effect might not generalize to self-reported PEB. The effect of social norms on PEB is generally less robust, as evidence for social norm effects has been generally mixed (Farrow et al., 2017; Gifford, 2014). A potential explanation for the non-significant findings in this research is that the instruction highlighting norms or the intention measure were too subtle to create an effect or were not properly read by participants. In line with that, the results of the manipulation reflections do not support our assumptions of participants' cognitive processes while answering the experimental manipulations, which suggests that the measures did not elicit the expected order effects. However, the manipulation reflections were ad-hoc and not previously validated. Also, Study 1's research design might account for these non-significant effects. Future research with a within-subjects design might better detect PEB overreporting but might also result in bias if participants repeatedly answer the same behavior questions within a short period.

Study 2 provided evidence for bias in self-reported PEB (H4) and showed an unusually large effect for experimental PEB research (Bamberg and Möser, 2007; Klöckner, 2013). Participants who responded to a standard measure of littering prevention behavior reported more PEB than participants who responded to a covert measure based on the Unmatched Count Technique. Moreover, if participants expressed a stronger environmentalist identity, they also reported more PEB in the standard condition. If participants were able to report anonymously, environmentalist identity did not predict reported past behavior. This provides support for the idea that social desirability is a form of expressing social identity, and that surveys are utilized as a low-cost and easy opportunity for performing this identity (Brenner and DeLamater, 2016). A previous meta-analysis found evidence for social desirability bias in PEB research but with a considerably smaller effect size (Vesely and Klöckner, 2020). Because those studies used individual difference scales of social desirability, the different research designs might account for this difference in effect sizes. In sum, the observed bias in Study 2 is consistent with the claim that many of the surveys that measured behavior among

a social group connected with this behavior (e.g., pro-environmental activism among politically left-leaning university students) could be distorted by measurement errors due to social desirability.

The effect found in Study 2 is also larger than in many UCT studies, though social desirability UCT effects are heterogeneous across different domains such as voting, support for authoritarian regimes, prejudice, or vote buying (G. Blair, Coppock, et al., 2020; Ehler et al., 2021). The variation in effect sizes can be partly explained with variations in sample size and study design. The UCT has high variance, so it requires large samples to detect smaller social desirability effects. Thus, many UCT studies are underpowered to detect social desirability. In the current study, we reduced variance by analyzing data with the regression approach and balancing control items (Blair et al., 2020). Moreover, conducting interviews face-to-face, adding sensitive control items, and whether the over- or underreporting is measured also explains heterogeneity in effect sizes (Ehler et al., 2021). However, this heterogeneity also indicates that a replication of this UCT study might find a different effect size.

5.1. Limitations

We measured thoughts and deliberations during Study 1. However, these manipulation reflections were neither previously validated, nor did they support our assumptions of the deliberate processes behind PEB responding. This suggests a potential disconnect that could be explained either by the manipulations or the manipulation reflections not functioning as expected. Notably, there is still no robust and consistent empirical evidence on the cognitive processes underlying the question-behavior effect or in which situations priming effects occur and why (Caruso et al., 2017; Molden, 2014; Spangenberg et al., 2016). This complicates the creation of robust and valid manipulation reflections. Both the experimental manipulations, the reflection items, and the underlying mechanisms should thus be interpreted with caution, as they can only offer face validity. Future research could test the experimental validations beforehand to establish construct validity and evaluate the manipulation reflections to further understand the cognitive processes behind misreporting (Chester and Lasko, 2021).

Second, the validity of the UCT has also been criticized despite its increasing popularity and its advantages compared to social desirability questionnaires (Gosen et al., 2019; Kramon and Weghorst, 2019). According to criticism, the UCT may yield inconsistent results, its accuracy decreases in more complex designs, and participants respond more affirmatively to covert items, especially with many items. In this study, several points support the validity of the UCT. The manipulation reflections in Study 2 demonstrated that participants overwhelmingly understood the instructions and felt anonymous in the covert condition. There was no evidence for design effects: participants' responses were independent of other items in the set and of their assigned condition (see Blair and Imai, 2012). Participants who reported PEB covertly did not respond more affirmatively than participants in the standard condition, even though the covert condition included more items. Employing the UCT study design might reduce social desirability bias, but we did not specifically measure social desirability in these studies. Self-reports are also subject to other measurement biases.

We assume that the generalization of the current findings is limited to the populations typically used in psychological research: well-educated, North American-based, or Western, politically left-leaning samples mostly of people in early or middle adulthood. As this characterization also applies to our samples, the effect sizes we found might be smaller or non-significant in the general population or the population of different countries and societies. The findings may well change in different samples, for example because their responses are influenced by different social identities and norms, e.g., some participants would find it socially desirable to report less PEB depending on their social identities (Brick et al., 2017). As these assumptions regarding generalizability are speculative, we encourage future research to replicate the design with different populations.

