nature human behaviour

Article

https://doi.org/10.1038/s41562-026-02629-8

Fifty-five crowdsourced designs assessing carbon pricing support in the USA

Received: 3 March 2026

Accepted: 30 September 2026

Published online: 25 November 2026

Check for updates

A list of authors and their affiliations appears at the end of the paper

A carbon price is an effective and cost-efficient policy to mitigate emissions, yet low public acceptance and limited political support remain major barriers to its widespread implementation. This crowdsourced 'many-designs' project presents results from 55 behavioural interventions on real-world support for carbon pricing, independently developed by international research teams. By implementing all designs simultaneously with almost 20,000 US residents, this preregistered study shows very small positive but statistically significant effects of behavioural interventions on a set of design-specific proxies of real-world support (Cohen's $d = 0.06$; 99.5% confidence interval (CI) 0.001 to 0.124) and on 4 out of 5 common measures of stated support (Cohen's d values 0.06–0.08; 99.5% CIs from 0.006 to 0.112 and from 0.020 to 0.133; between 1.7 and 2.4 percentage points in increased support across outcomes), including the willingness to endorse a carbon price that internalizes the social costs of US$120 per ton of CO_2 emissions ($d = 0.06$; 99.5% CI 0.012 to 0.118). Furthermore, the results reveal low-to-medium between-study heterogeneity ($\tau = 0.07$–0.12). Lastly, we identify strong overconfidence among research teams regarding the expected effects of their interventions and those of their peers, indicating a miscalibration of community expectations.

The use of behavioural interventions to change individual behaviour advancing towards climate neutrality has been widely advocated and tested[1–7]. However, meta-analyses and large-scale 'mega-studies' evaluating the impact of behavioural interventions in climate-related and other domains have raised concerns about publication bias[8,9], questioned the overall impact of behavioural interventions[10,11] and even documented instances of backfiring effects[8] with unintended consequences. An alternative path to advance towards climate neutrality is to test whether behavioural interventions can foster public support for regulations that in turn lead to broader and more profound behavioural change. Yet, considerably less attention has been paid to whether such interventions can increase support for system-changing policies[12,13]. This might offer a promising approach for tackling social concerns, such as climate change, on which public opinion is divided[14–17]. Minor changes in support for climate policies may be sufficient to change election outcomes or induce tipping points in public perceptions[18,19]. Recent innovative large-scale studies have advanced the field by investigating behavioural interventions designed to increase support for a variety of climate policies[20,21]. The present project focuses on carbon pricing, which is widely regarded by scientists as an effective[22,23] and cost-efficient policy instrument for reducing carbon emissions[24,25]. For example, the European Union (EU) Emission Trading System reduced 3.8% of total EU-wide emissions compared with a world without carbon pricing, despite the low trading prices for CO_2 permits[26]. Moreover, a worldwide analysis of climate policies showed that policy mixes including price-based instruments were particularly effective in reducing emissions[23]. However, despite carbon pricing being actively discussed by various governments, only 28% of global carbon emissions are covered, with an average carbon price of US$19 per tCO_2e in 2025[27], much lower than (conservative) estimates for a socially efficient carbon price (US$120 per tCO_2e (ref. 28)). Ongoing discussions revolve around the appropriate price level, given large uncertainties surrounding the social cost of carbon[29] and its implications for global justice[30]. Overall, political support appears to be insufficient for more widespread implementation[21].

e-mail: juergen.huber@uibk.ac.at

This study presents the results of a 'many-designs' crowdscience study on carbon pricing support (see, for example, Huber et al.[31] for the methodology). This 'many-designs' study contributes to the state of knowledge in several ways: first, it leverages the expertise of 55 independent research teams (RTs), randomly selected from 135 international applicants, to design individual randomized controlled trials. Thus, this 'many-designs' approach differs from the classical standard of 'one research team, one design, one dataset' by benefiting from multiple design ideas of, among others, environmental economists, psychologists and behavioural scientists. Second, it compiles evidence from 55 independent studies. Unlike in other projects aiming to foster support for climate policies[20,21,32], each of the 55 designs presented herein includes both a treatment and a control condition and an observed proxy measure for real-world policy support as the outcome. Third, it speeds up the scientific knowledge generation process, with independent studies on the same research question being developed simultaneously rather than sequentially. Fourth, it ensures consistency by using a common experimental participant pool and a unified econometric analysis, as opposed to traditional meta-analyses. Therefore, this 'many-designs' setup excludes population and analytical heterogeneity in evaluating the different interventions[33–37]. Moreover, it transparently reports the results of all 55 designs regardless of their findings, mitigating publication bias. It also does not exclude proposals ex ante on the basis of the subjective judgement of the principal investigators. We preregistered an extensive pre-analysis plan on OSF that outlines the entire structure, procedure and main part of the data analysis. This procedure prevents scientific malpractices such as *P*-hacking, HARKing and selective reporting. Finally, this project innovates by extending policy support outcomes to real-life measures of policy support proposed by the RTs. Previous literature commonly drew on intentions and self-reports to measure policy support[6,32], whereas real-life measures have only been used to capture pro-environmental individual behaviours[14,20,21].

Each RT's main task was to independently design a study to answer the general research question: "What is the impact of behavioural interventions on the real-world support for a price of carbon?". This required the RTs to develop an outcome measure with real-world implications for the support of carbon pricing. Some examples from the set of studies are donating to non-governmental organizations (NGO) that lobby for carbon pricing, signing petitions requesting the implementation of carbon pricing, writing letters to government officials asking to advance towards carbon pricing or publishing advocacy-oriented posts for carbon pricing on social media. We allowed RTs to choose the outcome measure to take advantage of the wisdom of the crowd in identifying suitable proxy variables for real-world support for carbon pricing, particularly because the literature has not yet converged on a standard measurement practice. Detailed descriptions of each study, along with an overview of intervention categories and outcome variables, are provided in Supplementary Table 1 and Supplementary Fig. 21; full details are available via the corresponding preregistrations in our OSF repository. Each RT developed the software for their online study and assessed its own expected effect size in Cohen's *d* units (self-assessment: "How large do you expect to be the effect size of the intervention that you propose?"; see the Methods for details). Furthermore, RTs were randomly assigned to assess ten other RTs' designs (peer assessment: "How large do you expect the effect size [measured as Cohen's d] of this experimental design to be?", "To what extent does the design provide an informative test [0 = not informative at all; 10 = very informative] of the research question" and "Under which type of behavioural intervention would you classify this proposal? Make sure you only state one type, and please provide a short reasoning for your classification [1–2 sentences]."; see the Methods for details). See Supplementary Figs. 19 and 20 for an overview of the participating RTs' geographical regions and research fields. From January to March 2025, we simultaneously ran all 55 studies on a US general population Prolific sample[38] with 19,558 participants (that is, approximately 356 participants per design—thus, 178 per condition—based on a preregistered power simulation for the meta-effect; Supplementary Figs. 22 and 23). Each participant was randomly assigned to either the TREATMENT or CONTROL condition of one of the 55 studies (that is, a random assignment to 1 of 110 (55 × 2) experimental conditions; see Supplementary Section 12 and Supplementary Table 17 for descriptive statistics on covariate balancing). We selected the USA because it is one of the world's largest emitters, lacks a federal carbon tax and exhibits highly heterogeneous attitudes towards such policies[39,40].

In each study, participants' support for carbon pricing was first measured through the real-world support metric for a carbon price proposed by the RTs. Next, participants completed a follow-up questionnaire ('policy survey') based on a subset of questions in the studies by Vlasceanu et al.[20] and Dechezleprêtre et al.[21], including five measures of stated support for carbon pricing, which were constant across all studies. Altogether, we used (1) the observed real-world policy support variable of each RT ('general support (observed)'), (2) a general stated support measure for carbon taxation ('general support (stated)'), (3) a stated support measure for a price increase of a carbon tax ('general price increase'), (4) a stated support measure for a carbon tax equal to the social cost of carbon of US$120 ('social price') based on the lower-bound estimates from the US Environmental Protection Agency (EPA)[28], (5) a stated support measure for a carbon tax with carbon dividends ('tax with rebates') and (6) a stated support measure for a carbon tax equal to US$120 with carbon dividends ('social price with rebates'; see Supplementary Section 15 for the full policy survey). This set of six variables constitutes the primary outcomes for hypotheses 1 and 2. Variables 2, 4, 5 and 6 were measured on five-point Likert scales (−2: strongly oppose, 2: strongly support). Variable 3 was measured on a continuous scale of 0–100.

We formulated ten preregistered hypotheses (see Supplementary Section 6 for the exact formulation). Hypothesis 1 states that behavioural interventions increase support for a price on carbon, as measured by the primary outcome variables. Hypothesis 2 claims that there is heterogeneity in the estimated effect sizes for the primary outcome variables across RTs. Hypothesis 3 concerns the impact of behavioural interventions on participants' perceptions about carbon pricing (that is, their fairness and effectiveness) and about climate change (that is, adoption of climate-friendly behaviour, worries about impacts, technical feasibility of net zero and the potential harm on one's own life), which we preregistered as 'secondary' outcome variables. Hypotheses 4–9 revolve around the associations between the observed effects on real-world support and the answers to the three questions from the peer assessment as potential moderators: the informativeness of the intervention (hypothesis 4), the expected effect size (hypothesis 6), the corresponding intervention category (hypothesis 8) and the residual heterogeneity after controlling for each of the three moderators (hypotheses 5, 7 and 9). Finally, hypothesis 10 states that RTs' expectations of their own effect sizes do not significantly differ from the observed effect sizes.

Results

Figure 1 provides a descriptive overview of the final design submissions. Figure 1a shows the behavioural intervention categories (Supplementary Section 10), together with subcategories (see Supplementary Section 11). Most RTs implemented information-based interventions (54.5%) and the most common within this category is explaining revenue recycling or distributional fairness, followed by interventions emphasizing policy effectiveness in terms of reductions in greenhouse gas emissions. A second large group used behavioural-biases or reflective-thinking interventions (36.4%), most often comparative imagining or self-generated persuasion tasks, such as prompts to consider future generations, compare future policy scenarios or generate arguments in favour of carbon pricing. Less frequent was the use of

https://doi.org/10.1038/s41562-026-02629-8

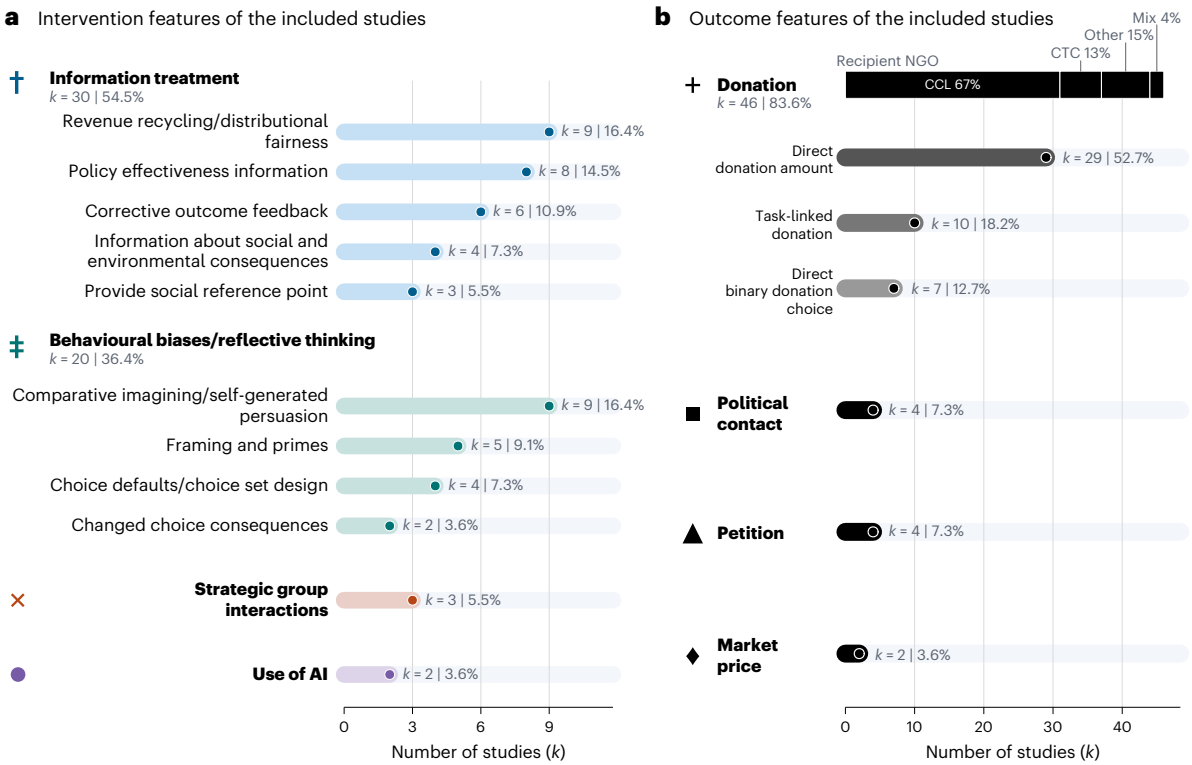

Fig. 1 | Design and outcome features of the studies. a, Intervention-design features across the 55 studies, grouped into four AI-classified main categories (Supplementary Section 10) and exploratory post hoc subcategories (Supplementary Section 11). 'Information treatment' includes informational interventions on revenue recycling or distributional fairness, policy effectiveness, corrective outcome feedback, social and environmental consequences, and social reference points. 'Behavioural biases/reflective thinking' includes comparative imagining or self-generated persuasion, framing and primes, choice defaults or choice-set design, and changed choice consequences. 'Strategic group interactions' covers strategic or social-dilemma settings, and 'Use of AI' covers AI-based tools. **b**, Real-world outcome measures used to capture support for carbon pricing. 'Donation' includes direct donation amounts, direct binary donation choices and task-linked donations. 'Political contact' depicts communication with political representatives or related policy actors; 'Petition' captures petition signing or willingness to sign; and 'Market price' outlines outcomes tied to real market-based carbon-pricing mechanisms, such as purchasing or removing CO_2 certificates or tradable permits. RT 'YQH52' used a composite outcome including both a donation and a petition signature; therefore, outcome-category counts sum to more than 55, while percentages use the 55 included studies as the denominator.

strategic group interactions (5.5%) or artificial intelligence (AI)-based interventions (3.6%). Figure 1b shows that most RTs used incentivized donations as proxies for real-world policy support (83.6%). The most common donation format was a direct endowment allocation, in which participants decided how much of a windfall payment of up to US$4 to keep or donate (52.7% of all RTs). Other donation outcomes were task linked, for example, to quizzes, real-effort tasks or other incentivized decisions (18.2%), or used binary take-or-donate choices (12.7%). Fewer RTs opted for active communication with political representatives (7.3%), petition signing with public or governmental communication (7.3%) or market-based decisions to purchase and remove CO_2 certificates (3.6%) (see Supplementary Table 1 for more details on the designs).

