HEIDELBERG UNIVERSITY
DEPARTMENT OF ECONOMICS

UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

Revealed preferences for policy experiments

Bettina Chlond, Timo Goeschl and Johannes Lohse

AWI DISCUSSION PAPER SERIES NO. 763
May 2025

Revealed preferences for policy experiments

By Bettina Chlond, Timo Goeschl and Johannes Lohse*

Randomized controlled trials remain underutilized in informing policy design, despite their potential. Moral objections to experimentation ("experiment aversion") have been proposed as an explanation. We present three studies with members of the general public and policy-makers that allow us to measure and compare moral approval, stated preferences as well as revealed preferences for policy experimentation, within the overarching context of a public assistance program. We find that evidence based on moral approval systematically underestimates revealed preferences for policy experimentation due to conceptual misalignment and hypothetical bias. People and policy-makers trade off possible moral objections against the benefits of policy experimentation.

I. Introduction

Randomized controlled trials (RCTs) are a cornerstone of the credibility revolution, providing causal evidence for key policy questions and informing programs aimed at poverty alleviation (Banerjee and Duflo, 2009; Egger et al., 2022), health care (Finkelstein et al., 2012), and social mobility (Chetty, Hendren and Katz, 2016). However, despite their popularity in economic research, most government programs are still designed without considering prior experimentation (Venkataramani, Underhill and Volpp, 2020; Lynch, Greiner and Cohen, 2020; Haushofer and Metcalf, 2020; Dur et al., 2024), with some policy-makers even openly opposing controlled trials (Heffetz and List, 2021).

Several factors could explain the lack of policy experimentation: Experiments being perceived as prohibitively costly, institutional inertia obstructing new processes (DellaVigna, Kim and Linos, 2024), legal mandates requiring equal treatment of citizens (Lynch, Greiner and Cohen, 2020), the benefits of causal evidence being undervalued (Vivalt and Coville, 2023), politicians having incentives for non-random policy implementation (Wang and Yang, 2024), or experts finding experiments entirely redundant due to strong priors about the status quo or the relative efficacy of policy alternatives to be evaluated experimentally.

Recent studies have introduced "experiment aversion", defined as a moral objection to randomization itself, as another significant barrier (Meyer et al., 2019; Heck et al., 2020). Evidence for experiment aversion primarily stems from studies

* Chlond: Heidelberg University and ZEW - Leibniz Centre for European Economic Research Mannheim (email: bettina.chlond@awi.uni-heidelberg.de); Goeschl: Heidelberg University and ZEW - Leibniz Centre for European Economic Research Mannheim (email: goeschl@uni-heidelberg.de); Lohse: Leuphana University and University of Birmingham (email: johannes.lohse@leuphana.de.)

employing vignette designs that elicit moral judgment. In these studies, participants appear to rate uniformly implementing either hypothetical policy A or B as morally more appropriate than their randomized assignment, despite each policy being individually uncontroversial or even desirable (Meyer et al., 2019).[1] If indeed capturing a genuine majoritarian preference among the general public, experiment aversion could impede policy experimentation either indirectly — via politicians implementing the preferences of the median voter (Black, 1958; Dur et al., 2024) — or directly, if policymakers are themselves averse to experimenting.

Moral judgments, however, do not easily map into preferences as understood by economists. They also need not dictate the decisions of those expressing them, rare cases of strong moral repugnance aside (Roth, 2007; Leuker, Samartzidis and Hertwig, 2021). In fact, across a wide variety of contexts, members of the general public and experts alike routinely trade off moral concerns against the benefits of an action rather than basing a decisions on moral judgments alone (Falk and Szech, 2013; Sharma et al., 2014; Bartling, Weber and Yao, 2015; Kirchler et al., 2016). An obvious example is human-subject experimentation in medical research. Similarly, when it comes to policy experiments, moral judgments need not lead people to reject experimentation. Therefore, both when people view negatively the process of experimentation or its outcomes, they may still decide in favor of its use because of the informational gains it generates.

A crucial open question, therefore, is whether evidence of experiment aversion, that is, moral objection to randomization in policy choice, tells us that people genuinely oppose policy experimentation, whether done by themselves or implemented by others. Providing evidence for the existence of such opposition is a matter of eliciting revealed preferences. Existing vignette designs are not well suited for delivering such evidence. First, conceptually, their focus on moral judgment is not designed to uncover either stated or revealed preferences. Second, methodologically, their hypothetical nature risks inviting biases compromising ecological validity, such as overstating moral concerns or understating the perceived usefulness of experiments. Such biases are common and have long been the focus of critiques aimed at the ecological validity of stated preference research (Rosenthal and Rosnow, 2008; Hausman, 2012; De Corte, Cairns and Grieve, 2021; Wuepper, Clemm and Wree, 2019). How problematic these conceptual or methodological issues are for accurately measuring actual support for policy experimentation – in absolute or relative terms – cannot be determined a priori.

This paper makes three main contributions. Its first contribution is to provide, for the first time, evidence on *revealed* preferences for policy experimentation, i.e., randomly allocating others to different policies. Employing a setting with tangible consequences and varying informational benefits, the paper reports on the revealed preferences of two groups of interest that could differ markedly in their choices, members of the general public and policy-makers, and compares

[1] Although commonly found, experiment aversion is not universally observed across all domains and study designs (Mislavsky, Dietvorst and Simonsohn, 2020; Mazar, Elbaek and Mitkidis, 2023).

them with evidence on moral judgments. Its second contribution is to offer a comparison of *moral* judgments and *stated* preferences regarding policy experiments. Recent contributions such as Dur et al. (2024) and Fischer et al. (2024) inform about stated preferences for policy experimentation, but lack the required control groups in order to determine whether moral judgments and stated preferences differ systematically. Our setting is designed to detect such differences, i.e. it can determine the size and direction of a measurement gap resulting from different conceptual approaches. The third contribution is to report on the gap measured between stated and revealed preferences for policy experimentation, i.e. a methodological measurement gap. This gap could be small or large, depending on the presence and direction of a hypothetical bias in stated preferences. Based on current knowledge, the size of this bias cannot be determined for the specific trade-offs inherent in decisions about policy experimentation (e.g., the implied consequences on others and the informational benefits of experimentation). In previous research, such as willingness to pay studies for non-market goods, hypothetical bias has been found to be large, leading participants to overstate their preferences by up to a factor of 3 on average (List and Gallet, 2001).

To make these contributions, we draw on three interconnected studies that progressively move from measuring moral approval of members of the public to revealed preferences of policy-makers. Study 1 measures moral approval of and stated preferences for conducting policy experiments through a set of hypothetical vignettes. Besides employing 5 vignettes used in earlier studies to elicit moral approval of experiments, it includes a new vignette on how to assign low-income households to one of two policies to increase take-up of a public assistance program. This vignette bridges to Study 2, which elicits revealed preferences in the same policy setting, but with real consequences and stakes. It employs a novel experimental design embedded within an ongoing field trial of the said public assistance program. Here, participants' choices determined whether 200 low-income households would be actually assigned to a single policy uniformly – or randomly to one of two different policies. Participants' potential earnings were linked to the quality of information generated. Study 3 extends this revealed-preference approach in the same setting to a separate sample of relevant policy experts and practitioners who are actively engaged in policy design and typically make decisions about policy experimentation.

Comparing moral approval and stated preferences elicited in Study 1 with revealed preferences elicited in Studies 2 and 3 forms the basis for deriving the three main results of this study. First, revealed preferences can evidence strong support for policy experimentation even when moral judgments in the same policy setting do not. In the context of the relevant bridging vignette in Study 1, only 39 percent of participants morally approve of policy experimentation. Yet, 72 percent of the participants choose to subject 200 low-income households to a policy experiment in Study 2. In Study 3, 67 percent of the policy experts do so.

Second, people's stated preferences can also evidence strong support for policy

experimentation even when moral judgments in the same context do not. Comparing moral judgments and stated preferences in Study 1, we observe that, across all vignettes, support for policy experimentation increases significantly when moving from eliciting moral approval to eliciting stated preferences. For the bridging vignette on the public assistance program, 61 percent of participants state that they would choose experimentation when faced with the choice. This not only affirms previous results on stated preferences for policy experimentation (Dur et al., 2024; Fischer et al., 2024). Additionally, it highlights the measurement gap that results from relying on a moral judgment concept rather than a preference concept in order to gauge support for policy experimentation.

Third, revealed preferences in favor of policy experiments do not appear to be weaker than stated preferences, but rather stronger. Comparing the stated preferences for experimentation by participants in Study 1 to the corresponding revealed preferences in Study 2, the share of participants choosing the policy experimentation increases from 61 percent to 72 percent when comparing among members of the general public. The share of policy-makers that choose to implement policy experimentation in Study 3 (67 percent) is also higher than the stated preferences in Study 1, but not significantly so. A corollary finding is that in the case presented, moral judgments underestimate revealed preferences mostly for conceptual reasons and less so for hypothetical bias.

Our findings challenge interpretations of stated moral objections as major barriers to policy experimentation. They suggest instead that when elicited as a stated or revealed preference, individuals readily trade off moral concerns about the consequences of experimentation against informational gains. Therefore, moral concerns against randomization are likely to represent a weaker barrier in practice than previously inferred from vignette studies alone. Our findings also speak to a larger literature on policymakers' willingness to generate and use evidence (Hjort et al., 2021; Vivalt and Coville, 2023; DellaVigna, Kim and Linos, 2024) as well as efforts to encourage experimentation through training or political incentives (Mehmood, Naseer and Chen, 2021; Corduneanu-Huci, Dorsch and Maarek, 2021). Our research most closely applies to the behavioral barrier at the decisive "go/no-go" moment for initiating experiments. Additionally, our results inform broader discussions on the usefulness and ethics of human subject experimentation, illustrating how individuals balance knowledge gains against concerns such as generating temporary inequality (List, 2008; Glennerster, 2017; Asiedu et al., 2021; Charness, Samek and Van de Ven, 2022; McDermott and Hatemi, 2020).

II. Methods and Design

A. Study 1

Study 1 (n = 600) links standard vignette-based designs used in the prior literature (Meyer et al., 2019) with the revealed preference framework we develop in Studies 2 and 3. Participants were recruited via Prolific.ac from a pool of German

citizens available on this platform. Each participant was randomly assigned two vignettes from six different policy contexts: five contexts previously studied (hospital safety, genetic testing, poverty alleviation, autonomous cars, and retirement savings), plus a newly created context (*public assistance program*) specifically designed as a bridge to Studies 2 and 3. That context was a program in Germany that provides low-income households with subsidy vouchers to replace old inefficient refrigerators and freezers with more efficient models. The vignette for this program was designed to present new, but untested policies of interacting with households such as to ensure high program take-up. See the Supplemental Appendix for the exact design of each vignette used.

The primary goal of Study 1 was to assess the extent to which the conceptual grounding of experiment aversion in moral judgments drives existing results. Participants were randomly assigned to one of three between-subjects treatments that systematically varied the framing of their decision, shifting from moral appropriateness alone towards a broader consideration of trade-offs, allowing subjects to express their stated preferences.

In each vignette, participants were prompted either to judge the moral appropriateness of, or (in the two alternative treatment conditions) to choose between, implementing one of two policies uniformly (Policy A or Policy B) or implementing a randomized A/B experiment testing both. For instance, the newly developed vignette on a public assistance program required participants to consider a Policy A (framing eligibility information letters in terms of losses from non-participation), a Policy B (installing physical reminders inside refrigerators), or a randomized A/B experiment testing both policies. Across the three between-subjects treatments, we varied the decision framing, asking participants (i) which option was morally most appropriate (conceptually replicating previous research), (ii) which option they would prefer a third-party policymaker to implement (*Stated Preference Other*), or (iii) which option they themselves would implement if they were the decision-maker (*Stated Preference Self*). These variations allow for a comparison of how moral judgments align with stated preferences when participants assume the roles of citizens evaluating policymakers or the role of the actual decision-maker.

To ensure consistency of the choice format across Studies, participants selected their most preferred option for each case, as common in stated and revealed preference research. This deliberately deviates from some existing vignette designs eliciting moral judgments by asking participants to rate the appropriateness of each option individually. As we will discuss in the results section this change in elicitation format does not change the main finding of previous studies that policy experiments are not approved of on moral grounds by a majority of participants.

B. Study 2

Study 2 (n = 500), also recruited from Prolific.ac among German citizens, introduced a revealed preference design of treatment arm (iii) from Study 1 integrated

into an actual policy evaluation trial. Hewing closely to the public assistance program vignette, participants made a decision regarding how a refrigerator replacement program for low-income households should be implemented. The four options were universal assignment of Policy A (eligibility information letter framed in terms of losses from missing out on replacement), universal assignment of Policy B (physical reminder of replacement installed inside the appliance), random assignment of A or B (randomized experiment), or maintaining the status quo (information letter framed in terms of gains from replacement and no physical reminder). Crucially, like in a public referendum, each participant's decision had a positive probability of being pivotal for determining the actual policy allocation of 200 real households enrolled in the field trial, thus reflecting consequential third-party impacts.

The referendum format is a common method for eliciting revealed preferences for public goods or policy proposals more generally – whether through observational studies of real-world voting behavior (Deacon and Shapiro, 1975; Vossler and Kerkvliet, 2003) or through experimentally designed referenda conducted as part of stated or revealed preference study designs (Vossler, Doyon and Rondeau, 2012). Theoretically, truthful voting is a dominant strategy as long as participants have a positive probability of influencing the outcome and the outcome has real consequences. In our design, this was achieved by informing participants that the selected option would be implemented if their vote was randomly chosen from all votes cast. Unlike in majority voting, pivotality in this setup is both positive and independent of the participant's preferred option or their beliefs about the preferred option of others — that is, each participant had a fixed and experimentally induced probability of determining the outcome regardless of which alternative they and others supported.