The Study 1 materials were designed to test for overreporting of behaviors that are typically seen as positive. As such, the materials are likely unsuitable to bias a population towards underreporting, or if the populations are characterized by strong aversions towards environmentalism. This is illustrated by the finding that none of the experimental conditions reported less PEB than the control condition. The generalizability of the behavioral measurement is also limited. The measures in both studies capture behaviors that are recurring, daily, and do not require large investments of time and money. Hence, they cannot be generalized to high-impact, lower-frequency PEBs like the purchase of an electric car or solar panels (Nielsen et al., 2021).

5.2. Future directions

Future research could directly measure the mechanisms of bias in self-reports and develop techniques to reduce the bias. We found that self-reported pro-environmental behavior (PEB) may have considerable overreporting. Future work could measure social desirability directly as a mechanism. Other studies could evaluate how different social identities affect PEB across different contexts and study designs such as objective vs. self-reported behavior. Furthermore, future research could focus on other kinds of measurement error. Self-reports can also be influenced by memory bias, inaccurate self-perception, subjective interpretations of items and response categories, or mood (Eisenhower et al., 1991; Kormos and Gifford, 2014; Lange and Dewitte, 2019; N. Schwarz and Clore, 1983). These sources of bias are treated as random error and supposedly become irrelevant in typical research designs where multiple items and scores are used. As these biases particularly affect self-report measures, they should be further investigated to explain the weak correlations between self-reported and objectively measured PEBs. This study was concerned with biases in self-report measures and did not address the measurement biases affecting objective PEB measures such as meter readings or observations (Kormos and Gifford, 2014). For instance, meter readings capture the energy use of a household, whereas questionnaires typically measure behavior for a single person. Additionally, reports on a person's behavior by trained observers or peers are often operationalized using questionnaires, making these measures also susceptible to response biases. Similarly, pro-environmental behavior observed in the lab can be affected by social desirability due to the presence of the researcher and other participants (Brick and Sherman, 2021). In sum, objective measures also have measurement error.

5.3. Conclusion

Overall, we showed that response patterns in pro-environmental behavior measures are partially explained by questionnaire design. In many domains, self-report scales lack accuracy compared to objectively measured data (Gorber et al., 2007; Kobayashi and Boase, 2012; Prince et al., 2008; Shields et al., 2008; Spielholz et al., 2001; Streff and Wagenaar, 1989). Therefore, further investigating sources of bias in other self-report instruments is worthwhile to increase the accuracy of such measures. Study 1 showed that measuring an identity led to slight overreporting of self-reported behavior. Study 2 showed that standard questionnaires can lead to considerable misreporting of PEB, and this finding suggests that personality-based social desirability scales may underestimate this impact. Future research could also use indirect questioning techniques to estimate and prevent measurement bias.

Techniques such as the UCT can potentially produce more accurate estimates of self-reported PEB but it remains unclear to what extent different sources of bias can be reduced with anonymizing techniques. We suggest that standard environmental psychology questionnaires about socially desirable behaviors could lead to overestimation, and this bias may not affect each individual equally.

One potential solution is to better separate research that is focused on how people think about and reflect on their pro-environmental behavior from research about their actual environmental impact. For projects depending on environmental impact, we recommend researchers consider whether traditional self-report scales of repeated, conscious behaviors are appropriate (Nielsen et al., 2021, 2022). In sum, environmental psychology should not solely rely on self-reported measures of behavior, consistent with recent calls for impact-focused measurement (Nielsen et al., 2021). Rather, these self-reported measures should be complemented with objective measures of PEB which are becoming increasingly available from digital trace data, such as energy consumption from meter readings. Combining survey PEB and independent measures from digital traces will allow environmental researchers to enhance the quality of their estimates, for instance by modeling the various errors in all sources simultaneously (using latent variable modeling) and obtaining error-corrected estimates. Overall, we encourage focusing on measurement validation to build a stronger basis for testing theories and developing interventions, innovations, and policies.

Financial disclosure/funding

We have no conflict of interest to disclose. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of Competing Interest

The authors declared no potential conflicts of interests with respect to the authorship and/or publication of this article.

CRediT authorship contribution statement

Katharina Koller: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization, Project administration. **Paulina K. Pankowska:** Conceptualization, Writing – review & editing. **Cameron Brick:** Conceptualization, Methodology, Validation, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Data availability

All data shared at OSF.

Acknowledgement

We thank Florian Lange and the anonymous reviewers for helpful feedback.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.cresp.2022.100087.

References

Alwin, D.F., 2007. Margins of Error: A Study of Reliability in Survey Measurement. Wiley-Interscience.