For all analyses, we preregistered a significance threshold of $\alpha = 0.5\%$[41]. Consequently, we refer to findings with a P value between 0.5% and 5% as suggestive evidence. Unless stated otherwise, all analysis specifications in this paper were preregistered, including the set of covariates—covering sociodemographic variables, private behaviour and trust in government—added to the multivariate ordinary least squares (OLS) regressions (see the Methods for details).

Figure 2a addresses hypotheses 1 and 2. It shows meta-analytical effect sizes (Cohen's d) and heterogeneity estimates (Cochran's Q, I^2 and τ) based on random-effects meta-analyses (DerSimonian–Laird model) for all six primary outcome variables 1–6. The effect sizes of the individual studies are obtained via univariate OLS regressions with an intervention treatment dummy variable (green diamond) or a multivariate model additionally including control variables (orange square). We find positive yet very small statistically significant meta-effects of behavioural interventions across most of the measures (Cohen's $d = 0.04$–0.08; Supplementary Table 4). Specifically, results show statistically significant positive meta-effects on real-world support ('general support (observed)': $d = 0.063$, 99.5% confidence interval (CI) 0.001 to 0.124, $z = 2.867$, $k = 55$, $P = 0.002$), stated general support for carbon pricing ('general support (stated)': $d = 0.076$, 99.5% CI 0.020 to 0.133, $z = 3.793$, $k = 55$, $P < 0.001$), stated support for an increase in the price of carbon ('general price increase': $d = 0.063$, 99.5% CI 0.007 to 0.119, $z = 3.156$, $k = 55$, $P < 0.001$), stated support for a carbon price internalizing the social cost of carbon at US$120 per ton of CO_2 emissions ('social price': $d = 0.065$, 99.5% CI 0.012 to 0.118, $z = 3.422$, $k = 55$, $P < 0.001$), and stated support for a carbon price with carbon dividends ('tax with rebates': $d = 0.059$, 99.5% CI 0.006 to 0.112, $z = 3.118$, $k = 54$, $P < 0.001$). Suggestive evidence also points to a positive effect on the support for a carbon price at the social cost of carbon with carbon dividends ('social price with rebates': $d = 0.038$, 99.5% CI −0.015 to 0.091, $z = 2.022$, $k = 55$, $P = 0.022$). Controlling for the preregistered covariates yields effect sizes ranging from $d = 0.04$ to $d = 0.09$, with qualitatively equivalent results. Placing these findings in perspective, exploratory post hoc OLS regressions indicate that the treatment conditions increased average support for carbon pricing by 1.1–2.4 percentage points relative to the control group, depending on the outcome measured (general support (observed): $\beta = 2.296$, 99.5% CI 0.726 to 3.867, $P < 0.001$; general support (stated): $\beta = 2.434$, 99.5%

https://doi.org/10.1038/s41562-026-02629-8

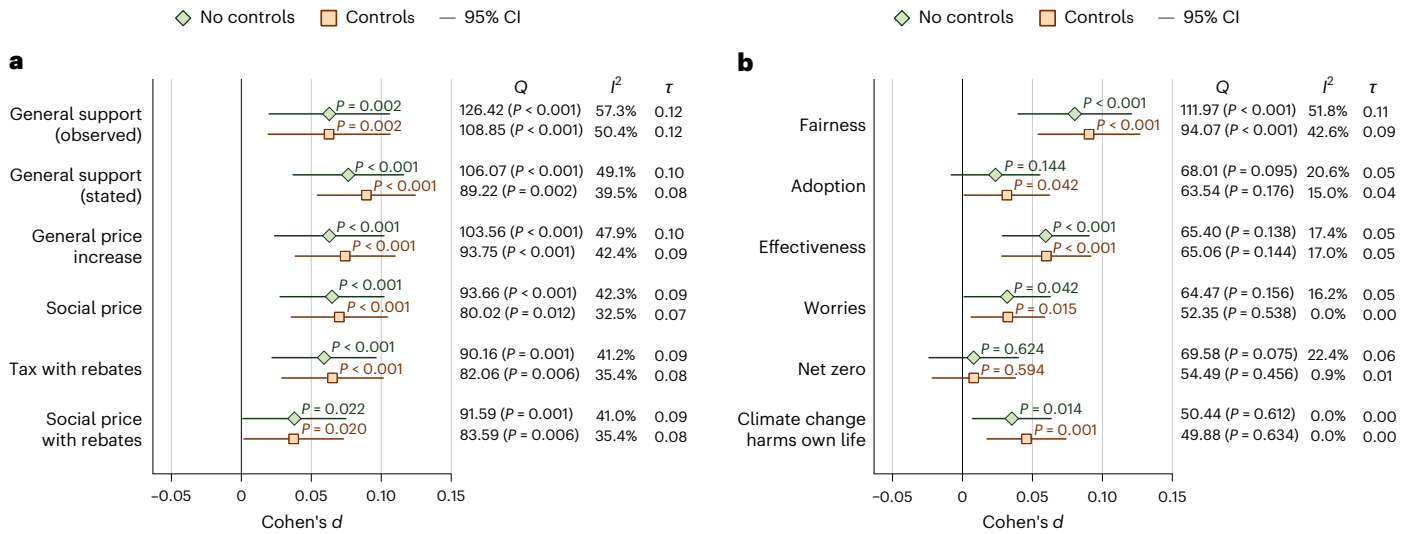

Fig. 2 | Meta-analytical results for primary and secondary outcomes. **a**, Meta-analytical results across primary outcomes for hypotheses 1 and 2. Meta-analytical point estimates and the 95% CIs of random-effects meta-analyses based on effect sizes obtained from univariate (green diamonds) and multivariate (orange rectangles; see Supplementary Section 9 for the set of covariates) OLS regressions for each of the 55 studies (hypothesis 1). 'General support (observed)' denotes the outcome variables suggested by the RTs, reflecting real-world support for carbon pricing, while the other variables were part of a common survey at the end of the experiment measuring stated support for a price on carbon (Supplementary Section 15). 'General support (stated)' measures an individual's general support for a carbon price. 'Social price' denotes support for a carbon price equal to the social price of carbon of US$120 tCO_2^{-1}. 'Tax with rebates' measures support for a general carbon tax with cash transfers as redistribution to households. 'Social price with rebates' denotes an individual's support for a carbon tax equal to the social price of carbon of US$120 tCO_2^{-1} with cash transfers as redistribution to households. All variables were measured on five-point Likert scales (−2: strongly oppose, 2: strongly support), except the variable 'general price increase', which was measured on a range slider from 0 (not at all) to 100 (very much). *P* values associated to meta-effects are based on one-sided *z*-tests (owing to the directionality of hypothesis 1). Cochran's *Q*, I^2 and τ indicate effect size heterogeneity (hypothesis 2). Detailed metrics are shown in Supplementary Table 4. **b**, Meta-analytical results across secondary outcomes for hypothesis 3. Meta-analytical point estimates and the 95% CIs of random-effects meta-analyses based on effect sizes obtained from univariate (green diamonds) and multivariate (orange rectangles; see Supplementary Section 9 for the set of covariates) OLS regressions for each of the 55 studies. All variables were part of a common survey battery at the end of each study measuring beliefs and perceptions about carbon pricing and climate change, as well as individuals' willingness to adopt climate-friendly behaviours (Supplementary Section 15). 'Fairness' measures an individual's agreement on whether a price on carbon is fair. 'Adoption' is an index measuring an individual's willingness to adopt climate-friendly behaviours, including the willingness to limit air travel, driving, beef consumption, heating or cooling and to get an e-vehicle. 'Effectiveness' is an index of an individual's belief in the effectiveness of a carbon tax in addressing climate change and its economic impacts. This comprises the beliefs on whether such a tax would reduce driving, air pollution and the use of fossil fuels, encourage building insulation and whether such a tax would have a negative effect on the US economy and its employment (reverse coded). 'Worries' is an index of an individual's perceived likelihood of climate change leading to the following events: droughts and heatwaves, volcanic eruptions, rising sea levels, declines in agricultural production, reduced standards of living, increased migration flows, more armed conflict and the extinction of humankind. 'Net zero' is an individual's belief about the technical feasibility to stop greenhouse gas emissions by the end of the century while maintaining satisfactory standards of living in the USA. 'Climate change harms own life' measures an individual's belief about whether climate change is already affecting, or will soon negatively affect, their personal life. *P* values associated to meta-effects are based on two-sided *z*-tests (owing to the non-directionality of hypothesis 3). Cochran's *Q*, I^2 and τ indicate effect size heterogeneity. See Supplementary Table 5 for detailed metrics.

CI 1.124 to 3.744, $P < 0.001$; general price increase: $\beta = 2.001$, 99.5% CI 0.687 to 3.314, $P < 0.001$; social price: $\beta = 2.151$, 99.5% CI 0.795 to 3.507, $P < 0.001$; tax with rebates: $\beta = 1.746$, 99.5% CI 0.441 to 3.052, $P < 0.001$; social price with rebates: $\beta = 1.107$, 99.5% CI −0.162 to 2.376, $P = 0.014$; Supplementary Figs. 2 and 3 and Supplementary Table 18).

Figure 2a reports various estimates of effect size heterogeneity relevant to hypothesis 2. Cochran's *Q* is a test statistic indicating potential heterogeneity in effect sizes. Based on this metric, we find statistically significant evidence of between-study heterogeneity for all primary outcome variables 1–6. The preregistered quantification measure of heterogeneity we use is τ, which is the estimated standard deviation (s.d.) of true effect sizes and represents the variation in effect sizes across RT designs other than sampling variation. The values of τ range from 0.074 for the general stated support of a social carbon price with control variables (99.5% CI 0.000 to 0.138) to 0.123 for the observed real-world support measure used by the RTs without control variables (99.5% CI 0.071 to 0.193). We find I^2 values between 33% and 57%. As I^2 is the share of true between-study heterogeneity in total variance, this implies that roughly one-third to half of the observed variation reflects true heterogeneity rather than within-study sampling variance. This can be classified as small-to-medium between-study heterogeneity[42,43]. All other preregistered heterogeneity estimates are presented in Supplementary Table 4.

Figure 2b provides the results for hypothesis 3 on the impact of behavioural interventions on secondary outcome variables measured via the policy survey. We calculate meta-analytical effect sizes on the basis of the same econometric and visual approach as for hypotheses 1 and 2. Owing to the non-directionality of hypothesis 3, we report results of two-sided *z*-tests (see Supplementary Fig. 4 for detailed results on each of the index items). We find statistically significant positive effects of behavioural interventions on perceptions about fairness of carbon pricing ($d = 0.08$, 99.5% CI 0.022 to 0.138, $z = 3.876$, $k = 55$, $P < 0.001$) and its effectiveness ($d = 0.06$, 99.5% CI 0.015 to 0.104, $z = 3.760$, $k = 55$, $P < 0.001$). In addition, we report suggestive positive evidence for the effectiveness of interventions on worries about climate change ($d = 0.03$, 99.5% CI −0.012 to 0.076, $z = 2.035$, $k = 55$, $P = 0.042$) and the belief in climate change harming one's own life ($d = 0.04$, 99.5% CI −0.005 to 0.076, $z = 2.458$, $k = 55$, $P = 0.014$). However, we do not find any statistical evidence for effects on the self-reported willingness to adopt individual climate-friendly behaviour ($d = 0.02$, 99.5% CI −0.022 to 0.069, $z = 1.462$, $k = 55$, $P = 0.144$) and for the belief in net zero being technically feasible ($d = 0.01$, 99.5% CI −0.038 to 0.054, $z = 0.490$,

$k = 55$, $P = 0.624$). Controlling for preregistered covariates yields effect sizes ranging from $d = 0.01$ to $d = 0.09$, with effects on willingness to adopt climate-friendly behaviour becoming positive suggestive evidence, and effects on perceived harms of climate change to one's own life becoming statistically significant. Moreover, we only find statistically significant evidence of between-study heterogeneity in fairness perceptions. The estimated τ ranges from 0.00 to 0.11 and I^2 from 0% to 52%.