Within this referendum format, Study 2 also implemented a 2×2 factorial design with two cross-cutting treatment dimensions to identify potential shift factors of revealed preferences for policy experimentation. First, we varied the relative probability that a participant's decision would become binding (low vs. high), scaling how consequential the own choice was for affected households by a factor of 10. Second, we manipulated participants' financial incentives to learn about policy effectiveness by varying the probability (10 percent vs. 90 percent) of receiving an invitation to a follow-up session. In this session, which was run after completing the field trial, participants would learn how the 200 households behaved under the assigned policy (or policy experiment), and where then asked to predict refrigerator replacement rates under Policies A and B, earning rewards for accurate predictions for the full sample. This made the information generated by the field trial (weakly or strongly) instrumental, allowing us to study how the trade-off between informational gains and possible moral concerns affects revealed preferences for experimentation. Both treatments were randomly assigned between subjects, and participants were informed about all procedures of the follow-up study and their exact probabilities of pivotality and receiving a

follow-up invitation before submitting their choice.

C. *Study 3*

Study 3 (n = 75) replicated the revealed preference methodology from Study 2 but recruited a more specialized sample of participants: policy-makers and practitioners from municipal welfare offices and an NGO managing the relevant assistance program. These experts were recruited via direct email outreach to government and NGO offices. Like participants in Study 2, these experts faced a non-zero probability that their decision would determine policy implementation for the same 200 households. Additionally, they knew they might participate in the follow-up prediction task, providing them a direct stake in the informational value of the randomized experiment.

D. *Procedures*

All analyses adhered to a pre-registered analysis plan (available at `https://aspredicted.org/54xs-mhpn.pdf`). Additional details about surveys, instructions, balance checks, and robustness analyses are available in the Supplemental Appendix. Deidentified data and replication files can be accessed at `https://doi.org/10.6084/m9.figshare.27605487`. Ethical approval was granted by the University of Birmingham (ERN21-1674).

III. Results

A. *Study 1*

Figure 1 presents the key results for condition (i) of Study 1. The study covered five vignettes (Vig. 1-5) previously used and the new bridging vignette focused on a public assistance program (Vig. 6) and elicited the choice of the morally most appropriate course of action with respect to the decision situation described in each vignette.

Each stacked bar in Figure 1 represents the distribution of responses for one vignette. It shows the percentage of participants who rate the uniform implementation of a policy (Policy A or B) or an A/B test as the morally most appropriate option. Averaging across all vignettes, 72 percent of participants favored the uniform implementation of a single policy on moral grounds while 28 percent favored the A/B experiment. At the average level, therefore, condition (i) of Study 1 returns general patterns that do not contradict previous evidence on experiment aversion. Clearly, A/B experiments are not universally disliked across all six vignettes. Support for A/B testing varies from as low as 8 percent for the morally most contentious vignette (genetic testing) to 46 percent for the least contentious (poverty alleviation). Still, even in the least contentious vignette, A/B testing would not secure a victory in a voting contest if such contest required an absolute majority and voting decisions were based on moral judgments alone. In

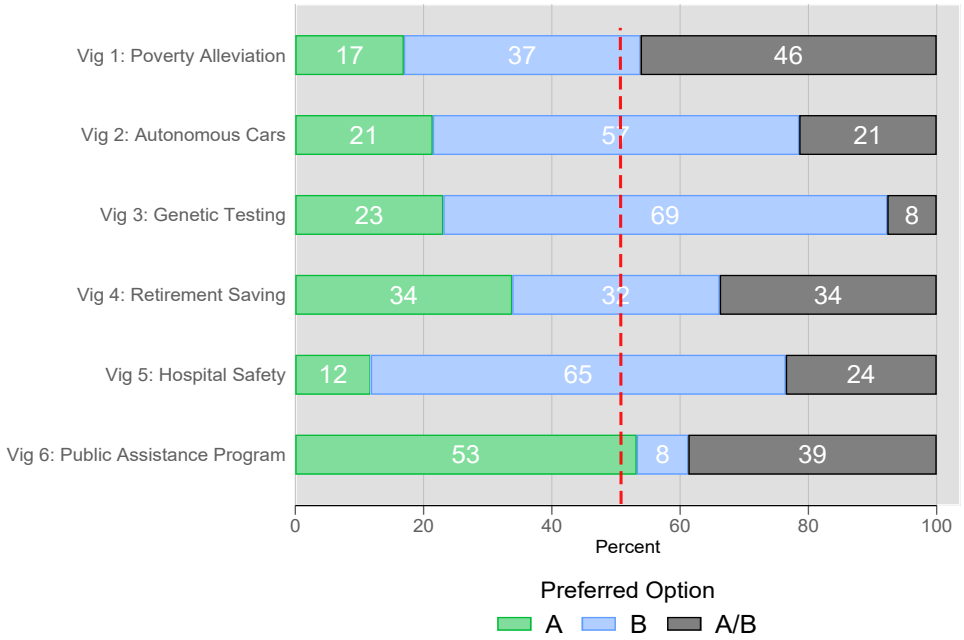


FIGURE 1. THIS FIGURE SUMMARIZES, FOR EACH OF THE FIVE ORIGINAL VIGNETTES (VIG. 1–5) AND THE BRIDGING VIGNETTE (VIG. 6: PUBLIC ASSISTANCE PROGRAM), THE FREQUENCY WITH WHICH PARTICIPANTS IN CONDITION (I) OF STUDY 1 RATED EITHER A UNIFORM POLICY ASSIGNMENT (A OR B) OR THE POLICY EXPERIMENT (A/B) AS MORALLY MOST APPROPRIATE. IN ALL VIGNETTES EXCEPT VIG. 1, ONE OF THE UNIFORM POLICIES WAS JUDGED AS MORALLY MORE APPROPRIATE THAN THE POLICY EXPERIMENT. FURTHERMORE, IN ALL VIGNETTES, THE POLICY EXPERIMENT WAS CHOSEN BY LESS THAN 50 PERCENT OF THE PARTICIPANTS AS THE MOST APPROPRIATE OPTION.

contrast, a uniform implementation of one of the two policies (A or B) would win such a contest in four out of the six vignettes. Similarly, in a voting contest requiring a relative majority, a uniform policy implementation would prevail in five out of the six vignettes, with the A/B test winning in only one (Vig. 1). In sum, condition (i) of Study 1 can be seen as affirming previous findings on experiment aversion: when judged purely on the basis of moral appropriateness, a majority of participants rates the uniform implementation of a single policy as significantly more appropriate than the randomized A/B test of two policies. The results do not change when dropping participants who failed a simple attention check (Supplemental Appendix Figure 1: Panel B).

B. Evidence across studies — from moral support to stated and revealed preferences

Figure 2 displays the stepwise progression from moral approval of policy experimentation to revealed preferences in its favor across two panels, each zooming

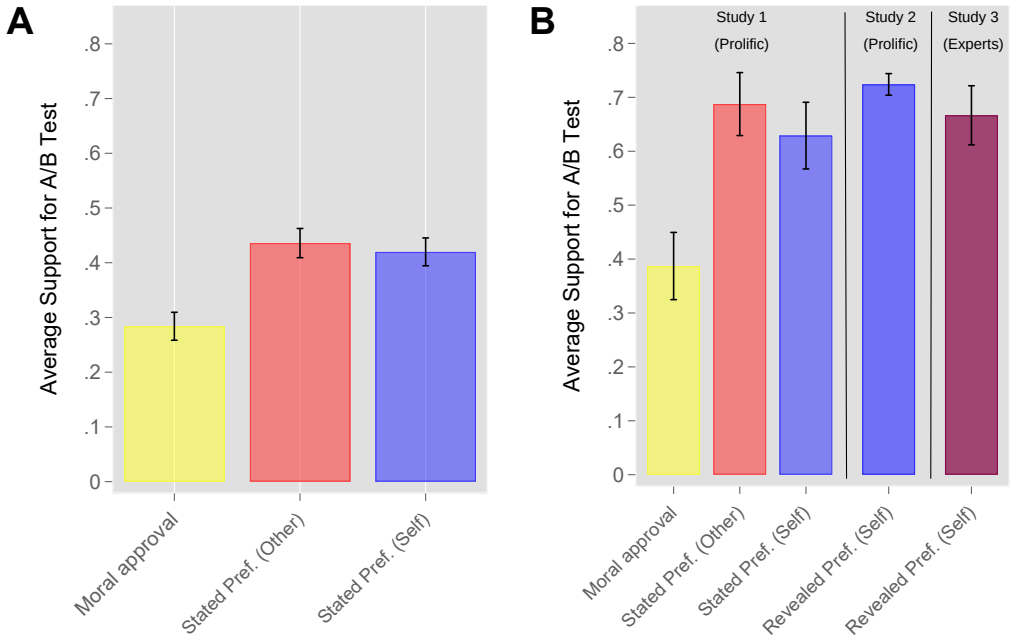


FIGURE 2. **Panel A** CONTAINS STUDY 1 DATA ONLY AND SHOWS SUPPORT FOR THE A/B-TEST IN EACH OF THE THREE CONDITIONS, AVERAGED ACROSS THE SIX VIGNETTES. IT INDICATES HOW AVERAGE SUPPORT INCREASES WHEN MOVING FROM MORAL APPROVAL (I) TO STATED PREFERENCES (OTHER) (II) AND (III) STATED PREFERENCES (SELF). **Panel B** INCLUDES DATA FROM THE THREE STUDIES COMBINED. THE FIRST THREE BARS SHOW SUPPORT FOR THE A/B-TEST IN VIG. 6 OF STUDY 1, INDICATING THAT SUPPORT AGAIN RISES SIGNIFICANTLY WHEN MOVING FROM MORAL APPROVAL (I) TO STATED PREFERENCES ((II) AND (III)). THE REMAINING TWO BARS SHOW THAT REVEALED SUPPORT FOR EXPERIMENTATION IS AS HIGH OR EVEN HIGHER THAN THE STATED SUPPORT FOR EXPERIMENTATION IN STUDY 1. THIS IS TRUE FOR BOTH THE GENERAL PUBLIC (STUDY 2) AND POLICY EXPERTS (STUDY 3).

in on the different studies and conditions within these studies. Panel A compares moral approval and stated preferences across the three conditions of Study 1. Panel B compares moral approval and stated preferences from Study 1 with revealed preferences from Studies 2 and 3, all within the same policy context of the public assistance program.

Panel A of Figure 2 reports for all three conditions of Study 1 the share of individuals who expressed moral approval of or a stated preference for the policy experiment (A/B), now aggregated across all vignettes. For condition (i) of Study 1, the leftmost bar shows the 28.4 percent share of participants who support A/B testing as morally most appropriate. The center bar reports the share of

participants with a stated preference in favor of the A/B test in condition (ii). Here, participants state which option they would prefer another decision-maker to implement. This changes the measurement concept from one based on moral approval to one of stated preferences, with the action to be taken by someone other than the participant. In this condition, the stated preference in favor of A/B testing is 43.6 percent. The difference in the share of participants from condition (i) to condition (ii) is 15.2 percentage points and is highly significant (Ranksum Test (199 vs 195), z-Score = -4.210, $p < 0.001$). As shown in the Supplemental Appendix (SA, Figure 2), this difference is driven by a substantial increase in support for A/B-testing in five of the six vignettes, with Vig. 1 being the only case of no increase. Finally, the rightmost bar in Panel A reports the share of participants preferring the A/B-test in condition (iii). In this second stated preference elicitation, participants are asked to declare their preferences when picturing themselves assuming the role of the decision-maker. Here, the A/B-test is the most preferred course of action for 42 percent of participants. The difference to condition (i) is 13.6 percentage points and is again highly significant (Ranksum test (199 vs. 181), z score = -4.077, $p < 0.001$). It is driven by an increase in support for the A/B-test in five of the six vignettes, with Vig. 1 again being the exception (SA, Fig. 2). The difference between conditions (ii) and (iii) is relatively small (1.6 percentage points) and not statistically significant (Ranksum Test (195 vs 181), z-Score = 0.326, $p = 0.7446$). This reflects small variations in support for A/B-testing in the six vignettes, with the largest difference in Vig. 3. Disaggregated results for all vignettes are available in SA Figure 2 and Table 2.

We next examined whether participant and vignette characteristics can explain variations in the support for A/B tests. For this purpose, we run a Logit model, with A/B support as the dependent variable (SA Table 3). The resulting evidence across conditions and vignettes in Study 1 shows that individual participant characteristics – such as gender, voting intentions, educational attainment, and scientific reasoning score – do not significantly explain support for A/B testing. Instead, variation in support is more closely related to vignette characteristics: the less morally charged participants rate the policy domain of the vignette, the higher the support for A/B testing (Predicted OR: 1.46 [1.18, 1.80], $p < 0.001$). In contrast, the higher they perceive existing knowledge about the policy impacts to be, the lower the support for A/B testing (Predicted OR: 0.709 [0.596, 0.842], $p < 0.001$). This suggests that, in line with standard theories on decision making under uncertainty (Ellsberg, 1961), potential learning gains may intuitively drive a preference for experimentation, which subjects readily trade-off against concerns that are related to the overall moral repugnance of the policy domain (Roth, 2007). Surprisingly, however, support does not systematically relate to the perceived impact of the policy on the target population (Predicted OR: 1.072 [0.913, 1.259], $p = 0.396$). After controlling for these three vignette characteristics, support for policy experimentation remains significantly influenced by idiosyncratic factors specific to each vignette as shown by significant

vignette fixed effects that indicate level differences between vignettes also visible in Figure 1.

These results align with the observation that participants' support for A/B testing varies significantly at the individual level for the two vignettes each participant encountered. In particular, under conditions (ii) and (iii), which are designed to elicit a stated preference for experimentation, a majority of subjects (43 percent and 51 percent) support experiments in one of the two vignettes, rather than consistently supporting or rejecting experiments in both. In contrast, under condition (i), where participants are asked to express their moral approval of experiments, a majority (57 percent) disapproves of experiments in both cases, while only a minority (30 percent) approves of experiments in one vignette but not in the other. This strongly suggests that conditions (ii) and (iii) make the trade-off between moral aversion to experimentation and the perceived usefulness of experiments that underlie the stated preferences more salient. Importantly, this appears to prompt participants to consider the factors relevant to this trade-off and making a decision on a case-by-case basis. Such variation in behavior across vignettes would not be expected if moral objections against randomly assigning subjects to different policies were universally overriding. In sum, experiment aversion does not appear to proxy well for stated preferences and instead tends to underestimate support for policy experimentation.