Bamberg, S., Möser, G., 2007. Twenty years after Hines, Hungerford, and Tomera: a new meta-analysis of psycho-social determinants of pro-environmental behaviour. J. Environ. Psychol. 27 (1), 14–25. doi:10.1016/j.jenvp.2006.12.002.

Barger, S.D., 2002. The Marlowe-Crowne affair: short forms, psychometric structure, and social desirability. J. Pers. Assess. 79 (2), 286–305. doi:10.1207/S15327752JPA7902_11.

Barr, S., 2007. Factors influencing environmental attitudes and behaviors: a U.K. case study of household waste management. Environ. Behav. 39 (4), 435–473. doi:10.1177/0013916505283421.

Beretvas, S.N., Meyers, J.L., Leite, W.L., 2002. A reliability generalization study of the Marlowe-Crowne social desirability scale. Educ. Psychol. Meas. 62 (4), 570–589. doi:10.1177/0013164402062004003.

Bhattacherjee, A., 2012. Social Science Research: Principles, Methods, and Practices, 2nd ed. Anol Bhattacherjee.

Biemer, P.P., 2010. Latent Class Analysis of Survey Error. John Wiley & Sons, Inc doi:10.1002/9780470891155.

Blair, G., Chou, W., Imai, K., Park, B., Coppock, A., 2020a. List: Statistical Methods for the Item Count Technique and List Experiment.

Blair, G., Coppock, A., Moor, M., 2020b. When to worry about sensitivity bias: a social reference theory and evidence from 30 years of list experiments. American Polit. Sci. Rev. 114 (4), 1297–1315. doi:10.1017/S0003055420000374.

Blair, G., Imai, K., 2012. Statistical analysis of list experiments. Polit. Anal. 20 (1), 47–77. doi:10.1093/pan/mpr048.

Brenner, P.S., 2014. Testing the veracity of self-reported religious practice in the muslim world. Soc. Forces 92 (3), 1009–1037. doi:10.1093/sf/sot120.

Brenner, P.S., 2017. Toward a social psychology of survey methodology: an application of the approach and directions for the future. Sociol. Compass. 11 (7). doi:10.1111/soc4.12491, e12491–e12491.

Brenner, P.S., DeLamater, J., 2016. Lies, damned lies, and survey self-reports? identity as a cause of measurement bias. Soc. Psychol. Q. 79 (4), 333–354. doi:10.1177/0190272516628298.

Brick, C., Lai, C.K., 2018. Explicit (but not implicit) environmentalist identity predicts pro-environmental behavior and policy preferences. J. Environ. Psychol. 58, 8–17. doi:10.1016/j.jenvp.2018.07.003.

Brick, C., Sherman, D.K., 2021. When does being watched change pro-environmental behaviors in the laboratory? Sustainability 13 (5). doi:10.3390/su13052766, 2766–2766.

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. doi:10.1016/j.jenvp.2017.04.004.

Carfora, V., Caso, D., Sparks, P., Conner, M., 2017. Moderating effects of pro-environmental self-identity on pro-environmental intentions and behaviour: a multi-behaviour study. J. Environ. Psychol. 53, 92–99. doi:10.1016/j.jenvp.2017.07.001.

Caruso, E.M., Shapira, O., Landy, J.F., 2017. Show me the money: a systematic exploration of manipulations, moderators, and mechanisms of priming effects. Psychol. Sci. 28 (8), 1148–1159. doi:10.1177/0956797617706161.

Chao, Y.L., 2012. Predicting people's environmental behaviour: theory of planned behaviour and model of responsible environmental behaviour. Environ. Educ. Res. 18 (4), 437–461. doi:10.1080/13504622.2011.634970.

Chao, Y.L., Lam, S.P., 2011. Measuring responsible environmental behavior: self-reported and other-reported measures and their differences in testing a behavioral model. Environ. Behav. 43 (1), 53–71. doi:10.1177/0013916509350849.

Chaplin, G., Wyton, P., 2014. Student engagement with sustainability: understanding the value-action gap. Int. J. Sustain. Higher Educ. 15 (4), 404–417. doi:10.1108/IJSHE-04-2012-0029.

Chester, D.S., Lasko, E.N., 2021. Construct validation of experimental manipulations in social psychology: current practices and recommendations for the future. Perspect. Psychol. Sci. 16 (2), 377–395. doi:10.1177/1745691620950684.

Cialdini, R.B., Kallgren, C.A., Reno, R.R., 1991. A focus theory of normative conduct: a theoretical refinement and reevaluation of the role of norms in human behavior. Adv. Exp. Soc. Psychol. 24, 201–234. doi:10.1016/S0065-2601(08)60330-5.

Comşa, M., Postelnicu, C., 2013. Measuring social desirability effects on self-reported turnout using the item count technique. Int. J. Public Opin. Res. 25 (2), 153–172. doi:10.1093/ijpor/eds019.