Next, we unpack the analysis and report estimated effects for each individual study. Figure 3 (left) shows a forest plot of effect sizes across all 55 interventions on the RTs' chosen real-world support measure (circles with 95% CIs), each RT's reported beliefs about the expected effect of their design ('self-assessment', blue triangles) and average reported beliefs by ten other RTs ('peer assessment', red rectangles with 95% CIs). We rank the studies from the lowest (top) to the largest effect sizes (bottom). One-sided t-tests show that most (44) studies do not result in any statistically significant increase in the support proxy variables nor exhibit suggestive evidence for an increase at a P value of 5%. Consequently, the rest of the studies (11) fall below the 5% threshold; five studies show a statistically significantly positive effect ($P < 0.005$) and six exhibit a suggestive positive effect ($0.005 < P < 0.05$). In addition, we ran exploratory two-sided t-tests to capture potential backfiring effects (not shown in Fig. 3; see Supplementary Table 2 for details). These analyses reveal three studies exhibiting statistically significant positive effects (all $P < 0.001$), five studies with suggestively positive effects ($0.005 < P < 0.05$), one study with a negative suggestive effect ($P = 0.028$) and one study entailing a negative statistically significant effect ($P < 0.001$). The predictive distribution of true effect sizes is shown below the meta-analytic effect, incorporating both within- and between-study heterogeneity. The predictive distribution extends across both the positive and negative domains, indicating that a newly implemented intervention from the same design space could plausibly produce effects in either direction, including potential backfiring effects. Specifically, while 69.2% of newly conducted studies are expected to yield positive effects (that is, Cohen's $d > 0$), 30.8% of newly conducted studies from the same design space are expected to yield negative or null effects (that is, Cohen's $d \leq 0$). Thus, the 95% prediction interval (PI), which is more commonly reported in meta-analyses, spans negative and positive values, indicated by the left and right borders of the white area (95% PI −0.182 to 0.308).

Concurrently, expectations about the effect sizes for one's own study, as well as the expectations by other RTs, are overly optimistic. Each RT expected that their intervention, as well as their peers' interventions, would have a positive effect. Their expectations significantly exceed the materialized effects of behavioural interventions (effect size in Mann–Whitney Utests observed versus self-assessed belief: $r = 0.668$, 99.5% CI 0.460 to 0.800, $P < 0.005$; observed versus peer-assessed beliefs: $r = 0.672$, 99.5% CI 0.440 to 0.820, $P < 0.005$; Supplementary Fig. 6), contradicting hypothesis 10.

In addition, Fig. 3 (right) shows an equivalent forest plot of effect sizes across designs on the stated general support measure ("Do you support or oppose a carbon tax in general?"). Again, we rank the studies from the smallest (top) to the largest effect sizes (bottom; see Supplementary Table 3 for details). Similarly, the predictive distribution spans both negative and positive values, although it is more skewed towards positive effects: 76% of newly conducted studies from the same design space are expected to produce positive effects and 24% are expected to exhibit negative or null effects. The 95% PI confirms this (95% PI −0.133 to 0.285).

We also connect RT's effects across measures with grey lines. Perfect correlation in effect sizes across measures would yield horizontal lines, which is in contrast to what we observe (Pearson's correlation coefficient between real-world and stated support: $r = 0.175$, 99.5% CI −0.209 to 0.512, $P = 0.202$). For example, the study with the largest observed real-world effect ('WFQ37') shows one of the lowest stated effects, while some studies with larger stated effects entail smaller and non-significant real-world effects (for example, 'SDB70' or 'RVM76'). Nevertheless, some studies show consistent positive effects across both outcome measures (for example, 'EUZ78', 'EEO59' and 'ATY04'). Supplementary Fig. 1 shows the comparisons of individual and meta-analytical effect sizes for all primary outcomes 1–6.

We next examine which study characteristics explain variations in effect sizes as measured by observed real-world support. Effects from meta-regressions in Fig. 4a show suggestive evidence that interventions targeting behavioural biases and reflective thinking are associated with positive effects ($d = 0.092$, 99.5% CI −0.012 to 0.196, $z = 2.488$, $k = 20$; $P = 0.013$). Cross-comparisons of the coefficients do not indicate any evidence of differences in observed effect sizes between intervention categories (all $P > 0.05$; Supplementary Table 6). Additional exploratory analyses in Supplementary Table 15 and Supplementary Fig. 17 using the subcategories of Fig. 1 show suggestive evidence that information interventions providing information on revenue recycling and distributional fairness are associated with positive effects ($d = 0.098$, 99.5% CI −0.040 to 0.236, $z = 1.992$, $k = 9$; $P = 0.046$). The cross-comparison of coefficients shows suggestive evidence that stressing revenue recycling/distributional fairness ($\beta = -0.207$, 99.5% CI −0.425 to 0.011, $z = -2.662$, $P = 0.008$) or highlighting the policy effectiveness in reducing CO_2 emissions ($\beta = 0.200$, 99.5% CI −0.024 to 0.423, $z = 2.510$, $P = 0.012$) are both associated with larger effects than relying on corrective outcome feedback. No evident differences for all other comparisons are found (all $P > 0.05$; Supplementary Table 15). Within the intervention category of behavioural bias and reflective thinking, there is suggestive evidence of interventions changing choice consequences being associated with positive effects ($d = 0.337$, 99.5% CI −0.001 to 0.674, $z = 2.798$, $k = 2$, $P = 0.005$) and larger effect sizes than for framing/primes ($\beta = -0.327$, 99.5% CI −0.726 to 0.073, $z = -2.296$, $P = 0.022$) and for choice defaults/choice-set designs ($\beta = -0.299$, 99.5% CI −0.713 to 0.115, $z = -2.027$, $P = 0.043$). All other comparisons show no evidence of differences (all $P > 0.05$). Moving to the relevance of the type of outcome variables, Fig. 4b shows suggestive evidence of positive effects being associated with designs using donations as outcome variables ($d = 0.052$, 99.5% CI −0.013 to 0.118, $z = 2.236$, $k = 45$, $P = 0.025$) and statistically significant evidence of positive effects being associated with market prices as outcome variables ($d = 0.351$, 99.5% CI 0.044 to 0.658, $z = 3.210$, $k = 2$, $P = 0.001$). Cross-comparisons of coefficients provide suggestive evidence that market-price outcomes are associated with larger effects than donation outcomes ($\beta = 0.299$, 99.5% CI −0.015 to 0.612, $z = 2.671$, $P = 0.008$) and petitions ($\beta = 0.358$, 99.5% CI −0.041 to 0.756, $z = 2.521$, $P = 0.012$). All other comparisons lead to no evidence for differences (all $P > 0.05$). Further unravelling heterogeneity in donation outcomes (Supplementary Table 15 and Supplementary Fig. 17), we report suggestive evidence of positive effects being associated with direct donations ($d = 0.063$, 99.5% CI −0.016 to 0.142, $z = 2.229$, $k = 29$, $P = 0.026$). Corresponding cross-comparisons within donation types show no statistical evidence for differences (all $P > 0.05$). Collectively, there is no evidence that the set of preregistered moderators, including intervention categories (Supplementary Table 6a), peer-assessed intervention informativeness (Supplementary Table 6c) and peer-assessed expected effect sizes (Supplementary Table 6d), explain observed treatment effects (Supplementary Table 6a: $Q_M(3) = 1.35$, $P = 0.718$; Supplementary Table 6c: $Q_M(1) = 0.30$, $P = 0.586$; Supplementary Table 6d: $Q_M(1) = 0.80$, $P = 0.373$; hypothesis 4, 6 and 8; see Supplementary Table 6 for full outputs). For the outcome categories as exploratory moderators, we report suggestive evidence in Supplementary Table 6b that differences in outcome features explain variation in effect sizes (Supplementary Table 6b: $Q_M(3) = 8.08$, $P = 0.044$). Nevertheless, substantial residual heterogeneity remains across all moderation analyses ($\tau = 0.115$–0.126 with all P values < 0.005; hypotheses 5, 7 and 9).

https://doi.org/10.1038/s41562-026-02629-8

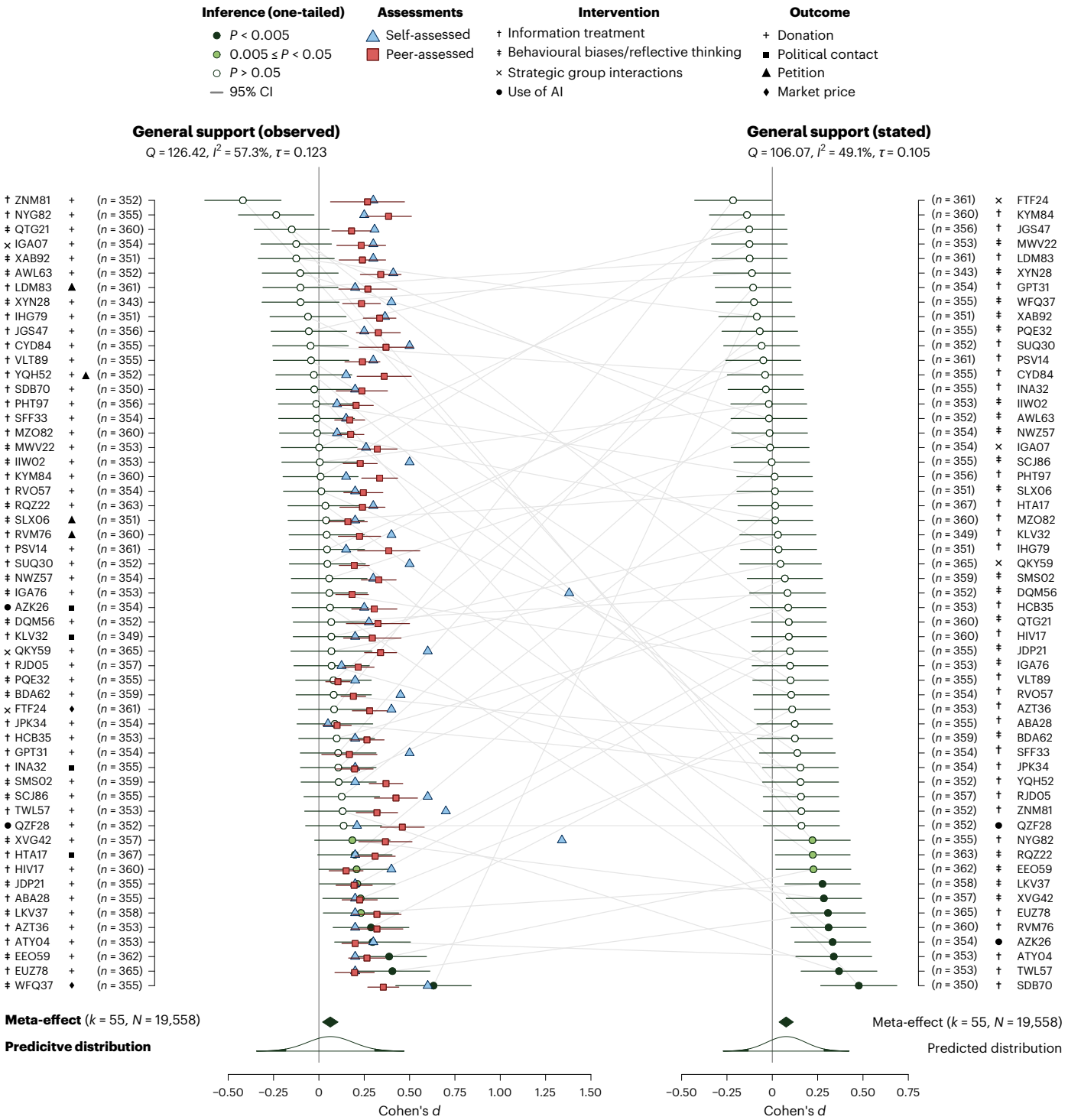

Fig. 3 | Forest plots of the meta-analytic results for real-world and general stated support for carbon pricing. The figure displays point estimates and 95% CIs for the treatment effects on RTs' real-world measure ('general support (observed)', left) and the stated general support measure common to all teams ('general support (stated)', right) across all 55 designs. Prefixes and suffixes attached to the RT IDs depict the type of behavioural intervention and the outcome variable used to measure real-world support for carbon pricing, respectively. RT 'YQH52' used an index combining a donation decision and a petition signature, and therefore displays two outcome-variable indicators. The colour shading of the circles indicates statistical significance (dark-green), suggestive evidence (light-green) and non-significance (white), and are based on one-sided *t*-tests (owing to the directionality of hypothesis 1). Dark-green diamonds represent the overall meta-analytic effects with the width of the diamond representing 95% CIs. Furthermore, the plot shows the predictive distribution, also indicating the 95% PIs by the white area under the curve. These metrics incorporate both the meta-effect's standard error and the between-study heterogeneity τ (all heterogeneity measures are shown at the top). For the 'observed' panel, blue triangles additionally indic ate each RT's belief about its own effect size and red rectangles (with 95% CIs) depict the average peer-belief estimates for the given study. Grey lines connect each RT's estimates across the two outcome variables. The project website (https://manydesignscarbon.online) includes a comprehensive and interactive version of this figure under the 'results' tab.