Panel B of Figure 2 combines evidence from Study 1, Study 2, and Study 3 for the public assistance context of Vig. 6. This scenario remains constant between all three studies. The three leftmost bars report the share of participants who preferred the A/B test under conditions (i), (ii), and (iii) in Study 1, respectively. The pattern observed is consistent with the aggregate patterns seen in Panel A: On moral grounds, experimentation does not gain majority support (condition i). When instead eliciting stated preferences for experimentation, there is a substantial and statistically significant increase in support for the A/B-test (by 30 percentage points ($p = 0.001$) and by 24 percentage points ($p = 0.012$), respectively) when moving – in a hypothetical scenario – from moral approval to stated preferences regarding another decision-maker's action (condition ii) or the participant's own action (condition iii). The difference in support for the A/B-test being enacted either by another decision-maker or oneself (6 percentage points) is statistically insignificant. Moving to the right, the adjacent bar reports revealed preferences for actual A/B testing in Study 2. It reports the share of participants who decided to collect data in the evaluation trial on the public assistance program in the form of an A/B-test by randomly assigning the 200 low-income households to either policy A or B. In this consequential setting, the revealed preference for A/B testing is 72 percent, even in the presence of a status quo option not available to participants in Study 1. The difference to the moral approval expressed in condition (i) of the hypothetical setting of Study 1 is 33 percentage points ($p < 0.001$). The differences to the stated preferences (other's action) elicited in condition (ii) ($p = 0.415$) and the stated preferences

(own action) in condition (iii) of Study 1 ($p = 0.192$) are insignificant.

The rightmost bar in Panel B reports revealed preferences for A/B-testing in Study 3. Support by the policy-makers for A/B-testing on 200 low-income households is 67 percent, again in the presence of a status-quo alternative as in Study 2. The difference in support between participants drawn from the general population in Study 2 and from policy experts in Study 3 is 5 percentage points and not statistically significant. The difference with moral approval elicited in condition (i) of Study 1 is at 29 percentage points significant ($p < 0.001$).[2]

Econometric analysis of the evidence adds further support to the findings above (see the SA for details). A first regression confirms our non-parametric results from above: Support for experimentation in condition (i) significantly differs from that in conditions (ii) and (iii) in Study 1 (see Table 3 in SA). This validates the simple comparison tests showing a significant gap between moral judgments and stated preferences. For Studies 2 and 3, we similarly use a regression-based test to investigate demographics and attitudes towards the assistance program as correlates of revealed preferences for experimentation (see Table 6 in SA). In line with the stated preference results, revealed preferences for policy experimentation are unrelated to standard demographics (gender, income, education). Within the same regression-based test we focus on revealed preferences for experimentation under the two treatment conditions, which vary the probability that the decision is pivotal (*consequences*) and the probability that participants are re-invited to the follow-up for a paid prediction task (*stakes*). Once consequential, the scale of *Stakes* and *Consequences* has an insubstantial impact on support for A/B-testing, both in separate t-tests (p = 0.148 and p = 0.172 respectively), as well as the interaction of both treatments (p = 0.976) and in a joint F-Test (p = 0.252, see text below Table 6 SA).

Discussion

The empirical evidence presented across the three studies that form the core of this paper provides new insights on the public's stance vis-à-vis policy experiments and allows to reinterpret existing evidence in a new light. We provide the first revealed preference evidence on policy experimentation, comparing it directly with standard measures of moral approval and stated preferences within a uniform framework and setting. The new insights are enabled by comparing two previously separate conceptual approaches, moral judgments and stated preferences, and contrasting it with revealed preferences in a consequential setting.

We find that measuring experiment aversion leads to a substantial and systemic underestimation of revealed preferences for policy experimentation. Our results

[2] Analogous to the sample in Study 2, the comparison of preferences for experimentation in conditions (ii) and (iii) of Study 1 is insignificant ($p = 0.902$ and $p = 0.801$, respectively). Moreover, aggregating preferences for experimentation over Studies 2 and 3, the comparison with moral approval in condition (i) of Study 1 is highly significant ($p < 0.001$), while the comparison with conditions (ii) and (iii) is insignificant ($p = 0.491$ and $p = 0.237$, respectively).

demonstrate a substantial gap between the different approaches measuring moral support for policy experiments and those measuring revealed preferences. In the policy context of our study, the size of the gap is around 30 percentage points, of which back-of-the envelope calculations attribute roughly two thirds to conceptual differences and one third to methodological factors such as hypothetical bias.

Our findings on the moral approval of experimentation, elicited under condition (i) of Study 1, are consistent with previous evidence indicating that a majority of individuals exhibit moral aversion to experiments across a diverse range of policy domains. This pattern extends to the results from the newly introduced vignette situated in the public assistance program context. These earlier findings are robust to our deliberate switch from an elicitation format asking participants to rate each option individually to one where they select the most appropriate option. The consistency of results across these two distinct elicitation formats suggests some generalizability of moral opposition to policy experimentation (Mazar, Elbaek and Mitkidis, 2023; Vogt et al., 2024).

At the same time, our results challenge the interpretation of moral views regarding policy experimentation as informative of the general public's preferences about policy experimentation. General population subjects in our studies state considerably more support for policy experimentation when the choice setting elicits preferences, even in the hypothetical contexts of Study 1. This increase in support for A/B-testing holds for five out of six vignettes. More importantly, moving from hypothetical stated preferences to consequential revealed preferences further increases or maintains this high level of support for experimentation, both among the general public and policy experts. Thus the gap between low moral approval and high revealed preference for experimentation appears driven primarily by conceptual misalignment (moral judgment vs. preference elicitation) rather than hypothetical bias alone. Conversely, stated preference elicitation for policy experimentation (Dur et al., 2024) would appear to incur only limited hypothetical bias, at least not in policy areas comparable to the one studied here. Also worth noting in this context is how the share of general population subjects whose stated preference is for a third party (such as a policy-maker) to be choosing the policy experiment compares with the share of policy-makers who actually reveal their preference for policy experimentation. These shares both come in around 70 percent and are therefore very similar.

Together, our findings suggest that experiment aversion can, but often need not be an impediment to policy experimentation. This leads to at least four practical implications for those wishing to promote experiments as the gold standard in evidence-based policy-making. For a start, our results at the level of individual vignettes suggest that detecting high levels of public experimentation aversion to a concrete policy experiment remains a valuable step (Meyer et al., 2019; Bas, Vosgerau and Ciulli, 2023; Vogt et al., 2024). The reason is that some policy areas appear inherently morally fraught. Detecting aversion early will help anticipate, and possibly preempt, public backlash against learning from experimentation in

public policy by identifying domains characterized by strong moral repugnance (Roth, 2007; Carattini, Dur and List, 2024).

Second, our findings suggest that, in order to fully gauge the public's views on how policy-makers ought to navigate the trade-offs implicit in policy experimentation, research should adopt a stated preference approach. Even relatively cheap and easy-to-administer hypothetical surveys do not suffer from larger hypothetical bias resulting in estimates of public support to experimentation that are reasonably close to revealed preferences for experimentation.

Third, for policy makers who wish to increase public support for a concrete policy experiment, it may be important to make salient the informational benefits from experimentation to "mitigate perceived experiment aversion" (Vogt et al., 2024). In our concrete example, this framing raised support for A/B-testing of two alternative policies with imperfectly understood impacts on a public assistance program helping low-income households from 39 percent of the participants approving on moral grounds to 69 percent approving of experimentation as what a policy-maker ought to do.

Finally, when examining why policy experimentation is not more common, it is essential to consider a wider range of explanations. Most real-world policies are still implemented without prior experimentation, even in a morally non-repugnant policy context in which support for policy experiments is likely to be large, such as the public assistance programs examined as part of this paper. We suspect that polic-ymakers often encounter additional obstacles that inhibit their ability or willingness to conduct more experiments. Three likely contributing factors include financial constraints and organizational inertia, legal mandates to treat citizens equally, and a reluctance to publicly acknowledge a lack of knowledge on which policy is likely to succeed. A better understanding of the relative contribution of each of these factors to a reluctance to conduct experiments for better policy design would be an important next research step.

REFERENCES

Asiedu, Elizabeth, Dean Karlan, Monica Lambon-Quayefio, and Christopher Udry. 2021. "A call for structured ethics appendices in social science papers." *Proceedings of the National Academy of Sciences*, 118(29): e2024570118.

Banerjee, Abhijit V., and Esther Duflo. 2009. "The Experimental Approach to Development Economics." *Annual Review of Economics*, 1(1): 151–178.

Bartling, Björn, Roberto A. Weber, and Lan Yao. 2015. "Do Markets Erode Social Responsibility?" *The Quarterly Journal of Economics*, 130(1): 219–266.

Bas, Burcak, Joachim Vosgerau, and Rachele Ciulli. 2023. "No Evidence That Experiment Aversion Is Not a Robust Empirical Phenomenon." *Proceedings of the National Academy of Sciences*, 120(50): e2317514120.

Black, Duncan. 1958. *The Theory of Committees and Elections.* Springer.

Carattini, Stefano, Robert Dur, and John List. 2024. "Policy evaluation and the causal analysis of public support." *Science*, 386(6721): 490–492.

Charness, Gary, Anya Samek, and Jeroen Van de Ven. 2022. "What is considered deception in experimental economics?" *Experimental Economics*, 25(2): 385–412.

Chetty, Raj, Nathaniel Hendren, and Lawrence F. Katz. 2016. "The Effects of Exposure to Better Neighborhoods on Children: New Evidence From the Moving to Opportunity Experiment." *American Economic Review*, 106(4): 855–902.

Corduneanu-Huci, Cristina, Michael T. Dorsch, and Paul Maarek. 2021. "The politics of experimentation: Political competition and randomized controlled trials." *Journal of Comparative Economics*, 49(1): 1–21.

Deacon, Robert, and Perry Shapiro. 1975. "Private preference for collective goods revealed through voting on referenda." *The American Economic Review*, 65(5): 943–955.

De Corte, Kaat, John Cairns, and Richard Grieve. 2021. "Stated Versus Revealed Preferences: An Approach to Reduce Bias." *Health Economics*, 30(5): 1095–1123.

DellaVigna, Stefano, Woong Kim, and Elizabeth Linos. 2024. "Bottlenecks for evidence adoption." *Journal of Political Economy*, 132(8): 2748–2789.

Dur, Robert, Arjan Non, Paul Prottung, and Benedetta Ricci. 2024. "Who's Afraid of Policy Experiments?" *The Economic Journal.*

Egger, Dennis, Johannes Haushofer, Edward Miguel, Paul Niehaus, and Michael Walker. 2022. "General Equilibrium Effects of Cash Transfers: Experimental Evidence From Kenya." *Econometrica*, 90(6): 2603–2643.

Ellsberg, Daniel. 1961. "Risk, ambiguity, and the savage axioms." *The Quarterly Journal of Economics*, 75(4): 643–669.

Falk, Armin, and Nora Szech. 2013. "Morals and Markets." *Science*, 340(6133): 707–711.

Finkelstein, Amy, Sarah Taubman, Bill Wright, Mira Bernstein, Jonathan Gruber, Joseph P. Newhouse, Heidi Allen, Katherine Baicker, and Oregon Health Study Group. 2012. "The Oregon Health Insurance Experiment: Evidence From the First Year." *The Quarterly Journal of Economics*, 127(3): 1057–1106.

Fischer, Mirjam, Elisabeth Grewenig, Philipp Lergetporer, Katharina Werner, and Hannah Zeidler. 2024. "The e-word—on the public acceptance of experiments." *Economics Letters*, 235: 111558.

Glennerster, Rachel. 2017. "The practicalities of running randomized evaluations: Partnerships, measurement, ethics, and transparency." In *Handbook of Economic Field Experiments*. Vol. 1, , ed. Abhijit V. Banerjee and Esther Duflo, 175–243. Amsterdam:North-Holland.

Haushofer, Johannes, and C. Jessica E. Metcalf. 2020. "Which Interventions Work Best in a Pandemic?" *Science*, 368(6495): 1063–1065.

Hausman, Jerry. 2012. "Contingent Valuation: From Dubious to Hopeless." *Journal of Economic Perspectives*, 26(4): 43–56.

Heck, Patrick R., Christopher F. Chabris, Duncan J. Watts, and Michelle N. Meyer. 2020. "Objecting to Experiments Even While Approving of the Policies or Treatments They Compare." *Proceedings of the National Academy of Sciences*, 117(32): 18948–18950.

Heffetz, Ori, and John List. 2021. "Who's Afraid of Evidence-Based Policymaking?" The Field Experiments Website Artefactual Field Experiments 00739.

Hjort, Jonas, Diego Moreira, Gautam Rao, and Juan Francisco Santini. 2021. "How research affects policy: Experimental evidence from 2,150 Brazilian municipalities." *American Economic Review*, 111(5): 1442–1480.

Kirchler, Michael, Jürgen Huber, Matthias Stefan, and Matthias Sutter. 2016. "Market Design and Moral Behavior." *Management Science*, 62(9): 2615–2625.

Leuker, Christina, Lasare Samartzidis, and Ralph Hertwig. 2021. "What makes a market transaction morally repugnant?" *Cognition*, 212: 104644.

List, John A. 2008. "Informed consent in social science." *Science*, 322(5902): 672.

List, John A, and Craig A Gallet. 2001. "What experimental protocol influence disparities between actual and hypothetical stated values?" *Environmental and resource economics*, 20: 241–254.

Lynch, H. Fernandez, D. J. Greiner, and I. G. Cohen. 2020. "Overcoming Obstacles to Experiments in Legal Practice." *Science*, 367(6482): 1078–1080.

Mazar, Nina, Christian T. Elbaek, and Panagiotis Mitkidis. 2023. "Experiment Aversion Does Not Appear to Generalize." *Proceedings of the National Academy of Sciences*, 120(16): e2217551120.

McDermott, Rose, and Peter K. Hatemi. 2020. "Ethics in field experimentation: A call to establish new standards to protect the public from unwanted manipulation and real harms." *Proceedings of the National Academy of Sciences*, 117(48): 30014–30021.