Corneille, O., Lush, P., 2022. Sixty years after Orne's American psychologist article: a conceptual framework for subjective experiences elicited by demand characteristics. Personal. Soc. Psychol. Rev. doi:10.1177/10888683221104368, 108886832211043.

Coutts, E., Jann, B., 2011. Sensitive questions in online surveys: experimental results for the randomized response technique (RRT) and the unmatched count technique (UCT). Sociol. Methods Res. 40 (1), 169–193. doi:10.1177/0049124110390768.

Crocker, L.M., Algina, J., 1986. Introduction to Classical and Modern Test Theory. Holt, Rinehart, and Winston.

Dalton, D.R., Wimbush, J.C., Daily, C.M., 1994. Using the Unmatched Count Technique (Uct) to estimate base rates for sensitive behavior. Pers. Psychol. 47 (4), 817–829. doi:10.1111/j.1744-6570.1994.tb01578.x.

Diamond, E.P., 2020. The influence of identity salience on framing effectiveness: an experiment. Polit. Psychol 41 (6), 1133–1150. doi:10.1111/pops.12669.

Dillman, D.A., Smyth, J.D., Christian, L.M., 2014. Internet, Phone, Mail, and Mixed-Mode Surveys: The tailored Design Method, 4th ed. Wiley.

Doherty, C., Kiley, J., Asheer, N., Jordan, C., 2022. Public's Top Priority for 2022: Strengthening the Nation's Economy. Pew Research Center.

Dono, J., Webb, J., Richardson, B., 2010. The relationship between environmental activism, pro-environmental behaviour and social identity. J. Environ. Psychol. 30 (2), 178–186. doi:10.1016/j.jenvp.2009.11.006.

Egleston, B.L., Miller, S.M., Meropol, N.J., 2011. The impact of misclassification due to survey response fatigue on estimation and identifiability of treatment effects. Stat. Med. 30 (30), 3560–3572. doi:10.1002/sim.4377.

Ehler, I., Wolter, F., Junkermann, J., 2021. Sensitive Questions in Surveys: a comprehensive meta-analysis of experimental survey studies on the performance of the item count technique. Public Opin. Q. 85 (1), 6–27. doi:10.1093/poq/nfab002.

Eisenhower, D., Mathiowetz, N.A., Morganstein, D, 1991. Recall error: sources and bias reduction techniques. In: Biemer, P.P., Groves, R.M., Lyberg, L.E., Mathiowetz, N.A., Sudman, S. (Eds.), Measurement Errors in Surveys. Wiley, pp. 125–144. doi:10.1002/9781118150382.ch8.

Erdfelder, E., FAul, F., Buchner, A., Lang, A.G., 2009. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behav. Res. Methods 41 (4), 1149–1160. doi:10.3758/BRM.41.4.1149.