https://doi.org/10.1038/s41562-026-02629-8

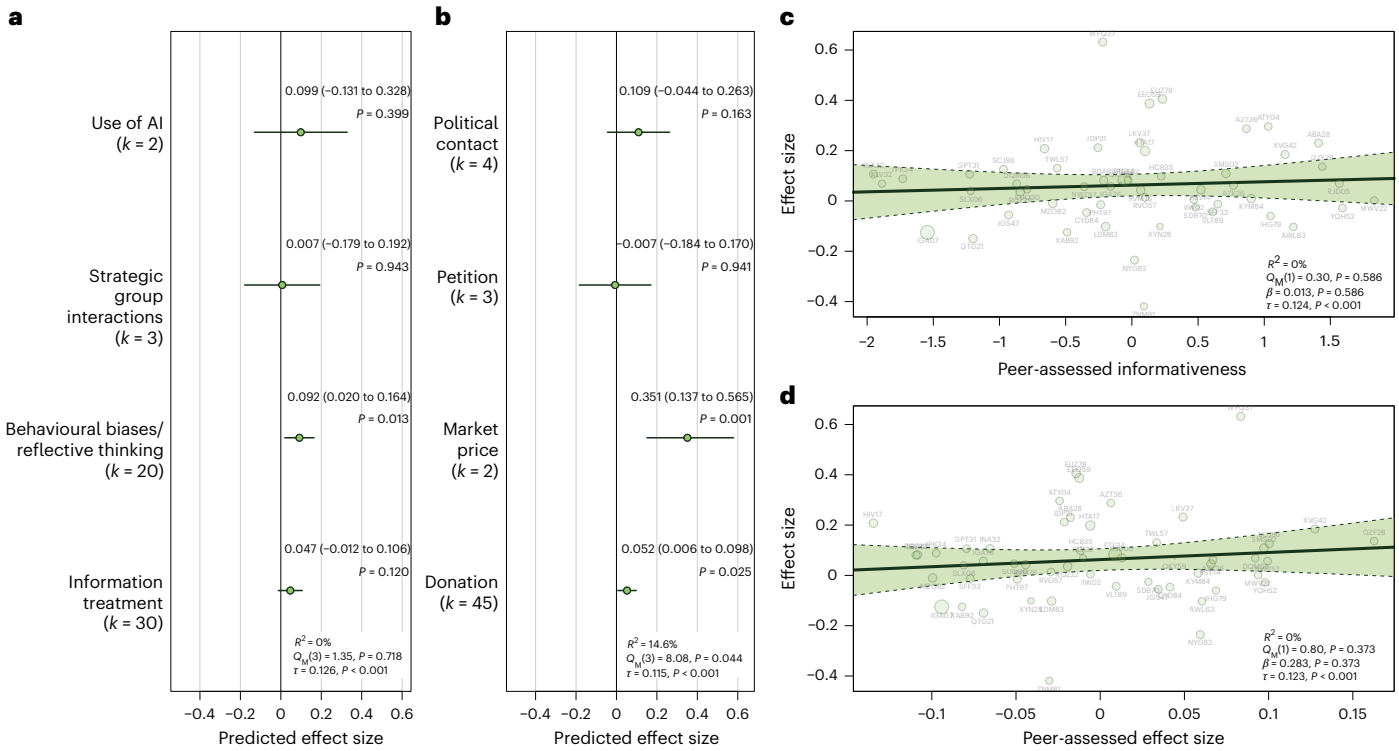

Fig. 4 | Results of meta-regression with potential moderators explaining observed effect sizes. a, Relationship between observed effect sizes and behavioural intervention categories (peer-assessed and AI-classified; hypotheses 8 and 9). Plotted are the predicted effect sizes with 95% CIs for each of the four categories, that is, 'Use of AI', 'Strategic group interactions', 'Behavioural biases/ reflective thinking' and 'Information treatment'. The estimates are based on a meta-regression using the uncontrolled OLS estimates of general observed policy support. *P* values shown are based on two-sided tests on whether the predicted meta-effect of the category is equal to zero. **b**, Relationship between observed effect sizes and RTs' suggested primary outcome variable (exploratory). Plotted are the predicted values and their 95% CIs of effect sizes for each of the four categories, that is, 'donation', 'market price', 'petition' and 'political contact'. The estimates are based on a meta-regression using the uncontrolled OLS estimates of general observed policy support. *P* values shown are based on two-sided tests on whether the predicted meta-effect of the category is equal to zero, respectively. RT 'YQH52' was excluded from this analysis, as the team proposed an index consisting of a donation and a petition signature. **c**, Relationship between observed effect sizes and peer-assessed informativeness (hypotheses 4 and 5). Plotted are the 55 estimated univariate effect sizes on real-world support (Cohen's *d*) over the average peer-assessed informativeness rating (demeaned) of the experimental designs. The green shaded area indicates 95% CIs of the regression line. Bubbles indicate the effects of the individual studies, with the size being inversely related to the standard errors of the effects. **d**, Relationship between observed effect sizes and peer-assessed effect size (hypotheses 6 and 7). Plotted are the 55 estimated univariate effect sizes (Cohen's *d*) over the average peer-assessed effect (demeaned) of the experimental designs. The green shaded area shows 95% CIs of the regression line. Bubbles indicate the effects of the individual studies, with the size being inversely related to the standard errors of the effects. Q_M and its associated *P* value indicate omnibus tests of whether the set of moderators jointly explains heterogeneity in effect sizes. τ indicates effect size heterogeneity and its associated *P* value is based on Cochran's *Q* tests for (residual) heterogeneity.

Complementary analyses for general stated support are reported in Supplementary Table 7 and Supplementary Fig. 5. These unveil a similar pattern whereby some individual design elements are statistically significantly associated with positive effects. However, for the cross-comparisons between elements, in most analyses, there is no statistical evidence for differences, and the aggregate effect of the preregistered moderator analyses is not explaining the variation in effect sizes. The same holds for associations within categories (Supplementary Table 16 and Supplementary Fig. 18): Individual associations between intervention attributes and effect sizes are not robust across the real-life carbon-pricing support measures and the general stated support measure. For example, Fig. 4a shows suggestive evidence of a positive effect for behavioural biases/reflective thinking when measured with real-life support outcomes. However, Supplementary Fig. 5a shows that for the general stated support outcome, the element that is significantly associated with positive effect sizes is information provision.

In addition, after data collection, we conducted an exploratory survey in October 2025 among RTs to gather further information on the study designs and researcher attributes. A total of 48 of the 55 RTs completed the survey (see Supplementary Section 3 for details of the questionnaire). First, using data clustering (Supplementary Section 3), we identified two distinct groups of RTs on the basis of observable characteristics elicited with the questionnaire (Supplementary Fig. 15). The first group/cluster is defined by the prevalence of present-focused interventions, interventions not yet established in previous literature, information treatments, belief and knowledge updating as the main mechanisms, and economists as researchers. Cluster 2 is characterized by interventions targeting behavioural biases and reflective thinking. Exploratory meta-regressions reveal no statistically significant associations between cluster assignment and observed effect sizes regarding observed general support ($\beta = 0.094$, 99.5% CI −0.047 to 0.234, $z = 1.865$, $P = 0.062$; Supplementary Table 13) and stated general support ($\beta = -0.068$, 99.5% CI −0.193 to 0.057, $z = -1.522$, $P = 0.128$; Supplementary Table 14).

Second, using exploratory factor analysis for mixed data, we identified two factors (Supplementary Fig. 16). Factor one mainly captures motivation-based and future-focused interventions, while factor two measures mostly past-focused interventions and designs contributed by researchers from organizational behaviour, finance and marketing. Applying exploratory meta-regressions again, no statistically significant associations between both factor scores and effects with respect

to the real world were found (score factor 1: $\beta = 0.023$, 99.5% CI −0.011 to 0.056, $z = 1.913$, $P = 0.056$; score factor 2: $\beta = 0.006$, 99.5% CI −0.032 to 0.045, $z = 0.458$, $P = 0.647$; Supplementary Table 13) and stated support (score factor 1: $\beta = -0.014$, 99.5% CI −0.044 to 0.015, $z = -1.362$, $P = 0.173$; score factor 2: $\beta = 0.006$, 99.5% CI −0.027 to 0.040, $z = 0.524$, $P = 0.600$; Supplementary Table 14).

Finally, we conducted several robustness checks (see Supplementary Section 2 for details). First, we ran a multiverse analysis with 32 preregistered analytical specifications on the studies' individual effect sizes of observed and stated (exploratory) carbon pricing support (Supplementary Table 8). Across all specifications, results regarding meta-effects on observed real-world and stated support and the associated between-study heterogeneity remain robust and qualitatively identical to the ones reported in the main analysis (median z-statistic observed: 2.637; stated: 4.254; Supplementary Figs. 7 and 9). In addition, we depict noteworthy but unsystematic variations in obtained Cohen's d estimates for each RT throughout all analysis specifications in the multiverse. For example, the inclusion of control variables increases effects in absolute terms for some RTs, while for other RTs, it attenuates the effects obtained by univariate OLS regressions (Supplementary Figs. 8 and 10). Second, we applied robustness checks for meta-research (that is, the leave-one-out method) to the results regarding real-world and stated support (Supplementary Figs. 11 and 13) and explored the variation of meta-effects and heterogeneity values across 1 million different random subsets of study combinations, also known as GOSH plots (Supplementary Figs. 12 and 14 and Supplementary Tables 9 and 11). All these robustness diagnostics support the main results. Next, we tested the robustness of our main results to potential mechanical and task-specific effects of the interventions on real-world outcomes, thus, studies in which responses to the intervention cannot be cleanly separated from responses to incentives, defaults or other features of the outcome measurement. In this exploratory sensitivity analysis, we excluded seven designs in which the intervention may have directly affected features of the behavioural outcome and found that the subset of non-affected studies also yields a detectable meta-effect ($d = 0.049$, 99.5% CI −0.012 to 0.111, $k = 48$, $P = 0.012$; see Supplementary Tables 10 and 12, columns 3, for stated support). This constitutes suggestive evidence under our preregistered significance threshold and indicates that the estimate remains small and positive, although with lower inferential strength. Finally, to provide a proxy for how our conclusions might differ in the presence of publication bias, we exploratorily repeated the analyses using only studies that show statistically significant evidence of a positive effect in any of our six primary outcome variables. The pooled meta-effect for real-world support is Cohen's $d = 0.219$ (99.5% CI 0.060 to 0.377, $z = 3.865$, $k = 12$, $P < 0.001$; Supplementary Table 10, column 2), while Cohen's $d = 0.256$ for stated support (99.5% CI 0.135 to 0.376, $z = 5.961$, $k = 12$, $P < 0.001$; Supplementary Table 12, column 2). Notably, these estimates are approximately three and a half times larger than the ones based on the full set of studies.

Discussion and conclusion

Carbon taxes are widely discussed political instruments for reducing carbon emissions to mitigate global warming. This project represents an effort to systematically leverage the creativity of international researchers to design behavioural interventions to foster observed real-world support for carbon pricing. Relying on a 'many-designs' methodology, we presented results of 55 experimental studies independently proposed by international RTs, without any preselection of designs. Compared with relying on isolated and uncoordinated studies, this approach encourages the simultaneous development of diverse ideas around a common research question. In addition to common self-reported measures for support included in all designs, this study measured the interventions' impact on a design-specific set of heterogeneous proxy measures of real-world policy support, advancing the external validity of the reported findings. Following our preregistered analysis plan, the findings are presented transparently, free from selective reporting, P-hacking, HARKing and forking[35,44–46].

The study yields two main findings: first, behavioural interventions produce statistically significant but very small increases in observed and stated support for carbon pricing in the USA (Cohen's $d = 0.06$–0.08; between 1.7 and 2.4 percentage points in increased support), along with modest between-study heterogeneity ($\tau = 0.07$–0.12). Second, the researchers participating in the study substantially overestimate the capacity of such interventions to shift public support for carbon pricing. As implications, the expectations among the scientific community on the capacity of behavioural interventions to foster support for carbon pricing should be tempered, results from individual studies should be interpreted with caution, and policy-makers should focus more strongly on meta-effects from classical meta-analyses, crowdscience studies and large-sample studies.

