

Dur, Robert; Non, Arjan; Prottung, Paul; Ricci, Benedetta

Working Paper

Who's Afraid of Policy Experiments?

Tinbergen Institute Discussion Paper, No. TI 2023-027/V

Provided in Cooperation with:

Tinbergen Institute, Amsterdam and Rotterdam

Suggested Citation: Dur, Robert; Non, Arjan; Prottung, Paul; Ricci, Benedetta (2023) : Who's Afraid of Policy Experiments?, Tinbergen Institute Discussion Paper, No. TI 2023-027/V, Tinbergen Institute, Amsterdam and Rotterdam

This Version is available at:

<https://hdl.handle.net/10419/273838>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

TI 2023-027/V
Tinbergen Institute Discussion Paper

Who's Afraid of Policy Experiments?

*Robert Dur*¹

*Arjan Non*¹

Paul Prottung

Benedetta Ricci

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and Vrije Universiteit Amsterdam.

Contact: discussionpapers@tinbergen.nl

More TI discussion papers can be downloaded at <https://www.tinbergen.nl>

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 598 4580

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900

Who's Afraid of Policy Experiments?*

Robert Dur[†], Arjan Non[‡], Paul Prottung, and Benedetta Ricci

May 2023

Abstract

In many public policy areas, randomized policy experiments can greatly contribute to our knowledge of the effects of policies and can thus help to improve public policy. However, policy experiments are not very common. This paper studies whether a lack of appreciation of policy experiments among voters may be the reason for this. Using unique survey data representative of the Dutch electorate, we find clear evidence contradicting this view. Voters strongly support policy experimentation and, in line with theory, particularly so when they do not hold a strong opinion about the policy. In a subsequent survey experiment among Dutch politicians, we find that politicians conform their expressed opinion about policy experiments to what we tell them the actual opinion of voters is. We conclude that voters are not afraid of policy experiments and neither are politicians when we tell them that voters are not.

Keywords: policy experiments, randomized controlled trials, voters, politicians, public policy, survey experiment, conformism.

JEL codes: C93, D72, D78

*We gratefully acknowledge comments and suggestions from seminar audiences at King's College London, Erasmus University Rotterdam, the University of Stavanger, the Dutch Ministry of Finance, the Netherlands Authority for the Financial Markets, and participants of the 6th Stavanger Behavioral Economics and Finance Workshop and the workshop Field Days: Experiments Outside the Laboratory 2022 in Cologne. We also thank Mareen Bastiaans, Anne Boring, Josse Delfgaauw, Peter John, Sacha Kapoor, Jori Korpershoek, Sandra Phlippen, Uri Simonsohn, Dana Sisak, Caroline Wehner, and Hide-Fumi Yokoo for helpful comments on an earlier draft of this paper. The voter survey data used in this paper were collected by Centerdata's LISS Panel, following an ODISSEI LISS Grant awarded to us in 2020. The data can be accessed at: https://www.dataarchive.lissdata.nl/study_units/view/1247. The data from the survey experiment among politicians will become available upon publication of this paper. The survey experiment is registered in the AEA RCT Registry (AEARCTR-0009142) and received approval from the Erasmus School of Economics IRB-Experimental (ETH2122-0500).

[†]Erasmus University Rotterdam, Tinbergen Institute, CESifo, and IZA. E-mail: dur@ese.eur.nl

[‡]Erasmus University Rotterdam, Tinbergen Institute, and CESifo. E-mail: non@ese.eur.nl

1 Introduction

The effects of public policies are often uncertain. Think, for instance, of a new training program for the unemployed, an advanced policing strategy to fight crime, or an innovative anti-discrimination policy. Not only *ex ante*, but also *ex post*, it is commonly hard to tell what the effects of such policies are. Even when we can observe how target outcomes (such as unemployment or crime) have developed since implementation, we do not know what would have happened if the policy had not been implemented. In other words, we lack knowledge about the counterfactual, which greatly complicates policy evaluation.