European Parliament, European Commission, & Directorate-General for Communication, 2021. Future of Europe: Report. Publications Office.
Farrow, K., Grolleau, G., Ibanez, L., 2017. Social norms and pro-environmental behavior: a review of the evidence. Ecol. Econ. 140, 1–13. doi:10.1016/j.ecolecon.2017.04.017.
Field, A.P., Miles, J., Field, Z., 2012. Discovering Statistics Using R. Sage.
Findor, A., Lášticová, B., Hruška, M., Popper, M., Váradi, L., 2020. The impact of response instruction and target group on the BIAS map. Front. Psychol. 11. doi:10.3389/fpsyg.2020.566725.
Freije, A.M., Naser, H.A., Abdulla, K.H., 2019. Attitudes and opinions towards public littering in the Kingdom of Bahrain. Arab J. Basic Appl. Sci. 26 (1), 354–361. doi:10.1080/25765299.2019.1628688.
Gervais, W.M., Najle, M.B., 2018. How many atheists are there? Soc. Psychol. Personal. Sci. 9 (1), 3–10. doi:10.1177/1948550617707015.
Gifford, R., 2014. Environmental psychology matters. Annu. Rev. Psychol. 65, 541–579. doi:10.1146/annurev-psych-010213-115048.
Gkargkavouzi, A., Halkos, G., Matsiori, S., 2019. Environmental behavior in a private-sphere context: integrating theories of planned behavior and value belief norm, self-identity and habit. Resour. Conserv. Recycl. 148 (January), 145–156. doi:10.1016/j.resconrec.2019.01.039.
Godin, G., Bélanger-Gravel, A., Amireault, S., Vohl, M.-C., Pérusse, L., 2011. The effect of mere-measurement of cognitions on physical activity behavior: a randomized controlled trial among overweight and obese individuals. Int. J. Behav. Nutr. Phys. Act. 8 (1), 2. doi:10.1186/1479-5868-8-2.
Godin, G., Bélanger-Gravel, A., Vézina-Im, L.A., Amireault, S., Bilodeau, A., 2012. Question-behaviour effect: a randomised controlled trial of asking intention in the interrogative or declarative form. Psychol. Health 27 (9), 1086–1099. doi:10.1080/08870446.2012.671617.
Gorber, S.C., Tremblay, M., Moher, D., Gorber, B., 2007. A comparison of direct vs. Self-report measures for assessing height, weight and body mass index: a systematic review. Obes. Rev. 8 (4), 307–326. doi:10.1111/j.1467-789X.2007.00347.x.
Gosen, S., Schmidt, P., Thörner, S., Leibold, J., 2019. Is the list experiment doing its job? In: Einstellungen Und Verhalten in Der Empirischen Sozialforschung. Springer Fachmedien Wiesbaden, pp. 179–205. doi:10.1007/978-3-658-16348-8_8.
Harland, P., Staats, H., Wilke, H.A.M., 1999. Explaining proenvironmental intention and behavior by personal norms and the theory of planned behavior. J. Appl. Soc. Psychol. 29 (12), 2505–2528. doi:10.1111/j.1559-1816.1999.tb00123.x.
Harrison, C.K., Stone, J., Shapiro, J., Yee, S., Boyd, J.A., Rullan, V., 2009. The role of gender identities and stereotype salience with the academic performance of male and female college athletes. J. Sport Soc. Issues 33 (1), 78–96. doi:10.1177/0193723508328902.
Holbrook, A.L., Krosnick, J.A., 2010. Measuring voter turnout by using the randomized response technique: evidence calling into question the method's validity. Public Opin. Q. 74 (2), 328–343. doi:10.1093/poq/nfq012.
Hothorn, T., Bretz, F., Westfall, P., Heiberger, R.M., Schuetzenmeister, A., Scheibe, S., 2021. Package 'multcomp.'.
Howard, S., Borgella, A., 2018. Sinking" or sinking? Identity salience and shifts in Black women's athletic performance. Psychol. Sport. Exerc. 39, 179–183. doi:10.1016/j.psychsport.2018.08.016.
Howell, A.P., Shaw, B.R., Alvarez, G., 2015. Bait shop owners as opinion leaders: a test of the theory of planned behavior to predict pro-environmental outreach behaviors and intentions. Environ. Behav. 47 (10), 1107–1126. doi:10.1177/0013916514539684.
Ibrahim, H., Mariapan, M., Lin, E.L.A., Bidin, S., 2021. Environmental concern, attitude and intention in understanding student's anti-littering behavior using structural equation modeling. Sustainability 13 (8). doi:10.3390/su13084301.
Jakobsen, M., Jensen, R., 2015. Common method bias in public management studies. Int. Public Manag. J. 18 (1), 3–30. doi:10.1080/10967494.2014.997906.
John, L.K., Loewenstein, G., Acquisti, A., Vosgerau, J., 2018. When and why randomized response techniques (fail to) elicit the truth. Organ. Behav. Hum. Decis. Process. 148, 101–123. doi:10.1016/j.obhdp.2018.07.004.
Kalinin, K., 2016. The social desirability bias in autocrat's electoral ratings: evidence from the 2012 Russian presidential elections. J. Elect. Public Opin. Part. 26 (2), 191–211. doi:10.1080/17457289.2016.1150284.
Kallgren, C.A., Reno, R.R., Cialdini, R.B., 2000. A focus theory of normative conduct: when norms do and do not affect behavior. Personal. Soc. Psychol. Bull. 26 (8), 1002–1012. doi:10.1177/01461672002610009.
Klar, S., 2013. The influence of competing identity primes on political preferences. J. Polit. 75 (4), 1108–1124. doi:10.1017/S0022381613000698.
Klöckner, C.A., 2013. A comprehensive model of the psychology of environmental behaviour-a meta-analysis. Glob. Environ. Change 23 (5), 1028–1038. doi:10.1016/j.gloenvcha.2013.05.014.
Kobayashi, T., Boase, J., 2012. No such Effect? The implications of measurement error in self-report measures of mobile communication use. Commun. Methods Meas. 6 (2), 126–143. doi:10.1080/19312458.2012.679243.
Kormos, C., Gifford, R., 2014. The validity of self-report measures of proenvironmental behavior: Ameta-analytic review. J. Environ. Psychol. 40, 359–371. doi:10.1016/j.jenvp.2014.09.003.
Kotzur, P.F., Veit, S., Namyslo, A., Holthausen, M., Wagner, U., Yemane, R., 2020. 'Society thinks they are cold and/or incompetent, but I do not': stereotype content ratings depend on instructions and the social group's location in the stereotype content space. Br. J. Soc. Psychol. 59 (4), 1018–1042. doi:10.1111/bjso.12375.
Kramon, E., Weghorst, K., 2019. Mis)Measuring sensitive attitudes with the list experiment. Public Opin. Q. 83 (S1), 236–263. doi:10.1093/poq/nfz009.
Kroesen, M., Chorus, C., 2018. The role of general and specific attitudes in predicting travel behavior – a fatal dilemma? Travel Behav. Soc. 10 (September 2017), 33–41. doi:10.1016/j.tbs.2017.09.004.