Specifically, for a sample of almost 20,000 US residents, we found positive and statistically significant meta-effects of behavioural interventions on the set of design-specific real-world proxies of carbon pricing support—mostly donations to organizations advocating for carbon pricing—as well as general self-reported support for carbon pricing, self-reported support for an increase in carbon pricing, a carbon price with rebates and self-reported support for a carbon price internalizing the social cost of carbon at US$120 per ton of CO_2 emissions. The reported meta-effect for real-world outcome measures is a Cohen's $d = 0.063$ (2.3 percentage-point increase in support for carbon pricing) and for self-reported outcomes, a Cohen's $d \approx 0.037$–0.089 (1.1–2.4 percentage-point increase in support). The effects, despite being small, are non-negligible. Indeed, even modest attitudinal shifts could affect social norms, tipping points and, ultimately, policy outcomes[12,13,18,19]. This result contrasts with interventions targeting individual behaviours, where there might be no comparable collective threshold or direct policy leverage[8,10]. When examining heterogeneity across both observed and stated effect sizes, we find suggestive exploratory evidence that variation in observed effects is partly accounted for by the type of outcome variable proposed by the RTs. By contrast, our preregistered moderators (type of interventions, expected informativeness of interventions by peers and expected effect sizes by peers) do not explain effect-size heterogeneity for both general support observed and stated. Moreover, when examining whether specific design features are associated with positive effects, we find no clear pattern across observed and stated outcome types. Which conclusions one draws from individual design characteristics producing larger or smaller effects thus depend on the outcome considered. Neither outcome type is, on its own, a complete measure of real-world policy support. The common stated measures provide a more standardized basis for comparing interventions across designs, whereas the behavioural proxies are consequential but capture heterogeneous and often task-specific responses that may also reflect features of the immediate choice environment. More generally, as the RTs independently defined their intervention condition and their real-world support proxy, each of the interventions was not evaluated for each of the alternative real-world support proxies. Thus, this study cannot determine whether the differences between studies reflect intervention efficacy, outcome sensitivity or the idiosyncratic combination of interventions and outcomes. Consequently, the results of an intervention on its proxy for real-world support for carbon pricing should not be taken to generalize to other measures of real-world support for carbon pricing.

Importantly, irrespective of the measure, our overall findings are smaller than those generally reported by previous behavioural literature. This could be explained by a combination of two factors. First, it could be that preferences about carbon pricing are rigid and thus not easily malleable. This might be associated with general opposition to taxes or with strong opinions on climate pricing related to political identities[21,47]. Related, it could be that behavioural interventions are

not sufficiently effective in modifying carbon pricing support. While behavioural interventions can trigger individual behavioural change (review by Byerly et al.[48]; meta-analysis by Bergquist et al.[49]; recent studies, for example, by Sinclair et al.[6] or Andre et al.[7]), it could be that fostering support for public policies affecting system-wide decisions might be inherently difficult. Vlasceanu et al.[20] found that 6 out of 11 behavioural interventions had a positively significant impact on stated support for climate policies. A letter to a member of the future generation, increasing policy support by 2.6%, had the largest average treatment effect. None of the interventions increased individual real-life behaviours, with 7 out of 11 interventions showing significant negative effects. Other studies present similar mixed results[21,32]. Second, our findings are consistent with concerns about systematic publication bias in literature addressing the impact of behavioural interventions on individual behavioural change. Evidence from psychology suggests that 'classical' one-team studies published in journals report statistically significant results in more than 95% of cases, compared with around 50% in preregistered studies and registered reports, where all data points and findings must be reported[50]. Similarly, at the individual-study level in this project, most effects are statistically not distinguishable from zero (two-sided *t*-tests): 51 out of 55 studies at the 0.5% level and 45 out of 55 at the 5% level in terms of real-world observed outcome. Eight studies showed statistically significant or suggestive positive evidence, while two showed statistically significant or suggestive negative evidence. This dispersion has important implications for how selective reporting may shape published literature. Selection based on statistically significant results in specific directions—whether driven by researchers' beliefs[51,52] (for a more critical perspective, see Auspurg and Brüderl[53])—editorial interest or both, would tend to over-represent non-null findings pointing in those directions. For example, if only the studies that reached our preregistered significance threshold in at least one primary outcome had been reported, the pooled meta-effect on real-world support would be Cohen's $d = 0.22$ and $d = 0.26$ for stated support, approximately three and a half times larger than the estimate based on the full set of studies. As a potential consequence of such selective reporting, we also document clear evidence of individual and peer overconfidence, a pattern already observed in domains outside climate action[37,54,55]. This finding further highlights the need to moderate the scientific community's expectations about the power of behavioural interventions, at least for support for carbon pricing, as this study shows.

Despite its strengths, this 'many-designs' project also has limitations. First, we required RTs to come up with proxy outcomes for real-world policy support to increase external validity compared with self-reported measures. This requirement may have led some teams to use consequential but design-specific proxies for real-world support for which responses to the intervention cannot be cleanly separated from responses to incentives, defaults or other features of the outcome measurement. As a consequence, these real-world effect estimates might be higher than their general stated counterparts. Still, such estimates remain informative about behaviour within the respective choice environment but might not necessarily identify genuine changes in policy support. An exploratory sensitivity analysis excluding seven designs in which the intervention may have directly affected features of the behavioural outcome still yields a detectable meta-effect. Furthermore, the weak correspondence between the stated and proxy measures for real-world support for carbon pricing may also partly reflect well-established differences between costless self-reports and consequential behaviour. A large body of research shows that stated attitudes and preferences often translate only imperfectly into actual behaviour, particularly when behavioural measures require participants to incur real costs or effort[56,57]. This discrepancy may be further amplified by social desirability concerns[58] and experimenter demand effects[59], all of which may inflate stated support relative to consequential real-world measures. Second, results of individual studies need to be treated with caution, as for each single study, we were underpowered to detect small effects with high probability (that is, Cohen's $d < 0.2$ with power ≥80%). As such, ceteris paribus, statistically insignificant results reported for single studies might become statistically significant with larger sample size. In any case, caution is always advisable against drawing policy implications from single lower-powered studies. Nevertheless, the main conclusions of our many-designs project remain stable, given the substantially greater statistical power achieved by pooling evidence across 55 studies and close to 20,000 participants. Third, we randomly selected 'only' 55 RTs from 135 applicant teams, all of whom met our inclusion criteria and represented a subset of the research community.

As implementing 135 studies turned out to be unfeasible, we considered this random selection approach the most appropriate way to avoid systematic selection effects and to meta-scientifically capture the creativity of the field in an unbiased manner. Finally, our study focuses on a single country, the USA, and further studies could investigate whether the results presented generalize to other cultures and societies.

Methods

Ethics

The study was preregistered and received approval by the Board for Ethical Questions in Science of the University of Innsbruck (certificate of good standing, 30/2024).

Study protocol

We preregistered an extensive pre-analysis plan on OSF (https://doi.org/10.17605/OSF.IO/35KRQ) outlining the entire structure, procedure and the main part of the data analysis. Furthermore, the pre-analysis plan included protocols for communication with the teams and the participation requirements. The preregistered study protocol, data and analysis files are available via the OSF repository at https://osf.io/ujvdn.

We then began an open call for collaboration (scientific email lists such as the Economic Science Association, Society of Experimental Finance, American and European Associations of Environmental and Resource Economists, and Judgment and Decision Making, as well as social media posts) while simultaneously launching the project website (https://manydesignscarbon.online), which was the main communication channel. The general research question that was presented on the project webpage and clearly communicated to the RTs was, "What is the impact of behavioral interventions on real-world support for a price of carbon?". The spirit of the project was to tackle the introduction of a carbon price or the substantial increases in the price of carbon, aiming to advance towards carbon pricing incorporating external damages from CO_2 emissions. This is referred to as the social price, and we refer to the lower-bound estimate by the EPA[28] of US$120 per ton of CO_2. RTs were not required to tailor their interventions specifically to this price, but it was mentioned as a reference value, explicitly included on the project website.

Overall, 135 international RTs indicated interest in the project, all of which complied with the prerequisites (maximum of two members, at least one with a PhD, and previous experience with experimental methods—at least one team member needed one experimental study in the social sciences that was either published or a working paper). We originally chose 42 random RTs with a preregistered STATA script to be fully funded by the project coordinators, on the basis of the ex ante power analysis. In addition, owing to the large number of applications by RTs (that is, 135 was much higher than our initial expectations) and the offer of some RTs to fund their design themselves, in addition to the already randomly selected RTs, we preregistered an Addendum A to the pre-analysis plan, specifying the random selection of 25 additional RTs using the same preregistered STATA script. In the few cases where these additional RTs could not self-fund their study, the core team covered the funding, and the availability of funding played no

role in being included in the study, thus, avoiding selection bias. Next to funding constraints, the upper bound of 67 invited RTs was defined by the technical feasibility on Prolific to simultaneously run the different treatments from all designs. In the end, 55 of the selected 67 RTs completed the full study protocol on time. Researchers from RTs whose designs are included in the project are full co-authors of this paper. RTs cannot be connected to the individual studies mentioned in the paper.

The RTs had several tasks to complete. First, they had to submit their design through the website. RTs described the intervention, control condition, suggested real-world outcome measure and details related to their study, such as expected duration and the incentives (see the preregistration template in the pre-analysis on the OSF repository). In addition, RTs were required to complete a self-assessment survey, in which we asked them about the effect size that they expected from their proposals, in Cohen's *d* units (Supplementary Section 14). The project coordinators carefully assessed all proposed designs to evaluate their fit to the project guidelines (that is, institutional review board criteria for fast-track approval, use of economic incentives as compensation, design qualifying as behavioural intervention and the suggested outcome having a real-world carbon pricing support). We notified the RTs that satisfied all criteria and asked them to start preparing the software (either in oTree or Qualtrics). In case proposals deviated from the project guidelines in any dimension, we contacted the RTs requesting revisions before developing their software. In all cases, we requested RTs to add a preprogrammed template of the 'policy survey' by the project coordinators (either an *.otreezip or *.qsf file) at the end of their software (see Supplementary Section 15 for the full policy survey). In the meantime, RTs were also randomly assigned ten other designs to assess: (1) the informativeness for answering the research question (Likert scale from 0 to 10); (2) the category that best described, in their view, the intervention; and (3) the expected effect sizes as measured in Cohen's *d* units. For the main data collection, we used Prolific with the following sample screening criteria: US based, English fluency and a 95% approval rate on the platform. We used an oTree-based balancer software from Huber et al.[31] to randomly assign the participants to one of the 110 experimental conditions (see the OSF repository for the preregistered balancer software code). Every participant from Prolific was directed to a common landing page with a general introduction and informed consent form, as well as a captcha task to prevent bots. Then, batches of four people were formed to permit group experiments of up to four people.

Data analysis

Data collection and analysis were not performed blinded to the conditions of the experiments. We only used completed observations on Prolific (either manually approved or flagged as completed by reaching the very last page of the study) and had no other exclusion criteria applied. We standardized the analytical approach across RTs (see below). There was only one prerequisite for the RTs. They needed to come up with an outcome measure that can be analysed via an OLS if continuous or ordinal, or with a logit model if binary. Irrespective of the level of measurement, we ran an OLS regression with the proposed outcome as the dependent variable and one dummy variable equal to 1 for observations from the treatment condition (0 otherwise), as registered (see Addendum C). We standardized the treatment effects across RTs to Cohen's *d*. We carried out a random effects meta-analysis on the basis of the standardized analytical approach across the teams[60]. As a robustness test and to address analytical heterogeneity, we conducted a multiverse analysis[35] for each individual study. This approach considers the flexibility inherent in research-design choices.

Standardizing the analytical approach across RTs. The effect size for a given design is based on the intervention treatment dummy variable coefficient in an OLS regression. For our main analysis, we ran a univariate model with only the treatment indicator as an independent variable, as well as a multivariate model including all control variables (see Supplementary Section 9 for the list of control variables). Thus, we obtained two Cohen's *d*'s per study for our main analysis. We used individual-level data and only included one observation per individual. In general, we estimated robust standard errors (HC3). However, in case RTs implemented interactive components via grouping participants, we used (CR2) cluster robust standard errors on the group level.

One exception to this common analytical approach was RT 'XYN28'. Based on informal conversations with this RT, we restrict the analysis to participants in the intervention who agreed to the favour ex ante, avoiding confounding by simultaneously differential information sets for those saying 'no'. Accordingly, we estimate the average treatment effect on the treated rather than the intention-to-treat effect. However, both (that is, the average treatment effect (ATE) and intention to treat (ITT)) deliver similar results (Cohen's *d* estimates obtained in univariate OLS regressions–ATE: $d = -0.101$, 99.5% CI −0.406 to 0.203; $t(341) = -0.933$; ITT: $d = -0.140$, 99.5% CI −0.437 to 0.157; $t(357) = -1.325$), thus not affecting the main conclusions of the manuscript given the small weight of each study in our random-effects meta-analyses based on the study's standard errors in effect sizes. Note that this restriction applies to all outcomes of 'XYN28'.

