

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (<i>n</i>) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give P values as exact values whenever suitable.</i>
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's <i>d</i> , Pearson's <i>r</i>), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	We used Qualtrics software (XM) to run our surveys, under a license from the University of Pennsylvania.
Data analysis	We used R and the R studio wrapper to run our analyses. These are open source tools available at https://www.r-project.org/ and https://posit.co/download/rstudio-desktop/ for both MAC and PC users.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Preregistration: Our hypotheses, methods, and analysis plan were pre-registered prior to data collection which spanned June-July 2023 and December 2024. There were no deviations from the preregistration in terms of the experimental design. There was one deviation in terms of data collection and, after receiving advice on early drafts, we simplified the analyses. See SI Table S1 for documentation of these changes. Copies of our pre-registrations are provided in the ResearchBox link

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	We asked participants to (optionally) report their gender, phrased as "With which gender do you most identify?" Options included: Male, Female, Non-binary, Not listed or prefer not to say. We did not use this variable in our analyses nor did we pre-register doing so. All data was collected with consent.
Reporting on race, ethnicity, or other socially relevant groupings	We asked participants to (optionally) report their ethnicity phrased as "With which ethnicity do you most identify?" Options included: American Indian/Alaska Native, Asian, Native Hawaiian / Other Pacific Islander, Black, Hispanic, White / Caucasian, Other (please specify), Prefer not to say. We did not use this variable in our analyses nor did we pre-register doing so. All data was collected with consent.
Population characteristics	We asked participants to (optionally) report their age phrased as "How old are you?" Options were inclusive of 18-99 years old. We did not use this variable in our analyses nor did we pre-register doing so. All data was collected with consent.
Recruitment	We recruited participants using the Prolific academic platform and detail our approach and the language used in the Supplement section 3D. The inclusion criteria we selected mimicked, as best as possible, filters applicable to the target population of each underlying field study. There is undoubtedly selection bias in who completes surveys like these on Prolific academic, however, the purpose of our study is to study exactly this setting: running hypothetical surveys in online environments with these online populations. So we are less concerned with selection undermining our inferences in this case.
Ethics oversight	This research complies with all relevant ethical regulations and received approval from the University of Pennsylvania Institutional Review Board (ID: 853845).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	We ran 20 randomized controlled experiments, each measuring the impact of a "nudge" (treatment group) relative to a control group on a hypothetical binary behavior (did or did not do the behavior). We look at the treatment effects and compare those to the treatment effects in the corresponding real-world field studies that the hypothetical studies imitate.
Research sample	We are interested in examining the generalizability of data collected in hypothetical settings with online participants, to real-world field settings. We used field data from prior field experiments as our benchmark; we collected data from online participants on Prolific, using inclusion criteria that (within reason) mimicked characteristics of the real-world contexts. Prolific is a common platform used by academics for online studies—including those with hypothetical scenarios. We selected five field studies to imitate with hypotheticals, from five distinct domains: sustainability (Myers & Souza, 2020: study 2), health (Polman et al., 2022: study 1), transportation (Kristal & Whillans., 2020: study 1), finance (Beshears et al., 2021; study 1), and consumer choice (Saccardo et al., 2020: study 2b). More details on our sample can be found in Section 3 of the Supplement.
Sampling strategy	More details on our sample can be found in Section 3 of the Supplement. Our power analyses can be found at the start of each Step 0 file, uploaded in ResearchBox; there is one file per hypothetical domain. Prior to the experiments, we conducted power analyses in order to have sufficient power to detect the original effect size of each experiment. To illustrate, for the sustainability study, our power analysis indicated needing $n=223$ per condition to detect the original effect with 80% probability (and an alpha of .05). We rounded up to collect $n=300$ or $N=2400$ total across the two conditions and four hypothetical designs. Where the small size of the field estimate required a larger sample than reasonably available (e.g., thousands per condition), we aimed to collect enough participants to detect six times the field estimate. This multiplier comes from Dellavigna and Linos (2022), who estimated that academic study results are approximately six times the size of practitioner studies of the same nudge technique (see SI section 3c for details).

Data collection	We used Qualtrics (an online survey software) to collect our data on hypothetical studies. All qsf files are available in ResearchBox, and a user-friendly walk through of each survey is in Supplement Section 8.
Timing	Table S5 in the supplement presents the start and stop dates of each experiment. We ran all but one experiment between June 13 and July 21, 2023. We ran one experiment -- the updated sustainability experiment -- on Dec 23, 2024.
Data exclusions	Across all studies, the sample size was N=17,573 (N=17,584 when using the original sustainability study data) before dropping any data. The sample size was N=16,114 (N=16,071 when using the original sustainability study data) after dropping failed attention checks, incompletes, duplicate IPs, and, where relevant, outlier data (per pre-registered specifications). The drops of each of these were pre-registered and can be reproduced using the Step 0 code files (one per hypothetical study) in our ResearchBox.
Non-participation	N=196 participants dropped out of our surveys. (No participant was able to take more than one survey.) These drop outs were either due to failing the attention check (per Prolific policy, an early explicit fail can be redirected to the end of the survey) or closing the survey without completing it.
Randomization	Participants were randomly allocated to treatment groups using the randomizer within Qualtrics. We selected the option to "evenly" balance conditions to ensure we didn't accidentally have imbalanced samples. In the finance study, participants were randomized within blocks of their birthday months, mimicking the field study's randomization strategy.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Plants

Seed stocks	NA
Novel plant genotypes	NA
Authentication	NA