LaBrie, J.W., Earleywine, M., 2000. Sexual risk behaviors and alcohol: higher base rates revealed using the unmatched-count technique. J. Sex Res. 37 (4), 321–326. doi:10.1080/00224490009552054.
Lalonde, R.N., Silverman, R.A., 1994. Behavioral preferences in response to social injustice: the effects of group permeability and social identity salience. J. Pers. Soc. Psychol. 66 (1), 78–85. doi:10.1037/0022-3514.66.1.78.
Lalot, F., Quiamzade, A., Falomir-Pichastor, J.M., Gollwitzer, P.M., 2019. When does self-identity predict intention to act green? A self-completion account relying on past behaviour and majority-minority support for pro-environmental values. J. Environ. Psychol. 61, 79–92. doi:10.1016/j.jenvp.2019.01.002.
Lange, F., Dewitte, S., 2019. Measuring pro-environmental behavior: review and recommendations. J. Environ. Psychol. 63, 92–100. doi:10.1016/j.jenvp.2019.04.009.
Lange, F., Dewitte, S., 2021a. Test-retest reliability and construct validity of the pro-environmental behavior task. J. Environ. Psychol. 73. doi:10.1016/j.jenvp.2021.101550, 101550–101550.
Lange, F., Dewitte, S., 2021b. The work for environmental protection task: a consequential web-based procedure for studying pro-environmental behavior. Behav. Res. Methods doi:10.3758/s13428-021-01617-2.
Lange, F., Steinke, A., Dewitte, S., 2018. The Pro-environmental behavior task: a laboratory measure of actual pro-environmental behavior. J. Environ. Psychol. 56, 46–54. doi:10.1016/j.jenvp.2018.02.007.
Lensvelt-Mulders, G.J.L.M., Hox, J.J., Van Der Heijden, P.G.M., Maas, C.J.M., 2005. Meta-analysis of randomized response research: thirty-five years of validation. Sociol. Methods Res. 33 (3), 319–348. doi:10.1177/0049124104268664.
Lenth, R.V., Buerkner, P., Herve, M., Love, J., Riebl, H., Singmann, H., 2021. Package 'emmeans.'.
Lippitt, M., Reese Masterson, A., Sierra, A., Davis, A.B., White, M.A., 2014. An exploration of social desirability bias in measurement of attitudes toward breastfeeding in public. J. Human Lact. 30 (3), 358–366. doi:10.1177/0890334414529020.
McCambridge, J., Day, M., 2008. Randomized controlled trial of the effects of completing the Alcohol Use Disorders Identification Test questionnaire on self-reported hazardous drinking. Addiction 103 (2), 241–248. doi:10.1111/j.1360-0443.2007.02080.x.
McCambridge, J., de Bruin, M., Witton, J., 2012. The effects of demand characteristics on research participant behaviours in non-laboratory settings: a systematic review. PLoS One 7 (6). doi:10.1371/journal.pone.0039116, e39116–e39116.
McGlone, M.S., Aronson, J., 2006. Stereotype threat, identity salience, and spatial reasoning. J. Appl. Dev. Psychol. 27 (5), 486–493. doi:10.1016/j.appdev.2006.06.003.
Milfont, T.L., 2009. The effects of social desirability on self-reported environmental attitudes and ecological behaviour. Environmentalist 29 (3), 263–269. doi:10.1007/s10669-008-9192-2.
Milfont, T.L., Zubielevitch, E., Milojev, P., Sibley, C.G., 2021. Ten-year panel data confirm generation gap but climate beliefs increase at similar rates across ages. Nat. Commun. 12 (1), 4038. doi:10.1038/s41467-021-24245-y.
Molden, D.C., 2014. Understanding priming effects in social psychology: what is "social priming" and how does it occur? Soc. Cogn. 32 (Supplement), 1–11. doi:10.1521/soco.2014.32.supp.1.
Morren, M., Grinstein, A., 2016. Explaining environmental behavior across borders: a meta-analysis. J. Environ. Psychol. 47, 91–106. doi:10.1016/j.jenvp.2016.05.003.
Morwitz, V.G., Fitzsimons, G.J., 2004. The mere-measurement effect: why does measuring intentions change actual behavior? J. Consum. Psychol. 14 (1–2), 64–74. doi:10.1207/s15327663jcp1401.
Moser, S., Kleinhückelkotten, S., 2018. Good intents, but low impacts: diverging importance of motivational and socioeconomic determinants explaining pro-environmental behavior, energy use, and carbon footprint. Environ. Behav. 50 (6), 626–656. doi:10.1177/0013916517710685.
Niaura, A., 2013. Using the theory of planned behavior to investigate the determinants of environmental behavior among youth. Environ. Res. Eng. Manag. 63 (1). doi:10.5755/j01.erem.63.1.2901.
Nielsen, K.S., Brick, C., Hofmann, W., Joanes, T., Lange, F., Gwozdz, W., 2022. The motivation–impact gap in pro-environmental clothing consumption. Nat. Sustain. doi:10.1038/s41893-022-00888-7.
Nielsen, K.S., Cologna, V., Lange, F., Brick, C., Stern, P.C., 2021. The case for impact-focused environmental psychology. J. Environ. Psychol. 74. doi:10.1016/j.jenvp.2021.101559, 101559–101559.
Ojedokun, O., 2016. Development and psychometric evaluation of the littering prevention behavior scale. Ecopsychology 8 (2), 138–152. doi:10.1089/eco.2015.0081.
Pankowska, P., Bakker, B., Oberski, D., Pavlopoulos, D., 2021. Dependent interviewing: a remedy or a curse for measurement error in surveys? Surv. Res. Methods 135–146. doi:10.18148/SRM/2021.V15I2.7640.
Pankowska, P., Pavlopoulos, D., Bakker, B., Oberski, D.L., 2020. Reconciliation of inconsistent data sources using hidden Markov models. Stat. J. IAOS 36 (4), 1261–1279. doi:10.3233/SJI-190594.
Perreault, W.D., 1975. Controlling order-effect bias. Public Opin. Quart. 39 (4), 544–551 JSTOR.
Pfattheicher, S., Sassenrath, C., Schindler, S., 2016. Feelings for the suffering of others and the environment: compassion fosters proenvironmental tendencies. Environ. Behav. 48 (7), 929–945. doi:10.1177/0013916515574549.
Ployhart, R.E., Ehrhart, M.G., 2003. Be careful what you ask for: effects of response instructions on the construct validity and reliability of situational judgment tests. Int. J. Sel. Assess. 11 (1), 1–16. doi:10.1111/1468-2389.00222.
Prince, S.A., Adamo, K.B., Hamel, M.E., Hardt, J., Connor Gorber, S., Tremblay, M., 2008. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. Int. J. Behav. Nutr. Phys. Act. 5. doi:10.1186/1479-5868-5-56.
R Core Team, 2021. R: A language and Environment For Statistical Computing. R Foundation for Statistical Computing.