Converting effect sizes to Cohen's *d*. The primary outcome variables included the measure of real-world support for a price of carbon proposed by the RTs. As RTs were expected to use different measures of support for a price of carbon, we standardized the treatment effects across RTs to conduct a meta-analysis. Thus, we converted all treatment effects to Cohen's *d* so that they are measured in terms of s.d. units of the outcome variables. To convert effect sizes to Cohen's *d*, we used the following formulas. We applied the same procedure as in the study by Huber et al.[31]:

$$\text{Cohen's } d = \frac{\text{Effect size}}{\text{Pooled s.d.}}.$$

$$\text{Standard error of Cohen's } d = \frac{\text{Standard error of effect size}}{\text{pooled SD}}$$

Here, the effect size for a given design is the estimated treatment effect before standardization; that is, the treatment dummy variable coefficient β_1 in the OLS regression; the standard error of the effect size is the standard error of the estimate β_1 for the treatment dummy variable in the OLS regression. The pooled s.d. for a given design is defined as

$$\text{Pooled s.d.} = \sqrt{\frac{(n_1 - 1) \times var_1 + (n_2 - 1) \times var_2}{(n_1 + n_2 - 2)}}.$$

Here, var_1 is the variance of the outcome variable in the control group of that design and n_1 is the sample size in the control group of the same design. Moreover, var_2 is the variance of the outcome variable in the intervention treatment, again, of the same design and n_2 is the sample size in the corresponding intervention treatment. We estimated the pooled s.d. (var_1, var_2, n_1 and n_2); this estimate of the pooled s.d. was used in the conversion of effect sizes to Cohen's *d* for a particular RT.

Coding of design categories and outcomes

To classify the behavioural intervention types, we used the peer-assessment responses in which RTs categorized other randomly selected designs and assigned each design to one of four categories (prespecified by us) using GPT-4o with temperature set to zero. The AI-generated classifications were then reviewed by independent research assistants, and final classifications were determined unanimously; full details, prompts, outputs and manual adjustments are reported in Supplementary Section 10. The resulting intervention-type

variable distinguishes between information treatments, strategic group interactions, behavioural biases/reflective thinking and the use of AI.

Subcategory coding principle

We coded each study by the active design feature that distinguished the treatment condition from the control condition. Information, tasks or incentives that were identical across treatment and control arms were not used to define the intervention subcategory. The main intervention categories were retained from the original AI classification and the subcategories refine these categories only where several clearly separable design features occurred within a main category. When a main category had too few studies or one clearly defining common feature, no additional subcategory was entered. The subcategory labels were chosen to be interpretable in Fig. 1 while remaining close to established terminology in behavioural-science and climate-policy literature. The main sources were the Behaviour Change Technique Taxonomy v1[61], the choice-architecture taxonomy in ref. 62 and carbon-pricing acceptance literature on perceived policy effectiveness, social norms, fairness and revenue recycling[63,64]. These subcategories are descriptive and exploratory. See Supplementary Section 11 for details on the subcategories.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The preregistered study protocol, data and analysis files are available via the OSF repository at https://osf.io/ujvdn. The preregistration is available at https://doi.org/10.17605/OSF.IO/35KRQ. The project website https://manydesignscarbon.online was our main channel of communication and also includes a comprehensive and interactive version of the paper's main results under the 'results' tab.

References

1. Allcott, H. Social norms and energy conservation. *J. Public Econ.* **95**, 1082–1095 (2011).
2. Allcott, H. & Rogers, T. The short-run and long-run effects of behavioral interventions: experimental evidence from energy conservation. *Am. Econ. Rev.* **104**, 3003–3037 (2014).
3. Tiefenbeck, V., Wörner, A., Schöb, S., Fleisch, E. & Staake, T. Real-time feedback promotes energy conservation in the absence of volunteer selection bias and monetary incentives. *Nat. Energy* **4**, 35–41 (2019).
4. Lohmann, P. M., Gsottbauer, E., Doherty, A. & Kontoleon, A. Do carbon footprint labels promote climatarian diets? Evidence from a large-scale field experiment. *J. Environ. Econ. Manag.* **114**, 102693 (2022).
5. Ebeling, F. & Lotz, S. Domestic uptake of green energy promoted by opt-out tariffs. *Nat. Clim. Change* **5**, 868–871 (2015).
6. Sinclair, A. H. et al. Behavioral interventions motivate action to address climate change. *Proc. Natl Acad. Sci. USA* **122**, e2426768122 (2025).
7. Andre, P., Boneva, T., Chopra, F. & Falk, A. Misperceived social norms and willingness to act against climate change. *Rev. Econ. Stat.* **108**, 1205–1221 (2026).
8. DellaVigna, S. & Linos, E. RCTs to scale: comprehensive evidence from two nudge units. *Econometrica* **90**, 81–116 (2022).
9. Maier, M. et al. No evidence for nudging after adjusting for publication bias. *Proc. Natl Acad. Sci. USA* **119**, e2200300119 (2022).
10. Kaiser, T., Kloidt, J., Mata, J. & Hertwig, R. *A Meta-Meta-Analysis of Behavior Change Interventions: Two Tales of Behavior Change* (CESifo, 2025).
11. Mertens, S., Herberz, M., Hahnel, U. J. J. & Brosch, T. The effectiveness of nudging: a meta-analysis of choice architecture interventions across behavioral domains. *Proc. Natl Acad. Sci. USA* **119**, e2107346118 (2022).
12. Connolly, D. J., Loewenstein, G. & Chater, N. An s-frame agenda for behavioral public policy research. *Behav. Public Policy* **9**, 593–613 (2025).
13. Granulo, A., Fuchs, C. & Böhm, R. Psychological reactance to system-level policies before and after their implementation. *Proc. Natl Acad. Sci. USA* **122**, e2409907122 (2025).
14. Andre, P., Boneva, T., Chopra, F. & Falk, A. Globally representative evidence on the actual and perceived support for climate action. *Nat. Clim. Change* **14**, 253–259 (2024).
15. Carson, R. T. et al. The public's views on climate policies in seven large global south countries. *Nat. Clim. Change* **15**, 954–962 (2025).
16. Umit, R. & Schaffer, L. M. Attitudes towards carbon taxes across Europe: the role of perceived uncertainty and self-interest. *Energy Policy* **140**, 111385 (2020).
17. Harring, N., Jönsson, E., Matti, S., Mundaca, G. & Jagers, S. C. Public acceptance of fossil fuel subsidy removal can be reinforced with revenue recycling. *Nat. Clim. Change* **13**, 214–215 (2023).
18. IPCC. *Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge Univ. Press, 2022).
19. Otto, I. M. et al. Social tipping dynamics for stabilizing Earth's climate by 2050. *Proc. Natl Acad. Sci. USA* **117**, 2354–2365 (2020).
20. Vlasceanu, M. et al. Addressing climate change with behavioral science: a global intervention tournament in 63 countries. *Sci. Adv.* **10**, eadj5778 (2024).
21. Dechezleprêtre, A. et al. Fighting climate change: international attitudes toward climate policies. *Am. Econ. Rev.* **115**, 1258–1300 (2025).
22. Hájek, M., Zimmermannová, J., Helman, K. & Rozenský, L. Analysis of carbon tax efficiency in energy industries of selected EU countries. *Energy Policy* **134**, 110955 (2019).
23. Stechemesser, A. et al. Climate policies that achieved major emission reductions: global evidence from two decades. *Science* **385**, 884–892 (2024).
24. Nordhaus, W. D. An optimal transition path for controlling greenhouse gases. *Science* **258**, 1315–1319 (1992).
25. Barrage, L. & Nordhaus, W. Policies, projections, and the social cost of carbon: results from the DICE-2023 model. *Proc. Natl Acad. Sci. USA* **121**, e2312030121 (2024).
26. Bayer, P. & Aklin, M. The European Union Emissions Trading System reduced CO_2 emissions despite low prices. *Proc. Natl Acad. Sci. USA* **117**, 8804–8812 (2020).
27. *State and Trends of Carbon Pricing 2025* (World Bank, 2025).
28. *Report on the Social Cost of Greenhouse Gases: Estimates Incorporating Recent Scientific Advances* https://www.epa.gov/system/files/documents/2023-12/epa_scghg_2023_report_final.pdf (US Environmental Protection Agency, National Center for Environmental Economics, Office of Policy, Climate Change Division, Office of Air and Radiation, 2023).
29. Tol, R. S. J. Social cost of carbon estimates have increased over time. *Nat. Clim. Change* **13**, 532–536 (2023).
30. Ankel-Peters, J., Bensch, G., Dabadge, A. et al. Tax carbon cautiously for sub-Saharan Africa. *Nat. Clim. Change* **15**, 7–9 (2025).
31. Huber, C. et al. Competition and moral behavior: a meta-analysis of forty-five crowd-sourced experimental designs. *Proc. Natl Acad. Sci. USA* **120**, e2215572120 (2023).

32. Woerner, A., Imai, T., Pace, D. D. & Schmidt, K. M. How to increase public support for carbon pricing with revenue recycling. *Nat. Sustain.* **7**, 1633–1641 (2024).
33. Wicherts, J. M. et al. Degrees of freedom in planning, running, analyzing, and reporting psychological studies: a checklist to avoid *P*-hacking. *Front. Psychol.* **7**, 1832 (2016).
34. Landy, J. et al. Crowdsourcing hypothesis tests: making transparent how design choices shape research results. *Psychol. Bull.* **146**, 451–479 (2020).
35. Simonsohn, U., Simmons, J. & Nelson, L. Specification curve analysis. *Nat. Hum. Behav.* **4**, 1208–1214 (2020).
36. Holzmeister, F. et al. Heterogeneity in effect size estimates. *Proc. Natl Acad. Sci. USA* **121**, e2403490121 (2024).
37. Menkveld, A. J. et al. Non-standard errors. *J. Finance* **79**, 2339–2390 (2024).
38. Palan, S. & Schitter, C. Prolific.ac—a subject pool for online experiments. *J. Behav. Exp. Finance* **17**, 22–27 (2018).
39. Anderson, S., Marinescu, I. & Shor, B. Can Pigou at the polls stop us melting the poles?. *J. Assoc. Environ. Resour. Econ.* **10**, 903–945 (2023).
40. Baldwin, M. & Lammers, J. Past-focused environmental comparisons promote proenvironmental outcomes for conservatives. *Proc. Natl Acad. Sci. USA* **113**, 14953–14957 (2016).
41. Benjamin, D. J. et al. Redefine statistical significance. *Nat. Hum. Behav.* **2**, 6–10 (2018).
42. Higgins, J. P., Thompson, S. G., Deeks, J. J. & Altman, D. G. Measuring inconsistency in meta-analyses. *BMJ* **327**, 557–560 (2003).
43. Pigott, T. *Advances in Meta-Analysis* (Springer Science & Business Media, 2012).
44. Munafò, M. et al. A manifesto for reproducible science. *Nat. Hum. Behav.* **1**, 0021 (2017).
45. Gelman, A. & Loken, E. The statistical crisis in science. *Am. Sci.* **102**, 460–465 (2014).
46. Rubin, M. When does harking hurt? Identifying when different types of undisclosed post hoc hypothesizing harm scientific progress. *Rev. Gen. Psychol.* **21**, 308–320 (2017).
47. Mildenberger, M., Lachapelle, E., Harrison, K. & Stadelmann-Steffen, I. Limited impacts of carbon tax rebate programmes on public support for carbon pricing. *Nat. Clim. Change* **12**, 141–147 (2022).
48. Byerly, H. et al. Nudging pro-environmental behavior: evidence and opportunities. *Front. Ecol. Environ.* **16**, 159–168 (2018).
49. Bergquist, M., Thiel, M., Goldberg, M. H. & van der Linden, S. Field interventions for climate change mitigation behaviors: a second-order meta-analysis. *Proc. Natl Acad. Sci. USA* **120**, e2214851120 (2023).
50. Scheel, A. M., Schijen, M. R. & Lakens, D. An excess of positive results: comparing the standard psychology literature with registered reports. *Adv. Methods Pract. Psychol. Sci.* **4**, 25152459211007467 (2021).
51. Borjas, G. J. & Breznau, N. Ideological bias in the production of research findings. *Sci. Adv.* **12**, eadz7173 (2026).
52. Rainero, V., Stolz, J. & Luijkx, R. The faith factor. How scholars' religiosity biases research findings on secularization. *Sociol. Sci.* **13**, 154–177 (2026).
53. Auspurg, K. & Brüderl, J. *Fragile Evidence for an Ideological Bias in the Production of Research Findings: Comment on Borjas and Breznau* https://www.econstor.eu/bitstream/10419/341575/1/I4R-DP303.pdf (Institute for Replication, 2026).
54. König-Kersting, C. et al. A *Many-designs Study of 516 Algorithms Challenges the Predictability of Wisdom of the Crowd Forecast Aggregation* https://www2.uibk.ac.at/downloads/c9821000/wpaper/2026-05.pdf (Faculty of Economics and Statistics, Universität Innsbruck, 2026).
55. Botvinik-Nezer, R. et al. Variability in the analysis of a single neuroimaging dataset by many teams. *Nature* **582**, 84–88 (2020).
56. Wicker, A. W. Attitudes versus actions: the relationship of verbal and overt behavioral responses to attitude objects. *J. Soc. Issues* **25**, 41–78 (1969).
57. Kraus, S. J. Attitudes and the prediction of behavior: a meta-analysis of the empirical literature. *Pers. Soc. Psychol. Bull.* **21**, 58–75 (1995).
58. Krumpal, I. Determinants of social desirability bias in sensitive surveys: a literature review. *Qual. Quant.* **47**, 2025–2047 (2013).
59. Zizzo, D. J. Experimenter demand effects in economic experiments. *Exp. Econ.* **13**, 75–98 (2010).
60. DerSimonian, R. & Laird, N. Meta-analysis in clinical trials. *Control. Clin. Trials* **7**, 177–188 (1986).
61. Michie, S. et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. *Ann. Behav. Med.* **46**, 81–95 (2013).
62. Münscher, R., Vetter, M. & Scheuerle, T. A review and taxonomy of choice architecture techniques. *J. Behav. Decis. Mak.* **29**, 511–524 (2016).
63. Drews, S. & van den Bergh, J. C. J. M. What explains public support for climate policies? A review of empirical and experimental studies. *Clim. Policy* **16**, 855–876 (2016).
64. Klenert, D. et al. Making carbon pricing work for citizens. *Nat. Clim. Change* **8**, 669–677 (2018).