A randomized policy experiment can be very helpful in these cases. By implementing a policy for a randomly chosen subset of units only (e.g., a random subset of individuals, regions, or firms), we can observe the counterfactual outcome by looking at the control group units, and compare it with the outcome for the treated units to see what the effect of the policy is. If some further conditions are met,¹ a policy experiment delivers a precise and unbiased estimate of the effect of the policy, enabling policy makers to take an informed decision about whether or not to implement the policy fully.

Despite the potential power of policy experiments, they are not yet very common in practice – or at least not as common as they could be. Indeed, as Heffetz and List (2021) note, “governments around the world still routinely oppose controlled trials of public policies”. One possible reason is that voters do not appreciate policy experiments and, consequently, politicians hesitate to use them. Voters’ opposition to policy experiments could arise for at least four reasons. First, an often-heard argument is that voters may consider experiments to be unfair because, for some period of time, people or firms are treated unequally. Second, for urgently needed policies, voters may feel that experimentation takes too much time; that instead of experimentation, immediate action at full scale is needed. Third, voters may

¹These conditions include among others that the experiment is of sufficient size, that the experimental sample is representative, that the measurement of relevant outcomes is sufficiently precise and comprehensive, and that the outcomes of the units are independent (no contamination). See e.g. Glennerster and Takavarasha (2013).

worry about a possible lack of informed consent of those involved in the policy experiment. Fourth, experimental findings may lack external validity. On top of these concerns, politicians may also worry that voters consider politicians who propose policy experiments as less knowledgeable or competent than politicians who do not, as the former apparently do not know what the effects of policies are, whereas the latter typically claim that they do.

This paper uses unique representative Dutch survey data to study voters' opinion about randomized policy experiments and voters' attitude towards politicians who propose them. To clarify to voters what we mean by a policy experiment, we start the survey by sketching what a randomized experiment could look like for a number of policies of which the effects are uncertain. We ask voters for each of these policies what they think about policy implementation, about policy experimentation, and about sticking to the status quo (i.e. no implementation). After this elicitation, we ask a series of general questions about policy experiments, including voters' general appreciation, their concerns, and whether they would be more or less likely to vote for a politician who proposes to run more policy experiments.

In addition to the survey among voters, we also ran a survey among politicians to study their beliefs on what voters think about policy experiments and what they think about policy experiments themselves. We invited all members of the Dutch national parliament (both first and second chamber) as well as members of regional parliaments ('Provinciale Staten') to take part in our survey. About one fifth of the invited politicians completed the survey. During the survey, we informed politicians about the main results we found in our survey among voters. We informed a randomly selected half of politicians before we asked for their opinion about policy experiments (the treatment group), whereas we informed the other half of politicians at the end of the questionnaire (the control group). This experimental design allows us to assess whether politicians' views about policy experiments are affected by learning what voters think about policy experiments.

The results are as follows. Our main finding is that voters highly appreciate policy experiments. For five out of the six policies for which we sketch an experiment, experimentation

meets approval from a majority of voters. For four out of the six policies, experimentation gets higher popular support than both implementation and no implementation. Moreover, two thirds of our respondents rate experimentation strictly higher (and 86% weakly higher) than both implementation and no implementation for at least one of the policies.

Using the data at the individual level, we find that voters’ support for experimentation is strongest when they have no strong opinion about the policy, i.e. when they find implementation neither very good nor very bad. This result is consistent with the idea that policy experimentation is most useful when the knowledge generated by the policy experiment may be decisive for future policy implementation, which is in line with a simple theoretical model we present in Section 3.

We also study what share of voters might be seen as systematically opposed to policy experiments. Rating experimentation strictly lower than both implementation and not implementation is perhaps the clearest indication of aversion to experimentation (Mislavsky, Dietvorst, & Simonsohn, 2020). We find that only 0.2% of the respondents consistently rates experimentation strictly lower than both implementation and no implementation. Overall, in only 4% of all cases that voters considered, experimentation was rated as strictly lower than both implementation and no implementation.