Redondo, I., Puelles, M., 2017. The connection between environmental attitude–behavior gap and other individual inconsistencies: a call for strengthening self-control. Int. Res. Geograph. Environ. Educ. 26 (2), 107–120. doi:10.1080/10382046.2016.1235361.

Rentfrow, P.J., Gosling, S.D., 2003. The Do Re Mi's of everyday life: the structure and personality correlates of music preferences. J. Pers. Soc. Psychol. 84 (6), 1236–1256. doi:10.1037/0022-3514.84.6.1236.

Saris, W.E., Gallhofer, I.N., 2007. Design, Evaluation, and Analysis of Questionnaires for Survey Research. John Wiley & Sons, Inc doi:10.1002/9780470165195.

Schmitt, M.T., Mackay, C.M.L., Droogendyk, L.M., Payne, D., 2019. What predicts environmental activism? The roles of identification with nature and politicized environmental identity. J. Environ. Psychol. 61, 20–29. doi:10.1016/j.jenvp.2018.11.003.

Schultz, P.W., Khazian, A.M., Zaleski, A.C., 2008. Using normative social influence to promote conservation among hotel guests. Soc. Influ. 3 (1), 4–23. doi:10.1080/15534510701755614.

Schwarz, A., Schwarz, C., & Rizzuto, T. (2008). Examining the "urban legend" of common method bias: nine common errors and their impact. 441–441. https://doi.org/10.1109/HICSS.2008.155

Schwarz, N., Clore, G.L., 1983. Mood, misattribution, and judgments of well-being: informative and directive functions of affective states. J. Pers. Soc. Psychol. 45 (3), 513–523. doi:10.1037/0022-3514.45.3.513.

Serenko, A., Bontis, N., 2013. First in, best dressed: the presence of order-effect bias in journal ranking surveys. J. Informetr. 7 (1), 138–144. doi:10.1016/j.joi.2012.10.005.

Sheeran, P., Webb, T.L., 2016. The intention–behavior gap. Soc. Personal. Psychol. Compass 10 (9), 503–518. doi:10.1111/spc3.12265.

Shields, M., Gorber, S.C., Tremblay, M.S., 2008. Estimates of Obesity Based on Self-Report Versus Direct Measures Catalogue No. 82-003-X Health Reports.