Acknowledgements

We thank participants at the Climate Action Workshop 2024 and 2025 in Innsbruck, the Experimental Finance Conference 2025 in Maastricht, the ESA Europe Meeting 2025 in Brno, the ICEP 2025 in Vilnius, the SABE 2025 in Trento, the Research Seminar at Montpellier Business School (2025), the 4th Behavioral Economics Workshop on Sustainability at the European University Valencia, and the 2026 World Congress of Environmental and Resource Economists (WCERE).

Author contributions

Project coordinators: A. Holzknecht, R. Schwaige, E. Blanco, J. Huber and M. Kirchler designed and organized the study (that is, the entire project). A. Holzknecht, R. Schwaiger, E. Blanco, J. Huber. and M. Kirchler assessed whether the proposals fulfil all guidelines. A. Holzknecht and R. Schwaiger managed the administration of the study and the correspondence with all participating RTs (that is, the entire project). A. Holzknecht, R. Schwaiger E.B., J.H. and M. Kirchler preregistered the study. A. Holzknecht and R. Schwaiger maintained the OSF repository. A. Holzknecht and R. Schwaige managed the entire data collection (that is, of both, the RTs and the main data collection of participants on Prolific). A. Holzknecht and R. Schwaiger programmed the website and the blueprints for the policy survey, and A. Holzknecht adapted the balancer software. A. Holzknecht and R. Schwaiger conducted the preregistered data analyses, A. Holzknecht implemented the entire main data analyses (that is, Figs. 1–4, Supplementary Figs. 1–24 and Supplementary Tables 1–18, as well as the robustness analyses, including the multiverse analyses). R. Schwaiger implemented the power analysis, the descriptive figures on designs and RTs as well as the clustering plus factor analysis on the exploratory survey. A. Holzknecht generated the replication kit for the data analyses. A. Holzknecht, R. Schwaiger, E. Blanco, J. Huber and M. Kirchler wrote the first draft of the manuscript. A. Holzknecht and R. Schwaiger prepared the Supplementary Information for the submission, and E. Blanco, J. Huber and M. Kirchler created the descriptive table for all designs. Participating RTs: A.L.A., A.D., A.G.-K., A. Herziger, A.K., A. Marcos, A. Mitra, A. Molnar, A. Richter, A. Rohmann, B.B., C.D., C.E., C.H., C.K., C.L., C.R.S., C.S., D.C., D.M.F., E.E.D., E.H., E.W., F.C., F.E., F.K., F.M., F.W., G.B., G.H., G.P., G.S., H.B., H.E., H.F.-R., H.H., H.Y.,

I. Mermelstein, I. Mussio, I.S., I.V., J.B.-H., J.C.S., J.P., J.Q., J.R., J. Schmitz, J. Šrol, J.V.Y., K.F.L., K.L.V., K.M., L. Bakalo, L. Burgstaller, L.F., L.M., L.E.N., M.A., M.B., M.D., M.G., M. Kasper, M.C.O., M.P., M.S.-O., M.S.-M., M.W., N.J., N.P., N.S., O.B., O.M., O.P., P. Basaglia, P. Brañas, P.M.G., P.H., P.M., P.N., R.C., R.G., R.P.K., R.R., R. Schmacker, S.B., S.J.G., S.G., S.N., S.H., S.L., S.W., S.S., T.J., T.R.K., T. König, T.P.-R., T.V., Ü.B., U.S., V.F., V.M., V. Vivek (independent researcher), W.F., W.L., X.Z., Y.R., Y.X. and Z.R.A. drafted and programmed their study, reviewed the final manuscript, gave feedback and approved the write-up.

Funding
This research was funded by the Austrian Science Fund (FWF) project SFB F63 (https://doi.org/10.55776/F63). Open access funding provided by University of Innsbruck and Medical University of Innsbruck.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41562-026-02629-8.

Correspondence and requests for materials should be addressed to Jürgen Huber.

Peer review information *Nature Human Behaviour* thanks Jörg Ankel-Peters, Christina Gravert and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2026

Armando Holzknecht[1,98], Rene Schwaiger[1,98], Esther Blanco[2,3,99], Jürgen Huber[1,99] ✉, Michael Kirchler[1,99], Anna Lou Abatayo[4], Magdalena Adamus[5,6], Lior Bakalo[7], Jesus Barreiro-Hurle[8], Burcak Bas[9], Piero Basaglia[10], Oindrila Bhattacharya[11], Ülkü Bıçakçı[12,13], Hayley Blunden[14], Michelle Blunt[15], Sandra Bohmann[16], Pablo Brañas-Garza[17], Georgia Buckle[18], Lilith Burgstaller[19], Romain Cadario[20], Juana Castro Santa[21], Fadong Chen[22], Doron Cohen[23], Elif E. Demiral[24], Andreas Drichoutis[25], Christoph Drobner[26], Moritz A. Drupp[27], Charles Ebert[28], Hande Erkut[29], Filippos Exadaktylos[30], Valeria Fanghella[31], William Fawson[32], Lars Felder[33,34,35], Hanna Fuhrmann-Riebel[36], Sebastian J. Goerg[37,38], Paul Matthias Gorny[39], Raphael Gottweis[40], Amélie Gourdon-Kanhukamwe[15], Manuel Grieder[41], Sven Gruener[42], Peter Haan[34,43], Hooman Habibnia[40], Sebastian Hafenbrädl[44], Emeric Henry[45], Atar Herziger[7], Geraldine Holenweger[46], Christoph Huber[47], Nicolas Jacquemet[48], Tao Jin[49], Fabian Kalleitner[16], Todd R. Kaplan[50,51], Matthias Kasper[19,52], Austėja Kažemekaitytė[53], Raghabendra Pratap KC[54], Tobias König[29,55], Cankut Kuzlukluoglu[35,56], Kyle Fiore Law[57], Christophe Lembregts[20], Shuzhen Li[58], Kristian López Vargas[59,60], Wolfgang Luhan[18], Vincent Mak[61], Aitor Marcos[62], Diego Marino Fages[63], Patrick Maus[64], Friederike Mengel[65,66], Ilana Mermelstein[67], Arjun Mitra[68], Andras Molnar[69], Laura Moritz[70], Kanchan Mukherjee[71], Oliver Musshoff[56], Irene Mussio[72], Sarah Necker[12,13], Petra Nieken[39], Laila Elina Nockur[73], Marina C. Orifici[74,75], Marco Palma[76], Navya Pandit[77], Tyler Parker-Rollins[78], Julien Picard[79], Ori Plonsky[7], Giulia Priolo[80], Julian Quandt[40], Zahra Rahmani Azad[81], Rastislav Rehák[82], Andries Richter[4], Anette Rohmann[74], Jens Rommel[83], Yefim Roth[84], Umit Saglam[24], Melania Salazar-Ordoñez[85], Renke Schmacker[29,43,86], Jan Schmitz[87], Claudia R. Schneider[11], Michael Schulte-Mecklenbeck[46,88], Garima Sharma[14], Christoph Siemroth[65], Nisheeth Srivastava[89], Jakub Šrol[5,90], Ivo Steimanis[91], Stylianos Syropoulos[92], Iris Vilares[49], Vivek[93], Tobias Vorlaufer[94], Matthew Walker[95], Francisco Wang[61], Elke Weidenholzer[65], Simon Weidenholzer[65], Yilong Xu[58], Han Yin[96], Junhao Vincent Yu[97] & Xian Zhao[78]

[1]Department of Banking and Finance, University of Innsbruck, Innsbruck, Austria. [2]Department of Public Finance, University of Innsbruck, Innsbruck, Austria. [3]The Ostrom Workshop, Indiana University, Bloomington, USA. [4]Environmental Economics and Natural Resources Group, Wageningen University and Research, Wageningen, the Netherlands. [5]Institute of Experimental Psychology, Slovak Academy of Sciences, Bratislava, Slovakia. [6]Faculty of Economics and Administration, Masaryk University, Brno, Czech Republic. [7]Faculty of Data and Decision Sciences, Technion – Israel Institute of Technology, Haifa, Israel. [8]European Commission, Joint Research Centre (JRC), Seville, Spain. [9]Department of Marketing, KU Leuven, Leuven, Belgium. [10]Bordeaux School of Economics, University of Bordeaux, Bordeaux, France. [11]School of Psychology, Speech and Hearing, University of Canterbury, Christchurch, New Zealand. [12]Ludwig Erhard ifo Center for Social Market Economy and Institutional Economics, ifo Institute, Munich, Germany. [13]Friedrich-Alexander University Erlangen-Nuremberg, Nuremberg, Germany. [14]Kogod School of Business, American University, Washington, DC, USA. [15]King's College London, London, UK. [16]Department of Sociology, LMU Munich, Munich, Germany. [17]Department of Finance, Universidad Loyola Andaluca, Sevilla, Spain. [18]Faculty of Business and Law, University of Portsmouth, Portsmouth, UK. [19]Walter Eucken Institut, Freiburg im Breisgau, Germany. [20]Rotterdam School of Management, Erasmus University, Rotterdam, the Netherlands. [21]Institute of Environmental Science and Technology,

Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Spain. [22]School of Management, Neuromanagement Laboratory, and State Key Laboratory of Brain-Machine Intelligence, Zhejiang University, Hangzhou, China. [23]Carnegie Mellon University, Pittsburgh, PA, USA. [24]Department of Management and Supply Chain, East Tennessee State University, Johnson City, TN, USA. [25]Agricultural University of Athens, Athens, Greece. [26]Department of Economics, Central European University, Wien, Austria. [27]Department of Management, Technology and Economics, ETH Zurich, Zurich, Switzerland. [28]Department of Marketing and Supply Chain Management, University of Wisconsin-Eau Claire, Eau Claire, WI, USA. [29]WZB Berlin Social Science Center, Berlin, Germany. [30]Department of Econometrics, Statistics and Applied Economics, University of Barcelona, Barcelona, Spain. [31]EDHEC Business School, Nice, France. [32]London Business School, London, UK. [33]Public Economics Department, DIW Berlin, Berlin, Germany. [34]Freie Universität Berlin, Berlin, Germany. [35]Berlin School of Economics, Berlin, Germany. [36]German Institute of Development and Sustainability, Bonn, Germany. [37]Nuremberg Institute for Market Decisions, Nuremberg, Germany. [38]TUM Campus Straubing for Biotechnology and Sustainability, Technical University of Munich, Munich, Germany. [39]Karlsruhe Institute of Technology, Karlsruhe, Germany. [40]Vienna University of Economics and Business, Vienna, Austria. [41]Faculty of Business and Economics, UniDistance Suisse, Brig, Switzerland. [42]Martin Luther University Halle-Wittenberg, Wittenberg, Germany. [43]DIW Berlin, Berlin, Germany. [44]IESE Business School, University of Navarra, Barcelona, Spain. [45]Department of Economics, Sciences Po, Paris, France. [46]Department of Consumer Behavior, University of Bern, Bern, Switzerland. [47]Aalto University School of Business, Espoo, Finland. [48]Centre d'Economie de la Sorbonne, Paris School of Economics and Université Paris 1 Panthéon-Sorbonne, Paris, France. [49]University of Minnesota, Minneapolis, MN, USA. [50]Department of Economics, University of Haifa, Haifa, Israel. [51]Department of Economics, University of Exeter, Exeter, UK. [52]World Bank, Washington, DC, USA. [53]Faculty of Economics and Management, Free University of Bozen-Bolzano, Bolzano, Italy. [54]Rollins College, Winter Park, FL, USA. [55]Department of Economics and Statistics, Linnaeus University, Växjö, Sweden. [56]Humboldt-Universität zu Berlin, Berlin, Germany. [57]Department of Psychology, University of Utah, Salt Lake City, UT, USA. [58]Utrecht School of Economics, Utrecht University, Utrecht, the Netherlands. [59]University of California Santa Cruz, Santa Cruz, CA, USA. [60]Universidad del Pacfico, Lima, Peru. [61]Cambridge Judge Business School, University of Cambridge, Cambridge, UK. [62]Faculty of Economics and Business, University of the Basque Country UPV/EHU, Bilbao, Spain. [63]Durham University, Durham, UK. [64]Alfred Weber Institute for Economics, University of Heidelberg, Heidelberg, Germany. [65]Department of Economics, University of Essex, Colchester, UK. [66]Erasmus School of Economics, Erasmus University Rotterdam, Rotterdam, the Netherlands. [67]School of Information, University of Michigan, Ann Arbor, MI, USA. [68]Department of Cognitive Science, Indian Institute of Technology Kanpur, Kanpur, India. [69]University of Michigan, Ann Arbor, MI, USA. [70]University of Trento, Trento, Italy. [71]Indian Institute of Management Bangalore, Bengaluru, India. [72]University of Leeds, Leeds, UK. [73]Department of Psychology and Behavioural Sciences, Aarhus University, Aarhus, Denmark. [74]Faculty of Psychology, FernUniversität in Hagen, Hagen, Germany. [75]Department of Psychology, University of Cologne, Cologne, Germany. [76]Texas A&M University, College Station, TX, USA. [77]University of Cologne, Cologne, Germany. [78]Ohio University, Athens, GA, USA. [79]Paris-Saclay Applied Economics, University of Paris-Saclay, Palaiseau, France. [80]Department of Management, Society and Communication, Copenhagen Business School, Frederiksberg, Denmark. [81]University of Basel, Basel, Switzerland. [82]Max Planck Institute for Behavioral Economics, Bonn, Germany. [83]Swedish University of Agricultural Sciences, Uppsala, Sweden. [84]Department of Human Services, University of Haifa, Haifa, Israel. [85]WEARE-Water, Environmental and Agricultural Resources Economics Research Group, Universidad de Córdoba, Córdoba, Spain. [86]Department of Public Economics, DIW Berlin, Germany; Department of Market Behavior, WZB Berlin Social Science Center, Berlin, Germany. [87]Radboud University, Nijmegen, the Netherlands. [88]Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin, Germany. [89]Department of Cognitive Science and Department of Computer Science and Engineering, Indian Institute of Technology Kanpur, Kanpur, India. [90]School of Psychology, University of Kent, Canterbury, UK. [91]Marburg University, Marburg, Germany. [92]School of Sustainability, Arizona State University, Tempe, AZ, USA. [93]Independent researcher, Bengaluru, India. [94]Leibniz Centre for Agricultural Landscape Research, Müncheberg, Germany. [95]Newcastle University Business School, Newcastle University, Newcastle upon Tyne, UK. [96]College of Industry-Entrepreneurs and Hexie Management Research Center, Xi'an Jiaotong-Liverpool University, Suzhou, China. [97]Farmer School of Business, Miami University, Oxford, OH, USA. [98]These authors contributed equally: Armando Holzknecht, Rene Schwaiger. [99]These authors jointly supervised this work: Esther Blanco, Jürgen Huber, Michael Kirchler. ✉e-mail: juergen.huber@uibk.ac.at