Mehmood, Saad, Shanza Naseer, and Daniel L. Chen. 2021. "Training policymakers in econometrics." *Working paper, Toulouse School of Economics.*

Meyer, Michelle N., Patrick R. Heck, Geoffrey S. Holtzman, Stephen M. Anderson, William Cai, Duncan J. Watts, and Christopher F. Chabris. 2019. "Objecting to Experiments That Compare Two Unobjectionable Policies or Treatments." *Proceedings of the National Academy of Sciences*, 116(22): 10723–10728.

Mislavsky, Robert, Berkeley Dietvorst, and Uri Simonsohn. 2020. "Critical Condition: People Don't Dislike a Corporate Experiment More Than They Dislike Its Worst Condition." *Marketing Science*, 39(6): 1092–1104.

Rosenthal, Robert, and Ralph L. Rosnow. 2008. *Essentials of Behavioral Research: Methods and Data Analysis.* McGraw-Hill.

Roth, Alvin E. 2007. "Repugnance as a Constraint on Markets." *Journal of Economic perspectives*, 21(3): 37–58.

Sharma, Eesha, Nina Mazar, Adam L. Alter, and Dan Ariely. 2014. "Financial Deprivation Selectively Shifts Moral Standards and Compromises Moral Decisions." *Organizational Behavior and Human Decision Processes*, 123(2): 90–100.

Venkataramani, Atheendar S., Kristen Underhill, and Kevin G. Volpp. 2020. "Moving Toward Evidence-Based Policy: The Value of Randomization for Program and Policy Implementation." *JAMA*, 323(1): 21–22.

Vivalt, Eva, and Aidan Coville. 2023. "How do policymakers update their beliefs?" *Journal of Development Economics*, 165: 103121.

Vogt, Randi L., Patrick R. Heck, Duncan J. Watts, Christopher F. Chabris, and Michelle N. Meyer. 2024. "Experiment Aversion Does Generalize, But It Can Also Be Mitigated." *Proceedings of the National Academy of Sciences*, 121(20): e2315439121.

Vossler, Christian A, and Joe Kerkvliet. 2003. "A criterion validity test of the contingent valuation method: comparing hypothetical and actual voting behavior for a public referendum." *Journal of Environmental Economics and Management*, 45(3): 631–649.

Vossler, Christian A, Maurice Doyon, and Daniel Rondeau. 2012. "Truth in consequentiality: theory and field evidence on discrete choice experiments." *American Economic Journal: Microeconomics*, 4(4): 145–171.

Wang, Shaoda, and David Yang. 2024. "Policy experimentation in China: The political economy of policy learning." *Journal of Political Economy*.

Wuepper, David, Alexandra Clemm, and Philipp Wree. 2019. "The Preference for Sustainable Coffee and a New Approach for Dealing With Hypothetical Bias." *Journal of Economic Behavior & Organization*, 158: 475–486.

Supplemental Appendix for

Preferring policy experiments despite moral aversion

Bettina Chlond[1,2] Timo Goeschl[1,2] Johannes Lohse[3,4]

Contents

[1] Heidelberg University
[2] ZEW-Leibniz Centre for European Economic Research Mannheim
[3] Institute for Economics, Leuphana University Lüneburg
[4] Department of Economics, University of Birmingham, Birmingham

Methods

Overview

In total, we conducted three studies. Participants for Studies 1 and 2 were recruited from the German-speaking pool of Prolific.ac, while participants for Study 3 were recruited through a targeted email campaign. The expert sample in Study 3 comprised two groups: experts from municipal social welfare offices and those working in local branches of the NGO administering the assistance program. The former were recruited via a systematic collection of municipal decision-makers' email addresses, while the latter were invited by federal-level NGO managers who contacted the local branch managers. All participants provided informed consent and received a standard participation fee, in addition to any experimental earnings. Participants in each study were compensated for their time, and those in Studies 2 and 3 were further incentivized to provide accurate estimates of the efficacy of the two policy alternatives tested in the field trial on public assistance programs.

Complete demographic information for each study is available in the results section of this supplement.

All three studies were conducted online via the Qualtrics survey platform, and randomizations were implemented using randomly drawn variables in Qualtrics. In Study 1, participants were randomly assigned to one of three treatment conditions (between-subjects) and encountered two random vignettes (from a set of six). In Studies 2 and 3, participants were randomly assigned to one of four treatment conditions in a 2x2 full factorial design, varying by Impact and Stakes. All main analyses of Studies 2 and 3 were pre-registered via AsPredicted (https://aspredicted.org/54xs-mhpn.pdf).

Studies

In total, we conducted three studies (n = 1,175).

Study 1 included 600 participants, with approximately 200 participants per condition. The three conditions varied in how the question was framed to elicit (i) moral approval of the experiments, (ii) a stated preference for the optimal actions of a third party, and (iii) a participants' own hypothetical actions. Each participant was presented with two vignettes and asked to indicate their support for an A/B test versus the uninformed implementation of either policy A or policy B. The primary comparison of interest was the level of support for the A/B test across the different framing conditions.

Using G*Power, we determined that t-tests with $\alpha = .05$ and $\beta = .20$ would detect standardized effect sizes of $d \geq 0.30$ for differences between the two experimental conditions.

Participants for this study were recruited via Prolific.ac, targeting German native speakers with a high Prolific rating (>95). They earned a flat fee of €1.50 for a study lasting approximately 15 minutes. The study included attention checks regarding the vignette contents to ensure data quality. Only 2.4 % of participants failed the attention check and this number was comparable across the different vignettes. After receiving general instructions, participants were presented with two randomly selected vignettes derived from the literature, along with one new vignette related to a public assistance program to match the context of Studies 2 and 3.

The full text of all vignettes is provided in the materials section below.

Each vignette described two policy options (A and B) as well as the possibility of conducting an A/B test. For each vignette they encountered, participants in condition (i) chose the morally most appropriate option (*Moral approval*), in condition (ii) they selected the most appropriate option to be taken by a third party (*Stated Preference Other*), and in condition (iii) they chose their preferred course of action if they were in the position of the decision-maker (*Stated Preference Self*).

For each vignette, we then elicited (a) how morally contentious participants found the setting and (b) whether they believed there was already substantial knowledge about the policy context. On the following screen, we collected standard demographic information (gender, age, education), along with participants' voting intentions and scientific reasoning abilities.

Study 2 involved 500 participants and Study 3 involved 75 participants. In both studies, subjects were asked about their preferences for experiments, participating in a referendum on their preferred policy to be implemented for an additional 200 households eligible for the assistance program, choosing between (I) the status quo, (II) Policy A only, (III) Policy B only, or (IV) a random allocation of Policies A and B. The stakes and impact of their vote varied in a 2x2 full factorial design. Participants receiving the low (high) stakes treatment were informed that the probability for invitation to a second round with a prediction task on household participation rates in the assistance program in Policies A and B is low (high) at 10% (90%). Participants receiving the low (high) impact treatment were informed that their decision had a low (high) weight of 1 (10) vote(s) in the referendum to determine which of the options (I) to (IV) would be implemented.

Participants for Study 2 were recruited via Prolific.ac, targeting German native speakers with a high Prolific rating (>95). They earned a flat fee of €2.50 for a study lasting approximately 15 minutes and had the opportunity to gain a bonus in a prediction task.

Participants for Study 3 were recruited via two channels: (i) in a systematic online search, a list of emails by heads of social welfare offices and climate protection agencies considering all German municipalities was created, and (ii) the assistance program's federal managers sent out study invitations to all managers of the NGO's local program branches. They earned a flat fee of €5.00 for a study lasting approximately 15 minutes and had the opportunity to gain a bonus in a prediction task. The higher fee for this sample represents its higher opportunity cost of time.

After receiving general instructions, participants were informed about the baseline participation rate in the program of interest (14/100 households) and were asked for their incentivized prediction on how each of the Policy options A and B fared in a pilot study. If they guessed a participation rate in the range of 3 percentage points around the true value for a randomly determined Policy option, they gained a bonus of €1.5.After, they were informed about their stakes and impact on the implementation, and subsequently asked for their vote on which option (I) to (IV) should be implemented.

Materials

Below is the complete set of materials for Studies 1-3.

Study 1

We used a total of **six** different vignettes. Vignettes 1 to 5 were sampled from the existing literature on experimental aversion and translated to German. Here we provide the German originals as well as their English translation. Bridging Vignette 6 provides a new context that mirrors the setting in Studies 2 and 3.

Vignette 1: Poverty Alleviation

German original:

Das Szenario beschreibt, wie sich der Abteilungsleiter in einem Ministerium bei der Verteilung von Sozialhilfe verhalten könnte.

Versuchen Sie, jede Handlungsmöglichkeit getrennt von den anderen zu betrachten. Bitte lesen Sie den gesamten Text auf der nächsten Seite sorgfältig durch und beziehen Sie sich bei der Beantwortung der Fragen auf den Text.

Das Sozialministerium hat die Aufgabe, Sozialhilfe- und Arbeitslosengelder an bedürftige Familien zu verteilen. Der Leiter der zuständigen Abteilung denkt über zwei verschiedene Ideen zur Verringerung von Armut und Arbeitslosigkeit nach.

Eine Idee ist, die Sozialleistungen für die ärmsten Familien, also diejenigen, die am weitesten unter der Armutsgrenze leben, deutlich zu erhöhen.

Eine weitere Idee ist, die Sozialleistungen für alle Familien, die unterhalb der Armutsgrenze leben, moderat zu erhöhen.

Im Folgenden sehen Sie verschiedene Handlungsmöglichkeiten, wie der Abteilungsleiter sich verhalten könnte.

Der Abteilungsleiter möchte die Zahl an Familien in Armut und Arbeitslosigkeit so effektiv wie möglich verringern. Er beschließt daher, ab sofort die Sozialleistungen für die ärmsten Familien, die am weitesten unter der Armutsgrenze leben, deutlich zu erhöhen.

Der Abteilungsleiter möchte die Zahl an Familien in Armut und Arbeitslosigkeit so effektiv wie möglich verringern. Er beschließt daher, ab sofort die Sozialleistungen für alle Familien, die unter der Armutsgrenze leben, moderat zu erhöhen.

Der Abteilungsleiter möchte die Zahl an Familien in Armut und Arbeitslosigkeit so effektiv wie möglich verringern. Er beschließt daher, ein Experiment durchzuführen, bei dem Landkreise nach dem Zufallsprinzip einer von zwei Gruppen zugewiesen werden. In der Hälfte der Landkreise werden die Sozialleistungen für die ärmsten Familien, die am weitesten unter der Armutsgrenze leben, deutlich erhöht werden. In der anderen Hälfte der Landkreise werden die Sozialleistungen für alle Familien, die unterhalb der Armutsgrenze leben, moderat erhöht werden. Nach einem Jahr wird das Sozialministerium für alle Landkreise diejenige Politik übernehmen, die Armut und Arbeitslosigkeit am effektivsten verringert.

English translation:

The scenario describes how the head of department in a ministry might behave when distributing social assistance.

Try to consider each possible action separately from the others. Please read the entire text on the next page carefully and refer to the text when answering the questions.

The Ministry of Social Affairs is responsible for distributing social welfare and unemployment benefits to struggling families. The head of the responsible department is considering two different ideas to reduce poverty and unemployment.

One idea is to significantly increase social benefits for the poorest families, i.e. those living furthest below the poverty line.

Another idea is to moderately increase social benefits for all families living below the poverty line.

Below you can see various possible courses of action that the head of department could take.

The head of department wants to reduce the number of families in poverty and unemployment as effectively as possible. He therefore decides to significantly increase social benefits for the poorest families living furthest below the poverty line with immediate effect.

The head of department wants to reduce the number of families in poverty and unemployment as effectively as possible. He therefore decides to moderately increase social benefits for all families living below the poverty line with immediate effect.

The head of department wants to reduce the number of families in poverty and unemployment as effectively as possible. He therefore decides to conduct an experiment in which counties are randomly assigned to one of two groups. In half of the counties, social benefits will be significantly increased for the poorest families living furthest below the poverty line. In the other half of the districts, social benefits for all families living below the poverty line will be increased moderately. After one year, the Department of Social Services will adopt the policy that most effectively reduces poverty and unemployment for all counties.

Vignette 2: Autonomous Cars

German original:

Das Szenario beschreibt, wie sich der Geschäftsführer eines Unternehmens aus der Automobilindustrie verhalten könnte.

Versuchen Sie, jede Handlungsmöglichkeit getrennt von den anderen zu betrachten. Bitte lesen Sie den gesamten Text auf der nächsten Seite sorgfältig durch und beziehen Sie sich bei der Beantwortung der Fragen auf den Text.

Viele Menschen sind von der Idee selbstfahrender Autos begeistert, die in der Lage sind, ohne Eingriffe eines menschlichen Fahrers zu fahren. Solche Autos können das Leben der Menschen erleichtern und die Zahl der Unfälle verringern. Manchmal kommt es aber vor, dass der Fahrer im Notfall nicht mehr die Kontrolle über sein Auto übernehmen kann. Der Geschäftsführer eines Unternehmens, das selbstfahrende Autos entwickelt, möchte, dass Fahrer so viel Freiheit wie möglich haben, während sie auf der Straße unterwegs sind, aber auch, dass sie sicher fahren. Der Geschäftsführer hat zwei Ideen, wie beide Ziele erreicht werden können.

Eine Idee ist, dass alle Autos des Unternehmens mit einem Hebel ausgestattet werden, mit dem der Fahrer zwischen dem selbstfahrenden und dem von Menschen gesteuerten Modus wechseln kann.

Eine weitere Idee ist, dass jede Betätigung der Bremsen, des Gaspedals oder des Lenkrads durch einen menschlichen Fahrer automatisch den Selbstfahrmodus der Autos des Unternehmens außer Kraft setzt.

Im Folgenden sehen Sie verschiedene Handlungsmöglichkeiten, wie der Geschäftsführer sich verhalten könnte.

Der Geschäftsführer möchte, dass die Menschen so viel Freiheit wie möglich haben, während sie auf der Straße unterwegs sind, aber auch sicher fahren. Daher beschließt er, dass alle Autos des

Unternehmens mit einem Hebel ausgestattet werden, mit dem der Fahrer zwischen dem selbstfahrenden und dem vom Menschen gesteuerten Modus wechseln kann.