Shih, M., Pittinsky, T.L., Ambady, N., 1999. Stereotype susceptibility: identity salience and shifts in quantitative performance. Psychol. Sci. 10 (1), 80–83. doi:10.1111/1467-9280.00111.

Shove, E., 2010. Beyond the ABC: climate change policy and theories of social change. Environ. Plann. A 42 (6), 1273–1285. doi:10.1068/a42282.

Soper, D., 2021. Regression Coefficient Confidence Interval Calculator. Free Statistics Calculators.

Spangenberg, E.R., Kareklas, I., Devezer, B., Sprott, D.E., 2016. A meta-analytic synthesis of the question–behavior effect. J. Consum. Psychol. 26 (3), 441–458. doi:10.1016/j.jcps.2015.12.004.

Spangenberg, E.R., Sprott, D.E., Grohmann, B., Smith, R.J., 2003. Mass-communicated prediction requests: practical application and a cognitive dissonance explanation for self-prophecy. J. Mark. 67 (3), 47–62. doi:10.1509/jmkg.67.3.47.18659.

Spielholz, P., Silverstein, B., Morgan, M., Checkoway, H., Kaufman, J., 2001. Comparison of self-report, video observation and direct measurement methods for upper extremity musculoskeletal disorder physical risk factors. Ergonomics 44 (6), 588–613. doi:10.1080/00140130118050.

St. John, F.A.V., Edwards-Jones, G., Gibbons, J.M., Jones, J.P.G, 2010. Testing novel methods for assessing rule breaking in conservation. Biol. Conserv. 143 (4), 1025–1030. doi:10.1016/j.biocon.2010.01.018.

Streff, F.M., Wagenaar, A.C., 1989. Are there really shortcuts? Estimating seat belt use with self-report measures. Accid. Anal. Prevent. 21 (6), 509–516. doi:10.1016/0001-4575(89)90066-3.

Stutzman, T.M., Green, S.B., 1982. Factors affecting energy consumption: two field tests of the Fishbein-Ajzen model. J. Soc. Psychol. 117 (2), 183–201. doi:10.1080/00224545.1982.9713427.

Transue, J.E., 2007. Identity salience, identity acceptance, and racial policy attitudes: American National Identity as a Uniting Force. Am. J. Pol. Sci. 51 (1), 78–91. doi:10.1111/j.1540-5907.2007.00238.x.

Tyson, A., Funk, C., Kennedy, B., 2022. Americans Largely Favor U.S. Taking Steps to Become Carbon Neutral by 2050. Pew Research Center.

van der Werff, E., Steg, L., 2016. The psychology of participation and interest in smart energy systems: comparing the value-belief-norm theory and the value-identity-personal norm model. Energy Res. Soc. Sci. 22, 107–114. doi:10.1016/j.erss.2016.08.022.

Van der Werff, E., Steg, L., Keizer, K., 2013. The value of environmental self-identity: the relationship between biospheric values, environmental self-identity and environmental preferences, intentions and behaviour. J. Environ. Psychol. 34, 55–63. doi:10.1016/j.jenvp.2012.12.006.

Verbeij, T., Pouwels, J.L., Beyens, I., Valkenburg, P.M., 2021. The accuracy and validity of self-reported social media use measures among adolescents. Comput. Hum. Behav. Rep. 3. doi:10.1016/j.chbr.2021.100090, 100090–100090.

Vesely, S., Klöckner, C.A., 2020. Social desirability in environmental psychology research: three meta-analyses. Front. Psychol. 11. doi:10.3389/fpsyg.2020.01395.

Vilar, R., Milfont, T.L., Sibley, C.G., 2020. The role of social desirability responding in the longitudinal relations between intention and behaviour. J. Environ. Psychol. 70. doi:10.1016/j.jenvp.2020.101457, 101457–101457.

Whitmarsh, L., O'Neill, S, 2010. Green identity, green living? The role of pro-environmental self-identity in determining consistency across diverse pro-environmental behaviours. J. Environ. Psychol. 30 (3), 305–314. doi:10.1016/j.jenvp.2010.01.003.

Wilding, S., Conner, M., Sandberg, T., Prestwich, A., Lawton, R., Wood, C., Miles, E., Godin, G., Sheeran, P., 2016. The question-behaviour effect: a theoretical and methodological review and meta-analysis. Eur. Rev. Soc. Psychol. 27 (1), 196–230. doi:10.1080/10463283.2016.1245940.

Wood, C., Conner, M., Miles, E., Sandberg, T., Taylor, N., Godin, G., Sheeran, P., 2016. The impact of asking intention or self-prediction questions on subsequent behavior: a meta-analysis. Personal. Soc. Psychol. Rev. 20 (3), 245–268. doi:10.1177/1088868315592334.