nature portfolio

Corresponding author(s): Dr. Jürgen Huber

Last updated by author(s): Sep 4, 2026

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 (*n*) 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 *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ 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. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒ 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 *d*, Pearson's *r*), 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

Data was collected with 19,558 participants on Prolific. In a basic oTree balancer (oTree version 5.11) app each participant was randomly assigned to one of 110 experimental conditions (i.e., either treatment or control condition of one of 55 experimental designs), which themselves were either programmed in oTree or Qualtrics. All experimental software is available on our OSF repository (https://osf.io/ujvdn/).

Data analysis

Data analysis presented in the manuscript was performed using R version 4.4.2. The analysis scripts, as well as additional scripts for the inital sampling procedure of applicant research teams and the AI-classifictation of designs, are available on our OSF repository (https://osf.io/ujvdn/). The following packages, including their dependencies) were used and pinned using groundhog 3.2.3 (date snapshot: 2025-04-10):

abind 1.4-8; askpass 1.2.1; backports 1.5.0; base64enc 0.1-3; bayestestR 0.15.2; berryFunctions 1.22.5; bit 4.6.0; bit64 4.6.0-1; blob 1.2.4; boot 1.3-31; broom 1.0.8; bslib 0.9.0; cachem 1.1.0; callr 3.7.6; car 3.1-3; carData 3.0-5; cellranger 1.1.0; checkmate 2.3.2; class 7.3-23; classInt 0.4-11; cli 3.6.4; clipr 0.8.0; cluster 2.1.8.1; codetools 0.2-20; coin 1.4-3; collections 0.3.7; colorspace 2.1-1; cols4all 0.8; commonmark 1.9.5; CompQuadForm 1.4.3; conflicted 1.2.0; corrplot 0.95; cowplot 1.1.3; cpp11 0.5.2; crayon 1.5.3; crosstalk 1.2.1; curl 6.2.2; data.table 1.17.0; datawizard 1.0.2; DBI 1.2.3; dbplyr 2.5.0; Deriv 4.1.6; dichromat 2.0-0.1; digest 0.6.37; distributional 0.5.0; doBy 4.6.26; dplyr 1.1.4; DT 0.33; dtplyr 1.3.1; e1071 1.7-16; ellipse 0.5.0; emmeans 1.11.0; estimability 1.5.1; estimatr 1.0.6; evaluate 1.0.3; FactoMineR 2.11; fansi 1.0.6; farver 2.1.2; fastmap 1.2.0; flashClust 1.01-2; fontawesome 0.5.3; forcats 1.0.0; formatR 1.14; Formula 1.2-5; fs 1.6.5; furrr 0.3.1; future 1.34.0; gargle 1.5.2; generics 0.1.3; geojsonsf 2.0.3; geometries 0.2.4; ggfittext 0.10.2; ggnewscale 0.5.1; ggplot2 3.5.2; ggpubr 0.6.0; ggrepel 0.9.6; ggridges 0.5.6; ggsci 3.2.0; ggsignif 0.6.4; globals 0.16.3; glue 1.8.0; googledrive 2.1.1; googlesheets4 1.1.1; gridExtra 2.3; gridtext 0.1.5; gtable 0.3.6; haven 2.5.4; highr 0.11; hms 1.1.3;

htmltools 0.5.8.1; htmlwidgets 1.6.4; httpuv 1.6.15; httr 1.4.7; ids 1.0.1; insight 1.1.0; isoband 0.2.7; jpeg 0.1-11; jquerylib 0.1.4; jsonify 1.2.2; jsonlite 2.0.0; kableExtra 1.4.0; KernSmooth 2.23-26; knitr 1.50; labeling 0.4.3; later 1.4.2; lattice 0.22-7; lazyeval 0.2.2; leafem 0.2.3; leafgl 0.2.2; leaflegend 1.2.1; leaflet 2.2.2; leaflet.providers 2.0.0; leafsync 0.1.0; leaps 3.2; libcoin 1.0-10; lifecycle 1.0.4; listenv 0.9.1; litedown 0.7; lme4 1.1-37; lmtest 0.9-40; logger 0.4.0; lubridate 1.9.4; lwgeom 0.2-14; magrittr 2.0.3; markdown 2.0; MASS 7.3-65; mathjaxr 1.6-0; Matrix 1.7-3; MatrixModels 0.5-4; matrixStats 1.5.0; memoise 2.0.1; meta 8.0-2; metadat 1.4-0; metafor 4.8-0; mgcv 1.9-3; microbenchmark 1.5.0; mime 0.13; minqa 1.2.8; modelr 0.1.11; modelsummary 2.3.0; modeltools 0.2-23; multcomp 1.4-28; multcompView 0.1-10; multiverse 0.6.2; munsell 0.5.1; mvtnorm 1.3-3; nlme 3.1-168; nloptr 2.2.1; nnet 7.3-20; numDeriv 2016.8-1.1; openssl 2.3.2; parallelly 1.43.0; parameters 0.24.2; patchwork 1.3.0; pbapply 1.7-2; pbkrtest 0.5.3; performance 0.13.0; pillar 1.10.2; pkgconfig 2.0.3; png 0.1-8; polynom 1.4-1; prettyunits 1.2.0; processx 3.8.6; progress 1.2.3; promises 1.3.2; proxy 0.4-27; ps 1.9.0; purrr 1.0.4; quantreg 6.1; R.cache 0.16.0; R.methodsS3 1.8.2; R.oo 1.27.0; R.utils 2.13.0; R6 2.6.1; ragg 1.3.3; rapidjsonr 1.2.0; rappdirs 0.3.3; raster 3.6-32; rbibutils 2.3; RColorBrewer 1.1-3; Rcpp 1.0.14; RcppEigen 0.3.4.0.2; Rdpack 2.6.4; readr 2.1.5; readxl 1.4.5; reformulas 0.4.0; rematch 2.0.0; rematch2 2.1.2; reprex 2.1.1; rlang 1.1.5; rmarkdown 2.29; rnaturalearth 1.0.1; rnaturalearthdata 1.0.0; rprojroot 2.0.4; rstatix 0.7.2; rstudioapi 0.17.1; rvest 1.0.4; s2 1.1.7; sandwich 3.1-1; sass 0.4.9; scales 1.3.0; scatterplot3d 0.3-44; selectr 0.4-2; servr 0.32; sf 1.0-20; sfheaders 0.4.4; shades 1.4.0; showtext 0.9-7; showtextdb 3.0; sp 2.2-0; spacesXYZ 1.5-1; SparseM 1.84-2; stars 0.6-8; stringdist 0.9.15; stringi 1.8.7; stringr 1.5.1; styler 1.10.3; survival 3.8-3; svglite 2.1.3; sys 3.4.3; sysfonts 0.8.9; systemfonts 1.2.2; tables 0.9.31; terra 1.8-42; textshaping 1.0.0; TH.data 1.1-3; this.path 2.6.0; tibble 3.2.1; tidyr 1.3.1; tidyselect 1.2.1; tidyverse 2.0.0; timechange 0.3.0; tinytable 0.8.0; tinytex 0.56; tmap 4.0; tmaptools 3.2; treemapify 2.5.6; tzdb 0.5.0; units 0.8-7; utf8 1.2.4; uuid 1.2-1; vctrs 0.6.5; viridis 0.6.5; viridisLite 0.4.2; vroom 1.6.5; withr 3.0.2; wk 0.9.4; xfun 0.52; XML 3.99-0.18; xml2 1.3.8; xtable 1.8-4; yaml 2.3.10; zoo 1.8-13

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

The study was pre-registered and received approval by the Board for Ethical Questions in Science of the University of Innsbruck (Certificate of good standing, 30/2024). We provide the pre-registered study protocol, data, and analysis files in an OSF repository at \url{https://osf.io/ujvdn}. The pre-registration is available at \url{https://doi.org/10.17605/OSF.IO/35KRQ}. The project website \url{https://manydesignscarbon.online} was our main channel of communication and also includes a comprehensive and interactive version of the paper's main results under the ``Results'' tab.

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 report gender ratios across experimental conditions in a Table in Supplementary Information. Gender is one of our control variables (measured as categorical variable).
Reporting on race, ethnicity, or other socially relevant groupings	We did not obtain data on race nor ethnicity, but only on nationality and participants country of birth. The former is used as a control variable in our analysis. However, none of this information is explicitly reported in the paper.
Population characteristics	As control variables we elicited age (mean: 38.08), gender (female: 57.57%; male: 40.73%; other: 1.70%), nationality, education, employment ,income, political orientation, trust in government, number of children, area of living (whether rural or city).
Recruitment	We recruited a balanced U.S. general population on Prolific. Our sample is not representative. We used this sample based on the relevance of the RQ and feasibility for the large required sample on Prolific. Researchers were not blinded to experimental condition and the study hypothesis. We did only include complete observations (completion code on Prolific) or manual approval after individual inspection. Sampling procedure was random, each participant had a 1/110 chance to be sampled to each condition (55 studies with two conditions each). No self selection relevant. Participants only showed the willingness to sign up on Platforms such as Prolific and accepted the study invitation.
Ethics oversight	We obtained full IRB approvals from the University of Innsbruck for both the entire project and substudies (Certificate of good standing, 30/2024). The approvals can be found in our OSF repository (https://osf.io/ujvdn/).

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 obtained quantitative data on individual behavior across all 55 experimental designs.
Research sample	General popluation U.S. sample on Prolific.
Sampling strategy	Our oTree balancer app randomized batches of four participants to each of 110 experimental conditions. No stratification or weighting allocation was applied.
Data collection	We conducted online experiments on Prolific, which were programmed either in oTree or Qualtrics.
Timing	Data collection started at the end of January 2025 and lasted until March 2025
Data exclusions	We only used completed observations from Prolific-either by manual approving submission upon indiviudal inspection or by participants entering a valid Prolific completion code. We excluded incomplete observations (i.e., missing variables due to coding errors by individual teams, or participants not reaching the end of the experiment).
Non-participation	We do not have any information provided by Prolfic on how many participants started our study but declined participation. The average attrition rate across all designs (i.e., participants starting the experiment but not finishing it) was 6.1%.
Randomization	We randomized participants in groups of four into one of 110 experimental design conditions. Thus, each participant had the same probability to be redirected to each of the 110 conditions.

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	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*