Der Geschäftsführer möchte, dass die Menschen so viel Freiheit wie möglich haben, während sie auf der Straße unterwegs sind, aber auch sicher fahren. Daher beschließt er, dass jede Betätigung der Bremsen, des Gaspedals oder des Lenkrads durch einen menschlichen Fahrer automatisch den Selbstfahrmodus der Autos des Unternehmens außer Kraft setzt.

Der Geschäftsführer möchte, dass die Menschen so viel Freiheit wie möglich haben, während sie auf der Straße unterwegs sind, aber auch sicher fahren. Er beschließt daher, ein Experiment durchzuführen, bei dem Autos nach dem Zufallsprinzip einer von zwei Gruppen zugewiesen werden. Die Hälfte der Autos, die das Unternehmen verkauft, wird mit einem Hebel ausgestattet, der es dem Fahrer ermöglicht, zwischen dem selbstfahrenden und dem vom Menschen gesteuerten Modus zu wechseln. Die andere Hälfte wird so programmiert, dass jede Betätigung der Bremsen, des Gaspedals oder des Lenkrads durch einen menschlichen Fahrer automatisch den Selbstfahrmodus der Fahrzeuge des Unternehmens außer Kraft setzt. Nach einem Jahr wird der Geschäftsführer alle Fahrzeuge des Unternehmens nach dem Konzept bauen lassen, das sich als am sichersten erweist.

English translation:

The scenario describes how the managing director of a company in the automotive industry might behave.

Try to consider each possible action separately from the others. Please read the entire text on the next page carefully and refer to the text when answering the questions.

Many people are excited by the idea of completely self-driving cars that are capable of navigating the road without any input from a human driver. These kinds of cars can make people's lives easier and reduce the number of accidents. Sometimes, however, some of them prevent the driver from taking control of their car in the case of an emergency. The CEO of a company developing self-driving cars wants drivers to have as much freedom as possible while on the road, but also wants them to drive safely. The CEO has two ideas on how to achieve both goals.

One idea is that all of the company's cars will have a lever that allows drivers to switch between self-driving and human-driven modes.

Another idea is that any use of the brakes, gas pedal or steering wheel by a human driver will automatically override the self-driving mode of the company's cars.

Below you can see various possible actions that the managing director could take.

The CEO wants people to have as much freedom as possible while travelling on the road, but also to drive safely. Therefore, he decides that all of the company's cars will have a lever that allows the driver to switch between self-driving and human-driven mode.

The CEO wants people to have as much freedom as possible while travelling on the road, but also to drive safely. Therefore, he decides that any use of the brakes, gas pedal or steering wheel by a human driver will automatically override the self-driving mode of the company's cars.

The CEO wants people to have as much freedom as possible while travelling on the road, but also to drive safely. He therefore decides to conduct an experiment in which cars are randomly assigned to one of two groups. Half of the cars the company sells will have a lever that allows the driver to switch between self-driving and human-driven mode. The other half will be programmed so that any use of the brakes, gas pedal or steering wheel by a human driver will automatically override the self-driving mode of the company's vehicles. After one year, the managing director will have all the company's vehicles built according to the concept that proves to be the safest.

Vignette 3: Genetic Testing

German original

Das Szenario beschreibt, wie sich der Geschäftsführer eines Unternehmens, das Gentests anbietet, verhalten könnte.

Versuchen Sie, jede Handlungsmöglichkeit getrennt von den anderen zu betrachten. Bitte lesen Sie den gesamten Text auf der nächsten Seite sorgfältig durch und beziehen Sie sich bei der Beantwortung der Fragen auf den Text.

Genetische Mutationen können zu gesundheitlichen Problemen führen. Manche sind so schwerwiegend, dass sie einen Menschen krank machen oder sogar zum Tod führen können. Viele dieser Erkrankungen können aber verhindert oder zumindest verlangsamt werden. Voraussetzung ist, dass bestimmte Maßnahmen ergriffen werden, sobald ein Befund vorliegt. Denken wir an eine Firma, die Gentests für die Ahnenforschung anbietet. Damit erhalten ihre Kunden Informationen zu ihrer Abstammung. Der Geschäftsführer der Firma möchte Menschen so viel wie möglich helfen. Er beschließt, allen Kunden, die einen Gentest machen, auch die Möglichkeit eines genetischen Befunds anzubieten. Dieser Befund informiert über ein erhöhtes genetisches Risiko für heilbare oder behandlungsfähige gesundheitliche Probleme. Der Geschäftsführer hat zwei Ideen, wie er dieses Angebot umsetzen könnte.

Eine Idee ist, die Möglichkeit eines genetischen Befunds nur dann anzubieten, wenn die Kunden ein erhöhtes genetisches Risiko für gesundheitliche Probleme haben, die auch verhindert oder abgemildert werden können. Die Kunden haben die Möglichkeit, diese Ergebnisse einzusehen oder nicht.

Eine weitere Idee ist, die Möglichkeit eines genetischen Befund immer anzubieten, wenn die Kunden ein erhöhtes genetisches Risiko für gesundheitliche Probleme haben und im Befund zu informieren, ob die Krankheit verhindert oder abgemildert werden kann. Die Kunden haben die Möglichkeit, diese Ergebnisse einzusehen oder nicht.

Im Folgenden sehen Sie verschiedene Handlungsmöglichkeiten, wie der Geschäftsführer sich verhalten könnte.

Der Geschäftsführer möchte Menschen so viel wie möglich helfen. Daher beschließt er, allen seinen Kunden die Möglichkeit eines genetischen Befunds nur dann anzubieten, wenn die Kunden ein erhöhtes Risiko für gesundheitliche Probleme haben, die verhindert oder abgemildert werden können. Die Kunden haben die Möglichkeit, diese Ergebnisse einzusehen oder nicht.

Der Geschäftsführer möchte Menschen so viel wie möglich helfen. Daher beschließt er, allen seinen Kunden die Möglichkeit eines genetischen Befunds anzubieten und im Befund zu informieren, ob die Krankheit verhindert oder abgemildert werden kann. Die Kunden haben die Möglichkeit, diese Ergebnisse einzusehen oder nicht.

Der Geschäftsführer möchte Menschen so viel wie möglich helfen. Daher beschließt er, ein Experiment durchzuführen, bei dem die Kunden des Unternehmens nach dem Zufallsprinzip einer von zwei Gruppen zugewiesen werden. Der einen Hälfte seiner Kunden bietet er die Möglichkeit eines genetischen Befunds nur dann an, wenn die Kunden ein erhöhtes Risiko für solche gesundheitlichen Probleme haben, die verhindert oder abgemildert werden können. Der anderen Hälfte bietet er die Möglichkeit eines genetischen Befunds an und informiert im Befund, ob die Krankheit verhindert oder abgemildert werden kann. Die Kunden beider Gruppen werden die Möglichkeit haben, die ihnen angebotenen zusätzlichen Ergebnisse einzusehen oder nicht. Nach einem Jahr wird der Geschäftsführer allen neuen Kunden die Option anbieten, die zur höchsten Kundenzufriedenheit führt.

English translation

The scenario describes how the managing director of a company that offers genetic testing might behave.

Try to consider each possible action separately from the others. Please read the entire text on the next page carefully and refer to the text when answering the questions.

Some genetic mutations can lead to health conditions. Some are so serious that they can make a person ill or even cause them to die. However, many of these diseases can be prevented or at least slowed down. The prerequisite is that certain measures are taken as soon as a finding is made. Consider a company that offers genetic tests for genealogical research. This provides its customers with information about their ancestry. The managing director of the company wants to help people as much as possible. He decides to offer all customers who take a genetic test the option of a genetic report. This report provides information about an increased genetic risk for curable or treatable health conditions. The managing director has two ideas on how he could realize this offer.

One idea is to only offer the option of a genetic test if customers have an increased genetic risk of health conditions that can be cured or treated. Customers have the option of viewing these results or not.

Another idea is to always offer the option of a genetic test if customers have an increased genetic risk of health conditions and to provide information in the test as to whether the disease can be cured or treated. Customers have the option of viewing these results or not.

Below you can see various possible courses of action that the manager could take.

The managing director wants to help people as much as possible. He therefore decides to offer all his customers the option of a genetic test only if the customers have an increased risk of health conditions that can be cured or treated. Customers have the option to view these results or not.

The managing director wants to help people as much as possible. He therefore decides to offer all his customers the option of a genetic test and to provide information in the results as to whether the condition can be cured or treated. Customers have the option of viewing these results or not.

The managing director wants to help people as much as possible. He therefore decides to conduct an experiment in which the company's customers are randomly assigned to one of two groups. He offers one half of his customers the option of a genetic test only if they have an increased risk of health conditions that can be cured or treated. The other half will be offered the option of a genetic test and will be informed in the results whether the condition can be cured or treated. Customers in both groups will have the option of viewing the additional results offered to them or not. After one year, the manager will offer all new customers the option that leads to the highest customer satisfaction.

Vignette 4: Retirement Savings

German original:

Das Szenario beschreibt, wie sich der Geschäftsführer eines Unternehmens bezüglich der betrieblichen Altersvorsorge für seine Mitarbeiter verhalten könnte.

Versuchen Sie, jede Handlungsmöglichkeit getrennt von den anderen zu betrachten. Bitte lesen Sie den gesamten Text auf der nächsten Seite sorgfältig durch und beziehen Sie sich bei der Beantwortung der Fragen auf den Text.

Die meisten Menschen sparen zu wenig für den Ruhestand. Dies kann später im Leben zu finanziellen Problemen führen. Die Teilnahme an einer betrieblichen Altersvorsorge ist eine gute Möglichkeit um zusätzlich zur gesetzlichen Rente ein Finanzpolster für den Ruhestand aufzubauen. Zu wenige Arbeitnehmer entscheiden sich jedoch dazu in die betriebliche Altersvorsorge einzuzahlen. Der Geschäftsführer eines Unternehmens möchte neu eingestellte Mitarbeiter dazu ermutigen, sich für die betriebliche Altersvorsorge anzumelden. Er hat zwei Ideen, wie er mehr Mitarbeiter ermutigen kann, sich anzumelden.

Eine Idee ist, die Anzahl der verfügbaren Investmentfonds, aus denen die Mitarbeiter wählen können, von 10 auf 15 zu erhöhen.

Eine weitere Idee ist, den Mitarbeitern Anmeldeunterlagen zur Verfügung zu stellen, in denen die beliebtesten der 10 vom Unternehmen angebotenen Investmentfonds hervorgehoben werden.

Im Folgenden sehen Sie verschiedene Handlungsmöglichkeiten, wie der Geschäftsführer sich verhalten könnte.

Der Geschäftsführer möchte so viele neu eingestellte Mitarbeiter wie möglich dazu ermutigen, sich bei der betrieblichen Altersversorge anzumelden. Daher beschließt er, für alle neu eingestellten Mitarbeiter die Anzahl der verfügbaren Investmentfonds von 10 auf 15 zu erhöhen.

Der Geschäftsführer möchte so viele neu eingestellte Mitarbeiter wie möglich dazu ermutigen, sich bei der betrieblichen Altersversorge anzumelden. Daher beschließt er, für alle neu eingestellten Mitarbeiter Anmeldeunterlagen zur Verfügung zu stellen, in denen die beliebtesten der 10 vom Unternehmen angebotenen Investmentfonds hervorgehoben werden.

Der Geschäftsführer möchte so viele neu eingestellte Mitarbeiter wie möglich dazu ermutigen, sich bei der betrieblichen Altersversorge anzumelden. Daher beschließt er, ein Experiment durchzuführen, bei dem die neu eingestellten Mitarbeiter nach dem Zufallsprinzip einer von zwei Gruppen zugewiesen werden. Der einen Hälfte der Mitarbeiter stellt er Anmeldeunterlagen zur

Verfügung, in denen die beliebtesten der 10 vom Unternehmen angebotenen Investmentfonds hervorgehoben werden. Für die andere Hälfte erhöht er die Anzahl der verfügbaren Investmentfonds von 10 auf 15. Nach einem Jahr wird der Geschäftsführer diejenige Vorgehensweise wählen, die dazu führt, dass sich die meisten Mitarbeiter für die betriebliche Altersvorsorge des Unternehmens anmelden.

English translation

The scenario describes how the managing director of a company might behave with regard to the company pension scheme for his employees.

Try to consider each possible action separately from the others. Please read the entire text on the next page carefully and refer to the text when answering the questions.

Most people save too little for their retirement. This can lead to financial problems later in life. Participating in a company retirement plan is a good way to build up financial savings for retirement in addition to the statutory pension. However, too few employees choose to enroll in a company retirement plan. The managing director of a company wants to encourage newly hired employees to enroll in the company retirement plan. He has two ideas on how to encourage more employees to enroll.

One idea is to increase the number of investment funds available for employees to choose from from 10 to 15.

Another idea is to provide employees with enrollment materials highlighting the most popular of the 10 investment funds offered by the company.

Below are several possible courses of action the manager could take.

The managing director wants to encourage as many newly hired employees as possible to enroll in the company retirement plan. He therefore decides to increase the number of available investment funds from 10 to 15 for all new employees.

The managing director wants to encourage as many newly hired employees as possible to enroll in the company retirement plan. He therefore decides to provide enrollment documents for all new hires, highlighting the most popular of the 10 investment funds offered by the company.

The managing director wants to encourage as many newly hired employees as possible to enroll in the company retirement plan. He therefore decides to conduct an experiment in which the newly hired employees are randomly assigned to one of two groups. For one half of the employees, he provides enrollment documents highlighting the most popular of the 10 investment

funds offered by the company. For the other half, he increases the number of investment funds available from 10 to 15. After one year, the manager will choose the approach that results in the most employees enrolling in the company's pension scheme.

Vignette 5: Hospital Safety

German original

Das Szenario beschreibt, wie sich der Direktor eines Krankenhauses zur Reduktion von Infektionen verhalten könnte.

Versuchen Sie, jede Handlungsmöglichkeit getrennt von den anderen zu betrachten. Bitte lesen Sie den gesamten Text auf der nächsten Seite sorgfältig durch und beziehen Sie sich bei der Beantwortung der Fragen auf den Text.

Bei einigen medizinischen Behandlungen muss ein Arzt einen Plastikschlauch in eine große Vene einführen. Diese Behandlungen können Leben retten, aber sie können auch zu tödlichen Infektionen führen. Der Direktor eines Krankenhauses möchte diese Infektionen reduzieren. Er hat zwei Ideen, wie das erreicht werden könnte.

Eine Idee ist, jedem Arzt, der dieses Verfahren durchführt, einen neuen Dienstausweis zu geben, auf dessen Rückseite eine Liste mit Standard-Sicherheitsvorkehrungen für das Verfahren aufgedruckt ist.
Eine weitere Idee ist, in allen Behandlungsräumen ein Poster mit einer Liste von Standard-Sicherheitsvorkehrungen für das Verfahren aufzuhängen.

Im Folgenden beschreiben wir Ihnen verschiedene Handlungsmöglichkeiten, wie der Direktor sich verhalten könnte.

Der Direktor möchte tödliche Infektionen so weit wie möglich reduzieren. Daher beschließt er, jedem Arzt, der dieses Verfahren durchführt, einen neuen Dienstausweis zu geben, auf dessen Rückseite eine Liste mit Standard-Sicherheitsvorkehrungen für das Verfahren aufgedruckt ist. Alle Patienten, die sich diesem Verfahren unterziehen, werden dann von Ärzten behandelt, die diese Liste an ihrer Kleidung tragen.

Der Direktor möchte tödliche Infektionen so weit wie möglich reduzieren. Daher beschließt er, in allen Behandlungsräumen ein Poster mit einer Liste von Standard-Sicherheitsvorkehrungen für dieses Verfahren aufzuhängen. Alle Patienten, die sich diesem Verfahren unterziehen, werden dann in Zimmern behandelt, in denen diese Liste an der Wand hängt.

Der Direktor möchte tödliche Infektionen so weit wie möglich reduzieren. Daher beschließt er, ein Experiment durchzuführen, bei dem die Patienten nach dem Zufallsprinzip einer von zwei Gruppen zugewiesen werden. Die eine Hälfte der Patienten wird von Ärzten behandelt, die einen neuen Dienstausweis erhalten haben, auf dessen Rückseite eine Liste mit Standard-Sicherheitsvorkehrungen über das Verfahren aufgedruckt ist. Die andere Hälfte wird in Räumen behandelt, in denen ein Poster mit denselben Standard-Sicherheitsvorkehrungen an der Wand hängt. Nach einem Jahr wird der Direktor alle Patienten auf die Weise behandeln lassen, bei der die Überlebensrate am höchsten ist.

English translation

The scenario describes how the director of a hospital could act to reduce infections.

Try to consider each possible action separately from the others. Please read the entire text on the next page carefully and refer to the text when answering the questions.

Some medical treatments require a doctor to insert a plastic tube into a large vein. These treatments can save lives, but they can also lead to deadly infections. The director of a hospital wants to reduce these infections. He has two ideas on how this could be achieved.

One idea is to give every doctor who performs this procedure a new ID badge with a list of standard safety precautions for the procedure printed on the back.

Another idea is to hang a poster with a list of standard safety precautions for this procedure in all treatment rooms.
In the following, we describe various possible actions that the director could take.

The director wants to reduce deadly infections as much as possible. He therefore decides to give each doctor who performs this procedure a new ID badge with a list of standard safety precautions for the procedure printed on the back. All patients undergoing this procedure will then be treated by doctors with this list attached to their clothing.

The director wants to reduce deadly infections as much as possible. He therefore decides to put up a poster in all treatment rooms with a list of standard safety precautions for this procedure. All patients undergoing this procedure will then be treated in rooms with this list posted on the wall.

The director wants to reduce deadly infections as much as possible. He therefore decides to conduct an experiment in which patients are randomly assigned to one of two groups. One half of the patients will be treated by doctors who have received a new ID badge with a list of standard safety precautions about the procedure printed on the back. The other half are treated in rooms

with a poster listing the same standard safety precautions on the wall. After one year, the director will have all patients treated in the way with the highest survival rate.

Vignette 6: Public Assistance Program

German original

Das Szenario beschreibt, wie sich der Manager einer wohltätigen Organisation verhalten könnte.

Versuchen Sie, jede Handlungsmöglichkeit getrennt von den anderen zu betrachten. Bitte lesen Sie den gesamten Text auf der nächsten Seite sorgfältig durch und beziehen Sie sich bei der Beantwortung der Fragen auf den Text.

Eine wohltätige Organisation möchte Haushalten mit geringem Einkommen helfen, ihre Stromrechnung zu senken. Dafür betreibt die Organisation ein Austauschprogramm für Kühlgeräte: Einkommensschwache Haushalte, die ein altes und ineffizientes Kühlgerät besitzen, erhalten bei einem Hausbesuch ein Informationsschreiben. Das Schreiben erklärt, wieviel ein bestimmter Haushalt sparen kann, wenn er sein Kühlgerät austauscht. Und es lädt den Haushalt ein, einen Rabattgutschein für ein neues Kühlgerät anzufordern.

Ein Manager der Organisation ist für das Austauschprogramm zuständig. Er findet, dass zu wenige bedürftige Haushalte den Gutschein anfordern und ihr Gerät austauschen. Er denkt über zwei verschiedene Ideen nach, um mehr Haushalte zu überzeugen.

Eine Idee ist, das Informationsschreiben umzuformulieren. Das neue Schreiben soll nicht mehr betonen, dass der Haushalt Geld spart, wenn er das Kühlgerät austauscht. Stattdessen soll es betonen, dass dem Haushalt Geld entgeht, wenn er das Kühlgerät nicht austauscht.
Eine weitere Idee ist, die Haushalte an die Austauschmöglichkeit zu erinnern. Dazu wird beim Hausbesuch ein Anhänger mit dem Logo des Programms im Inneren des alten Geräts angebracht. Diesen Anhänger sieht man jedes Mal, wenn man die Gerätetür öffnet.

Im Folgenden sehen Sie verschiedene Handlungsmöglichkeiten, wie der Manager sich verhalten könnte.

Der Manager möchte die Anzahl der Haushalte, die ihr Kühlgerät ersetzen, erhöhen. Er beschließt daher, dass ab sofort alle Haushalte beim Hausbesuch das umformulierte Infoschreiben mit Betonung der entgangenen Ersparnis ohne Gerätetausch erhalten werden.

Der Manager möchte die Anzahl der Haushalte, die ihr Kühlgerät ersetzen, erhöhen. Er beschließt daher, dass ab sofort bei allen Haushalten beim Hausbesuch der Anhänger im Kühlgerät angebracht wird.

Der Manager möchte die Anzahl der Haushalte, die ihr Kühlgerät ersetzen, erhöhen. Er beschließt daher, ein Experiment durchzuführen. Haushalte werden ab sofort nach dem Zufallsprinzip einer von zwei Gruppen zuordnet. Bei einer Hälfte der Haushalte wird das umformulierte Informationsschreiben mit Betonung der entgangenen Ersparnis ohne Gerätetausch ausgehändigt. Bei der anderen Hälfte wird der Anhänger im Kühlgerät angebracht. Nach einem Jahr vergleicht der Manager, bei welcher Gruppe mehr Kühlgeräte ausgetauscht wurden und wendet diese in Zukunft auf alle weiteren Haushalte an.

English translation:

The scenario describes how the manager of a charitable organization might behave.

Try to consider each possible action separately from the others. Please read the entire text on the next page carefully and refer to the text when answering the questions.

A charitable organization wants to help low-income households reduce their electricity bills. To this end, the organization runs a replacement program for cooling appliances: low-income households that have an old and inefficient cooling appliance receive an information letter during a home visit. The letter explains how much a particular household can save by replacing its cooling appliance. And it invites the household to request a discount voucher for a new cooling appliance.

A manager from the organization is responsible for the exchange program. He finds that too few households in need request the voucher and replace their appliance. He is thinking about two different ideas to convince more households.

One idea is to reformulate the information letter. The new letter should no longer emphasize that the household will save money if it replaces the cooling appliance. Instead, it should emphasize that the household will lose money if it does not replace the cooling appliance.

Another idea is to remind households of the replacement option. To do this, a tag with the program logo is placed inside the old appliance during the home visit. This tag can be seen every time the appliance door is opened.

Below you can see various possible actions the manager could take.

The manager would like to increase the number of households that replace their cooling appliance. He therefore decides that, as of now, all households will receive the reworded information letter during the home visit, emphasizing the savings lost without replacing the appliance.

The manager would like to increase the number of households that replace their cooling appliance. He therefore decides that, as of now, the tag will be attached to all households during the home visit.

The manager wants to increase the number of households that replace their cooling appliance. He therefore decides to conduct an experiment. Households are now randomly assigned to one of two groups. Half of the households will receive the rephrased information letter emphasizing the lost savings without replacing the appliance. For the other half, the label is attached to the refrigerator. After one year, the manager compares which group has had more refrigerators replaced and applies this to all other households in future.

Wording of treatments

Condition	German Original	English Translation
(i) Moral Approval	Wir präsentieren Ihnen gleich mehrere Handlungsmöglichkeiten. Dann bitten wir Sie zu bewerten, welche Handlungsmöglichkeit das moralisch angemessenste Verhalten beschreibt.	We will present you with several possible courses of action. Then we will ask you to evaluate which option describes the most morally appropriate behavior.
(ii) Injunctive Norm	Wir präsentieren Ihnen gleich mehrere Handlungsmöglichkeiten. Dann bitten wir Sie zu bewerten, wie sich jemand in	We will present you with several possible courses of action. Then we will ask you to evaluate how someone
	der Entscheidungssituation verhalten sollte.	should behave in the decision-making situation.
(iii) Hypothetical Action	Wir präsentieren Ihnen gleich mehrere Handlungsmöglichkeiten. Dann bitten wir Sie zu bewerten, welche der Handlungsmöglichkeiten am besten beschreibt, wie Sie sich in der Entscheidungssituation verhalten würden.	We will present you with several possible courses of action. Then we will ask you to evaluate which option best describes how you would behave in the decisionmaking situation.

Science literacy scale

German original:

1. Zwei Wissenschaftlerinnen wollen wissen, ob ein neuer Dünger das Pflanzenwachstum steigert. Die erste Wissenschaftlerin möchte den neuen Dünger an 1000 Pflanzen geben und sehen, wie viele von ihnen größer werden als vor der Gabe des neuen Düngers. Die zweite Wissenschaftlerin möchte 500 Pflanzen mit dem neuen Dünger versorgen und weitere 500 Pflanzen mit dem normalen Dünger und beobachten, ob die Pflanzen in der ersten Gruppe mehr wachsen als die Pflanzen in der zweiten Gruppe. Welche ist die bessere Methode, um den neuen Dünger zu testen?

Den neuen Dünger allen 1000 Pflanzen geben.
Den neuen Dünger 500 Pflanzen geben und weiteren 500 Pflanzen den normalen Dünger geben.

2. Ein Arzt teilt einem Paar mit, dass es aufgrund seiner genetischen Veranlagung eine Chance von eins zu vier hat, ein Kind mit einer Erbkrankheit zu bekommen. Bedeutet dies, dass, wenn ihr erstes Kind die Krankheit hat, die nächsten drei nicht erkranken werden?

Nein, die nächsten drei Kinder können die Krankheit trotzdem erben. Ja,
keines der nächsten drei Kinder wird die Krankheit erben.

3. Der Erdmittelpunkt ist sehr heiß.
Richtig
Falsch

4. Jegliche Radioaktivität wird vom Menschen verursacht.
Richtig
Falsch

5. Laser funktionieren durch die Bündelung von Schallwellen.
Richtig
Falsch

6. Elektronen sind kleiner als Atome.
Richtig
Falsch

7. Die Kontinente haben sich seit Millionen von Jahren verschoben und werden sich auch weiterhin bewegen.
Richtig
Falsch

8. Es ist das Gen des Vaters, das darüber entscheidet, ob ein Baby ein Junge oder ein Mädchen ist.
Richtig
Falsch

9. Antibiotika töten sowohl Viren als auch Bakterien.
Richtig
Falsch

English translation:

1. Two scientists want to know if a certain new plant food will increase plant growth. The first scientist wants to give the new food to 1,000 plants and see how many of them grow larger than they were before they received the new food. The second scientist wants to give the new food to 500 plants and give the normal, standard food to another 500 plants, and see whether the plants in the first group grow more than the plants in the second group. Which is the better way to test this plant food?

Give the new food to all 1,000 plants. Give the new food to 500 plants and give the normal, standard food to another 500 plants.

2. A doctor tells a couple that their genetic makeup means that they have a one in four chance of having a child with an inherited illness. Does this mean that if their first child has the illness, the next three will not?

No, the next three still might inherit the illness.
Yes, none of the next three children will have the illness.

3. The center of the earth is very hot.
True
False

4. All radioactivity is man-made.
True
False

5. Lasers work by focusing sound waves.
True
False

6. Electrons are smaller than atoms.
True

False

7. The continents have been moving their location for millions of years and will continue to move.
True
False

8. It is the father's gene that decides whether the bay is a boy or a girl. True
False

9. Antibiotics kill viruses as well as bacteria.
True
False

Demographics

German original:

Mit welchem Geschlecht identifizieren Sie sich?
Männlich
Weiblich
Divers

Was ist Ihr höchster Bildungsabschluss?
Hauptschulabschluss
Mittlere Reife
Allgemeine Hochschulreife/Fachhochschulreife
Bachelor- oder Masterabschluss/Diplom
Promotion

Haben Sie eine abgeschlossene Berufsausbildung?
Ja
Nein

Haben Sie einen Hochschulabschluss in einem naturwissenschaftlichen oder technischen Bereich? Ja
Nein

Welcher Altersgruppe gehören Sie an?

Unter 20 Jahre
20-29 Jahre
30-39 Jahre
40-49 Jahre
50-59 Jahre
60-60 Jahre 70
Jahre und älter

Welche Partei würden Sie wählen, wenn am kommenden Sonntag Bundestagswahl wäre?
SPD
CDU/CSU
FDP
Grüne
Die Linke
AfD
Sonstige
Nichtwähler/in

Ist Deutsch Ihre Muttersprache?
Ja
Nein

English translation:

What gender do you identify with?
Male
Female
Other

What is your highest level of education? Basic
secondary school diploma
Intermediate secondary school diploma
Specialized higher education entrance qualification
Bachelor or Master
PhD

Do you have a vocational degree?
Yes

No

Do you have a science degree?
Yes
No

Which age bracket do you belong to?
Below 20 years
20-29 years
30-39 years
40-49 years
50-59 years
60-60 years 70
years and older

If there was a Federal Election coming up this Sunday, which party would you vote for?
SPD
CDU/CSU
FDP
Grüne (Green Party)
Die Linke
AfD
Other
Not voting

Are you a German native speaker?
Yes
No

Studies 2 and 3

German original:

Wir zeigen Ihnen jetzt Aussagen zu den beiden Methoden.

Hier ist nochmal eine Erinnerung, welche Methoden wir testen wollen:

	Methode A: Betonung der Mehrkosten ohne Kühlgerätetausch im Infoschreiben (statt Betonung der Ersparnis durch Kühlgerätetausch)
	Methode B: Stecky Erinnerung als Anhänger im Kühlschrank zusätzlich zu Infoschreiben mit Betonung der Ersparnis durch Kühlgerätetausch
	Zum Vergleich: Das bisherige Infoschreiben betont die Ersparnis durch den Kühlgerätetausch.

In einer Vorstudie haben wir bereits beide Methoden im Zeitraum von April bis Juli mit jeweils 100 Haushalten getestet.

Auf dem nächsten Bildschirm bitten wir Sie zu schätzen, wie viele Haushalte (von 100) den Gutschein innerhalb von 90 Tagen nach Erhalt des Infoschreibens angefordert und eingelöst haben.

Wir möchten, dass Sie genau über diese Frage nachdenken. Für jede richtige Einschätzung erhalten Sie einen Betrag von 1,50€ als Amazon-Gutschein (Prolific-Bonus) zusätzlich zur Aufwandsentschädigung. Eine Einschätzung ist richtig, wenn Sie die Anzahl der Haushalte in einer Spanne von +/- 3 um die korrekte Anzahl herum nennen.

Im bisherigen Programm haben die meisten Haushalte ein Infoschreiben erhalten, in dem wir betonen, welche Vorteile (Ersparnis) sich daraus ergeben, den Gutschein einzulösen.

Im bisherigen Programm haben 14 von 100 Haushalten den Gutschein innerhalb von 90 Tagen nach Erhalt des Infoschreibens eingelöst.

Auf diesem Bildschirm bitten wir Sie die Anzahl Haushalte zu schätzen, wenn wir stattdessen eine der neuen Methoden nutzen.

Mit dem bisherigen Infoschreiben sind es 14 (von 100) Haushalte, die den Gutschein anfordern und einlösen.

	Mit Methode A (Infoschreiben mit Betonung der Mehrkosten) sind es: ___
	Mit Methode B (Infoschreiben mit Betonung der Ersparnis + Stecky Erinnerung als Anhänger im Kühlschrank) sind es: ___

Vielen Dank für Ihre Einschätzungen!

In einigen Wochen werden wir manche Teilnehmer dieser Befragung zu einer zweiten Befragung einladen.

Ihre persönliche Wahrscheinlichkeit zum zweiten Termin eingeladen zu werden liegt bei 10% (90%).

In der zweiten Befragung in ein paar Wochen werden wir Sie noch einmal bitten, für Methode A und B abzuschätzen, wie viele Haushalte von 100 ihren Gutschein angefordert und eingelöst haben. Liegen Sie mit Ihrer Abschätzung in Teil 2 richtig, erhalten Sie einen Bonus von 5,50 Euro. Neben diesem Betrag verlosen wir unter allen Teilnehmern, die bei einer zufällig ausgewählten Frage richtig liegen, auch fünf Amazongutscheine im Wert von 50 Euro.

In der Zwischenzeit werden wir noch einmal Daten von 200 zusätzlichen Haushalten sammeln. Als Teil dieses Forschungsprojekts können Sie mitentscheiden, wie wir diese Daten sammeln werden. Wir werden nach dieser Befragung eine/n Teilnehmer/in auslosen und die 200 zusätzlichen Daten so sammeln, wie diese/r Teilnehmer/in es auf dem nächsten Bildschirm bestimmt.

Jede/r Teilnehmer/in kann eine Stimme abgeben. Ihre Stimme zählt einfach (zehnfach). Das bedeutet, dass Ihre Auswahl das Endergebnis schwächer (starker) beeinflusst als die Auswahl anderer Teilnehmer/innen.

Bevor Sie im zweiten Teil der Befragung eine weitere Abschätzung abgeben, werden wir Ihnen dann zeigen, wie sich die 200 zusätzlichen Haushalte entscheiden.

Auf diesem Bildschirm können Sie festlegen, wie wir zusätzliche Daten von 200 Haushalten erheben sollen. Bitte wählen Sie eine der vier Erhebungsmethoden:

	Daten für alle 200 Haushalte sollen mit **Methode A** (Betonung der Mehrkosten) erhoben werden
	Daten für alle 200 Haushalte sollen mit **Methode B** (Betonung der Ersparnis + Stecky Erinnerung als Anhänger im Kühlschrank) erhoben werden
	Daten für 100 zufällig ausgewählte Haushalte sollen mit **Methode A** (Betonung Mehrkosten) und Daten für zufällig ausgewählte 100 Haushalte sollen mit **Methode B** (Betonung der Ersparnis + Stecky Erinnerung als Anhänger im Kühlschrank) erhoben werden
	Daten für 200 Haushalte sollen erhoben werden, **wie im Programm bisher üblich** (Betonung der Ersparnis), ohne dass eine der neuen Methoden zum Einsatz kommt.

English translation:

We will now show you statements about the two methods.

Here is a reminder of the methods we want to test:

	Method A: Emphasizing the additional costs without refrigerator replacement in the info letter (instead of emphasizing the savings through refrigerator replacement)
	Method B: Stecky reminder as a tag in the refrigerator in addition to the information letter emphasizing the savings from replacing the refrigerator
	For comparison: The original info letter emphasizes the savings made by replacing the refrigerator.

In a preliminary study, we have already tested both methods with 100 households each in the period from April to July.

On the next screen, we ask you to estimate how many households (out of 100) requested and redeemed the voucher within 90 days of receiving the info letter.

We would like you to think carefully about this question. For each correct estimate, you will receive a €1.50 Amazon voucher (Prolific bonus) in addition to the expense allowance. An

estimate is correct if you state the number of households within a range of +/- 3 around the correct number.
In the program to date, most households have received an info letter in which we emphasize the benefits (savings) of redeeming the voucher.

In the previous program, 14 out of 100 households redeemed the voucher within 90 days of receiving the information letter.

On this screen, we ask you to estimate the number of households if we use one of the new methods instead.
With the previous information letter, 14 (out of 100) households request and redeem the voucher.

	With method A (info letter with emphasis on the additional costs) ___ households request and redeem the voucher
	With method B (information letter emphasizing the savings + Stecky reminder as a tag in the refrigerator) ___ households request and redeem the voucher

Thank you for your feedback!
In a few weeks, we will invite some of the participants in this survey to a second survey.
Your personal probability of being invited to the second appointment is 10% (90%).
In the second survey in a few weeks, we will ask you once again to estimate for methods A and B how many households out of 100 have requested and redeemed their voucher. If your estimate in part 2 is correct, you will receive a bonus of 5.50 euros. In addition to this amount, we will also be giving away five Amazon vouchers worth 50 euros to all participants who get a randomly selected question right.

In the meantime, we will collect data from 200 additional households. As part of this research project, you will have a say in how we collect this data. We will draw a participant after this survey and collect the 200 additional data as this participant determines on the next screen. Each participant can cast one vote. Your vote counts once (ten times). This means that your selection has a weaker (stronger) influence on the final result than the selection of other participants.

Before you provide a further estimate in the second round of the survey, we will show you how the 200 additional households decide.

On this screen you can specify how you would like us to collect additional data from 200 households. Please select one of the four survey methods:

	Data for all 200 households should be collected using **Method A** (emphasis on additional costs)
	Data for all 200 households should be collected using **method B** (emphasis on savings + Stecky reminder as a tag in the refrigerator)
	Data for 100 randomly selected households should be collected using **Method A** (emphasis on additional costs) and data for 100 randomly selected households should be collected using **Method B** (emphasis on savings + Stecky reminder as a tag in the refrigerator)
	Data for 200 households should be collected **as usual in the program so far** (emphasis on savings), without using one of the new methods.

Accompanying survey

In Study 3, the experiment was followed by a short survey on the experts' work area and potential relation to the public assistance program.

Additional Results

Study 1

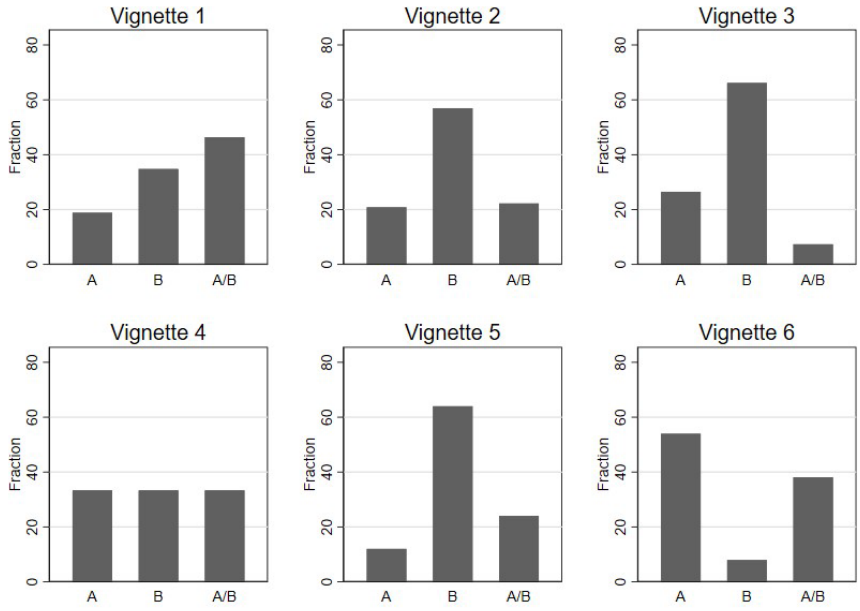
SA Table 1: Balancing Table (Study 1)

	Mean	*SD*
Treatment 1: Moral approval (N = 208)		
Gender (Male = 1)	.529	.5
High Education Level	.572	.496
Age	30.793	9.579
Scientific Reasoning Score	7.346	.961
Conservative Party	.029	.168
Right Wing Party	.038	.193
Green Party	.423	.495
Liberal Party	.087	.282
Treatment 2: Stated Preference Other (N = 204)		
Gender (Male = 1)	.456	.499
High Education Level	.564	.497
Age	31.838	9.709
Scientific Reasoning Score	7.294	.889
Conservative Party	.054	.226
Right Wing Party	.044	.206
Green Party	.392	.489
Liberal Party	.113	.317
Treatment 3: Stated Preference Self (N = 188)		
Gender (Male = 1)	.5	.501
High Education Level	.495	.501
Age	30.426	8.704
Scientific Reasoning Score	7.436	.872
Conservative Party	.08	.272
Right Wing Party	.021	.145
Green Party	.463	.5
Liberal Party	.08	.272

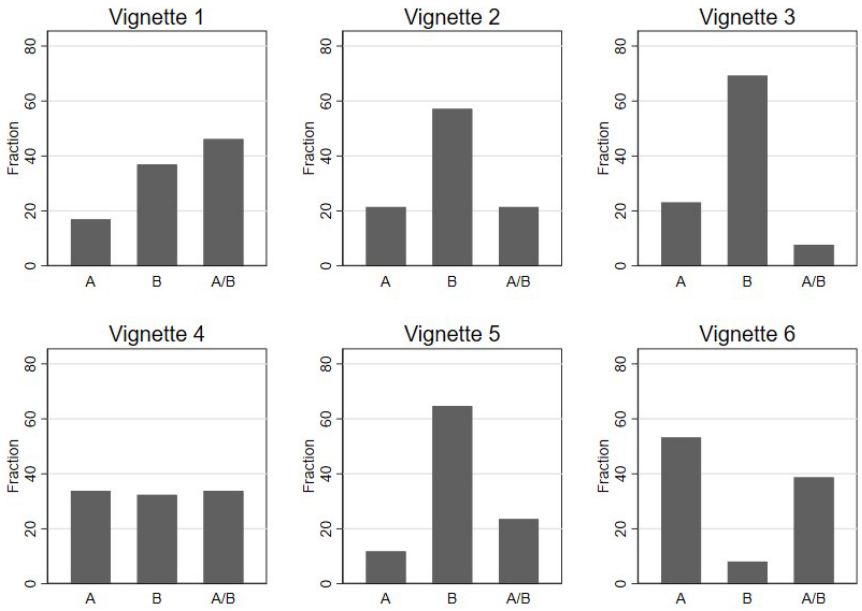
SA Table 1 summarizes the demographic attributes of participants across the three treatment conditions of the study. As expected from random assignment, attributes are strongly balanced across conditions. This is also confirmed by formal tests. Neither individual nor joint tests find evidence for treatment imbalances at $p<0.05$.

SA Figure 1: Moral approval of policies (Condition 1)

Panel A: Full sample



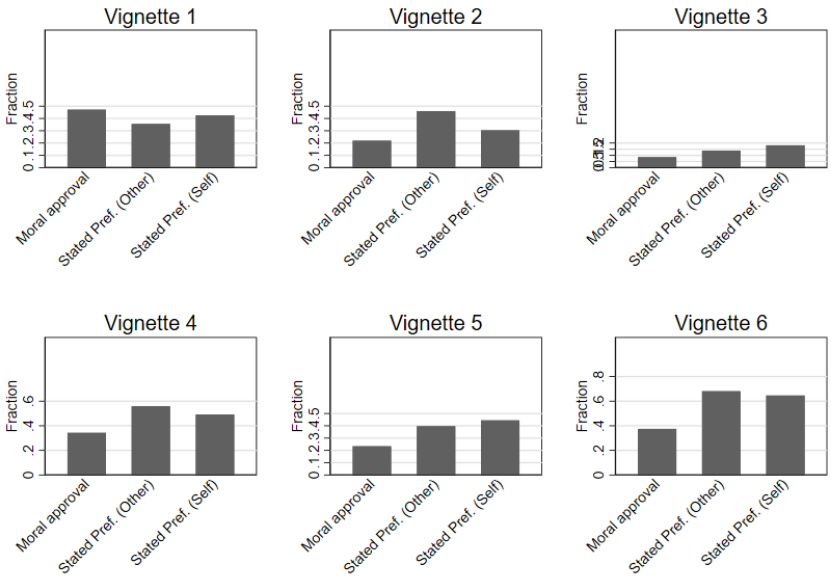
Panel B: Sample without inattentive participants



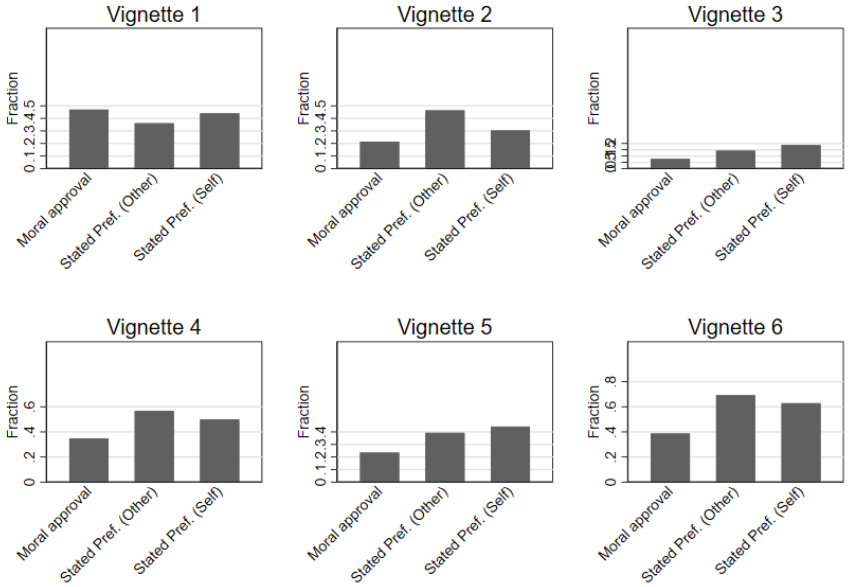
This figure shows, for each vignette, the fraction of participants who morally approve most of choosing policy A, policy B, or the A/B test. The upper panel shows the results for the full sample (600 participants), the lower panel for only the participants who answered neither of both attention checks incorrectly (575). In Vignettes 2, 3, 5, and 6, a single policy is clearly perceived as the most morally appropriate action. In Vignette 4, none of the options is clearly preferred. Only in the case of Vignette 1 (Poverty Alleviation) is the A/B test viewed as the most morally appropriate option in relative terms. However, in absolute terms, the A/B test does not receive majority support (>50%) in any of the vignettes.

SA Figure 2: Support for experiments across vignettes and treatment conditions 1-3

Panel A: Full sample



Panel B: Sample without inattentive participants



SA Figure 2 shows the fraction of participants in each condition (1, 2, 3) who supported the A/B test. The upper panel shows the results for the full sample (600 participants), the lower panel for only the participants who answered neither of both attention checks incorrectly (575). In all vignettes except Vignette 6, there is a clear increase in support for A/B testing as we move from

the elicitation of moral approval (condition 1) to the elicitation of an injunctive norm (condition 2) and then to the elicitation of a hypothetical action (condition 3). Vignette 1 is an exception, where support for experimentation is already high in condition 1. The corresponding statistical tests for these treatment differences can be found in SA Table 2 (non-parametric) and SA Table 3 (logit regression).

SA Table 2: Non-Parametric Tests of Condition 1 vs 2/3

Panel A: Full sample

	Vig. 1 **Poverty Alleviation**	**Vig. 2** **Autonomous Cars**	**Vig. 3** **Genetic Testing**	**Vig. 4** **Retirement Saving**	**Vig. 5** **Hospital Safety**	**Vig. 6** **Public Assistance Program**
Cond 1 vs Cond 2	*Pearson chi2(1) = 1.7643* *Pr = 0.184*	*Pearson chi2(1) = 8.5581* *Pr = 0.003*	*Pearson chi2(1) = 1.4877* *Pr = 0.223*	*Pearson chi2(1) = 7.0500* *Pr = 0.008*	Pearson chi2(1) = 3.7591 Pr = 0.053	Pearson chi2(1) = 11.4779 Pr = 0.001
Cond 1 vs Cond 3	*Pearson chi2(1) = 0.1755* *Pr = 0.675*	*Pearson chi2(1) = 1.12234* *Pr = 0.269*	*Pearson chi2(1) = 3.6786* *Pr = 0.055*	*Pearson chi2(1) = 3.4957* *Pr = 0.062*	Pearson chi2(1) = 6.8410 Pr = 0.009	Pearson chi2(1) = 9.0091 Pr = 0.003

Panel B: Sample without inattentive participants

	Vig. 1 **Poverty Alleviation**	**Vig. 2** **Autonomous Cars**	**Vig. 3** **Genetic Testing**	**Vig. 4** **Retirement Saving**	**Vig. 5** **Hospital Safety**	**Vig. 6** **Public Assistance Program**
Cond 1 vs Cond 2	*Pearson chi2(1) = 1.4256* *Pr = 0.232*	*Pearson chi2(1) = 9.0599* *Pr = 0.003*	*Pearson chi2(1) = 1.4277* *Pr = 0.232*	*Pearson chi2(1) = 7.1410* *Pr = 0.008*	Pearson chi2(1) = 3.5871 Pr = 0.058	Pearson chi2(1) = 11.4397 Pr = 0.001
Cond 1 vs Cond 3	*Pearson chi2(1) = 0.0529* *Pr = 0.818*	*Pearson chi2(1) = 1.4535* *Pr = 0.228*	*Pearson chi2(1) = 3.5817* *Pr = 0.058*	*Pearson chi2(1) = 3.4947* *Pr = 0.062*	Pearson chi2(1) = 6.6168 Pr = 0.010	Pearson chi2(1) = 7.2600 Pr = 0.007

SA Table 2 presents the results of non-parametric Chi-squared tests comparing condition 1 (Moral Approval) to the other conditions 2 (Stated Preference Other) and 3 (Stated Preference Self) for the full samples and the samples without inattentive participants. For most vignettes, with the exception of Vignette 1, there is significantly greater support for experimentation in conditions 2 and/or 3.

SA Table 3: Regression based evidence: Experimental Approval

	Sample w/o inattentive participants		Full sample	
Treatment = Stated Pref. Other	2.212***	(0.448)	2.170***	(0.431)
Treatment = Stated Pref. Self	2.068***	(0.405)	2.082***	(0.402)
Vignette Score (Moral)	1.462***	(0.158)	1.469***	(0.152)
Vignette Score (Impact)	1.072	(0.088)	1.044	(0.084)
Vignette Score (Knowledge)	0.709***	(0.062)	0.716***	(0.062)
Scientific Reasoning Score	1.099	(0.094)	1.091	(0.092)
Gender (Male = 1)	0.958	(0.158)	0.995	(0.160)
High Education Level (1=Yes)	0.991	(0.161)	1.040	(0.165)
Conservative Party (1=Yes)	1.483	(0.468)	1.345	(0.421)
Right Wing Party (1=Yes)	0.734	(0.355)	0.703	(0.333)
Green Party (1=Yes)	1.223	(0.224)	1.285	(0.230)
Liberal Party (1=Yes)	1.119	(0.309)	1.145	(0.308)
Vignette=2	0.338***	(0.093)	0.332***	(0.090)
Vignette=3	0.262***	(0.073)	0.270***	(0.074)
Vignette=4	0.067***	(0.023)	0.066***	(0.022)
Vignette=5	0.484**	(0.117)	0.464**	(0.111)
Vignette=6	0.325***	(0.090)	0.324***	(0.087)
Constant	0.230	(0.176)	0.240	(0.182)
Observations	1150		1200	
Prob > chi2	0.0000		0.0000	

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

SA Table 3 provides additional evidence on the factors influencing support for experimentation, using a panel logit regression with support for experimentation as the dependent variable. Robust standard errors are clustered at the individual level, and the results are reported as odds ratios.

The first two rows indicate that support is significantly higher in both stated preference conditions 2 and condition 3 compared to moral approval in condition 1. The next three variables capture participants' ratings of each vignette on the dimensions of the morality of the situation, the extent of impact on others, and existing knowledge of the policy's effectiveness. Experimentation is more strongly supported in vignettes perceived as less morally charged and receives less support when there is already substantial knowledge about the policy. None of the individual characteristics, including education or scientific reasoning abilities, drive support for experimentation. Support for experimentation also varies across vignettes, as reflected by the significant vignette fixed effects.

Studies 2 and 3

SA Table 4: Balancing Table (Study 2)

	Mean	*SD*
T1: Low Stakes – Low Consequences (N = 125)		
Gender (Male = 1)	0.496	.502
High Education Level	.544	.5
Low Income	0.16	0.368
Attitude Score: Program	384.128	80.026
Attitude Score: Policy A	120.784	51.532
Attitude Score: Policy B	89.744	56.889
T2: High Stakes – High Consequences (N = 127)		
Gender (Male = 1)	.488	.502
High Education Level	.551	.499
Low Income	0.126	0.333
Attitude Score: Program	379.472	84.821
Attitude Score: Policy A	110.850	53.840
Attitude Score: Policy B	99.890	56.619
T3: High Stakes – Low Consequences (N = 125)		
Gender (Male = 1)	.488	.502
High Education Level	.496	.502
Low Income	0.152	0.360
Attitude Score: Program	387.664	73.405
Attitude Score: Policy A	118.616	56.477
Attitude Score: Policy B	99.768	61.191
T4: Low Stakes – High Consequences (N = 123)		
Gender (Male = 1)	.496	.502
High Education Level	.480	.502
Low Income	0.146	0.355
Attitude Score: Program	388.919	84.388
Attitude Score: Policy A	108.862	58.648
Attitude Score: Policy B	100.382	61.042

SA Table 5: Balancing Table (Study 3)

	Mean	*SD*
T1: Low Stakes – Low Consequences (N = 17)		
Gender (Male = 1)	.529	.514
High Education Level	.294	.470
Attitude Score: Program	414.941	79.745
Attitude Score: Policy A	119.177	57.977
Attitude Score: Policy B	126.588	66.766
T2: High Stakes – High Consequences (N = 20)		
Gender (Male = 1)	.800	.41
High Education Level	.700	.47
Attitude Score: Program	440.9	66.167
Attitude Score: Policy A	112.15	59.582
Attitude Score: Policy B	145.3	40.684
T3: High Stakes – Low Consequences (N = 17)		
Gender (Male = 1)	.588	.507
High Education Level	.647	.493
Attitude Score: Program	422.941	60.512
Attitude Score: Policy A	89.059	64.368
Attitude Score: Policy B	133.941	42.615
T4: Low Stakes – High Consequences (N = 21)		
Gender (Male = 1)	.429	.507
High Education Level	.571	.507
Attitude Score: Program	401.857	76.831
Attitude Score: Policy A	92.476	61.505
Attitude Score: Policy B	117.238	56.119

SA Table 4 and 5 summarizes the demographic attributes of participants across the four treatment conditions of Study 2 and 3. As expected from random assignment, attributes are strongly balanced across conditions. This is also confirmed by formal tests for Study 2. Neither individual nor joint tests find evidence for treatment imbalances at $p<0.05$. For Study 3 observation numbers are lower and hence unsurprisingly there are some gender imbalances across treatments. When

investigating treatment effects below we combined both samples to achieve greater balance and control for demographic variables.

SA Table 6: Regression-based evidence: Experimental Approval

	(1) Main	(2) +Demographics	(3) +Views
Left-out category: Low stakes & consequences			
Stakes (High = 1)	-0.382	-0.385	-0.396
	(-1.44)	(-1.45)	(-1.47)
Consequences (High = 1)	-0.361	-0.362	-0.338
	(-1.37)	(-1.37)	(-1.26)
Stakes # Consequences	0.00853	0.00689	0.0155
	(0.03)	(0.02)	(0.06)
Male (1=Yes)		-0.261	-0.290
		(-1.36)	(-1.49)
High Education (1=Yes)		0.201	0.245
		(1.08)	(1.29)
Low Income (1=Yes)		0.0919	0.155
		(0.32)	(0.53)
Score Program			0.000816
			(0.66)
Score Policy A			0.00321*
			(1.85)
Score Policy B			0.00219 (1.25)
Constant	1.117****	1.140****	0.225
	(5.73)	(4.52)	(0.42)
Observations	575	575	575
Prob > chi2	0.255	0.306	0.136

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

SA Table 6 provides additional evidence on the factors influencing support for experimentation, using a logit regression with support for experimentation as the dependent variable. T statistics based on robust standard errors and coefficient estimates are transformed into odds ratios.

Model (1) shows that varying the stakes and consequences of participants' votes for policy implementation does not impact support for experimenting significantly. The first treatment dimension varies the stakes that participants have in the information derived from the experiment (high or low likelihood to be invited to the second prediction task, 10% vs 90%), while the second dimension varies the consequentiality of their decision, i.e., the likelihood of its implementation (low or high weight of own vote in lottery for policy implementation, 1 vs. 10 votes. Model (2): Demographics do not significantly impact support either. Model (3): Of the attitudes to the public assistance program elicited before the experiment, only a positive attitude to Policy A positively affects support, with weak statistical significance.

We complement the analysis contained in SA Table 6 with an ANNOVA. In line with Table 6, we do not find significant effects of varying stakes and consequences on preferences for A/B-testing, both in separate t-tests ($p = 0.148$ and $p = 0.976$ respectively) as well as the interaction of both treatments ($p = 0.172$) and in a joint F-Test ($p = 0.252$).