Broad voter support for policy experiments is also found in the answers to the general questions. A majority of the electorate (60%) rates it as ‘good’ or ‘very good’ if one of their favorite political parties at the next elections would be in favor of running policy experiments much more often. Only 15% finds this ‘bad’ or ‘very bad’. The modal voter also says that she is more likely to vote for the party when it is in favor of much more policy experimentation.²

What concerns (if any) do voters have about policy experiments? We asked voters to give their opinion about four often-raised concerns: unfairness because of unequal treatment during the experiment, lack of informed consent, lack of external validity, and policy experiments being too time consuming. We find that on all four concerns voters’ opinions are quite

²40.4% says she more likely votes for the party, 40.1% says equally likely, and 19.6% says less likely.

split: about a third agrees it is a concern, another third does not think it is a concern, and the final third neither agrees nor disagrees. The answers to these four questions do a fairly good job in explaining voters' general opinion about policy experiments: about 25% of the variation is explained by these four concerns.

Lastly, we asked voters why they think that politicians are not often in favor of running policy experiments. Many voters think that electoral reasons play a role. Despite the popularity of policy experiments among voters that we uncover in our study, a substantial share of voters think that politicians fear to lose votes when proposing to run more policy experiments.

The survey experiment that we ran among politicians shows quite some variation in politicians' opinions about policy experiments. On average, the responding politicians are slightly more positive about policy experiments than voters are. However, we need to interpret this result with care, because in contrast to our sample of voters, our sample of politicians is not representative.³ Hence, regarding the survey among politicians, we mainly focus on the experimental results.

Strikingly, politicians in the control group (those not yet informed by us about voters' opinion) tend to think that voters share their opinion about policy experiments. Politicians in the treatment group respond strongly to learning what voters actually think about policy experiments: they conform to voters' appreciation of policy experiments to a large extent. In line with this, our findings suggest that, as compared to control politicians, treated politicians are more likely to say that they would actively support an initiative within their political party to strive for more policy experimentation. These results indicate that publicizing voters' high appreciation of policy experiments which we find in this paper may persuade politicians to support policy experiments.

³The left-wing parties are overrepresented and the right-wing parties are underrepresented, particularly the radical right parties. As the invitation to participate in the survey was signed by a professor (the first author on this paper) and the stated goal of the survey was academic research, we suspect that politicians who are more sceptical about academic research responded to our survey in smaller numbers. Clearly, this may result in an upward bias in the appreciation for policy experiments that we measure in the survey among politicians.

The remainder of this paper is structured as follows. The next section briefly discusses the related literature. Section 3 presents a simple theoretical model of voters’ preferences for policy experimentation and derives a few predictions. Next, in Section 4 we describe the results of the survey among voters. Section 5 shows the results of the survey experiment among politicians. Section 6 concludes.

2 Related literature

We build on a recent strand in the literature that uses survey data to study people’s appreciation of or aversion to experimentation. Our paper is inspired by Mislavsky et al. (2020), who investigate people’s opinions about experiments run by companies. They find that people tend to evaluate experiments positively, unless the experiment contains a treatment they object to. Hence, they conclude that “Experiments are not unpopular; unpopular policies are unpopular.” (p.1092). In contrast, based on a series of vignette studies in a variety of domains, Meyer et al. (2019a) conclude that people are averse to experimentation.⁴ However, in a recent replication study, Mazar, Elbaek, and Mitkidis (2023) show that Meyer et al. (2019a)’s findings are neither generalizable nor robust. A distinguishing feature of our study is its focus on public policy. Another important contribution to the literature is that we use a representative sample of voters instead of a convenience sample.⁵

Our survey experiment among politicians builds on studies about politicians’ (mis)perceptions of voters’ opinions (Broockman & Skovron, 2018) and politicians’ responsiveness to infor-

⁴The research design in Meyer et al. (2019a) has been forcefully criticized by Mislavsky, Dietvorst, and Simonsohn (2019) for not being diagnostic about experiment aversion. This critique was subsequently rebutted by Meyer et al. (2019b) and Heck, Chabris, Watts, and Meyer (2020).

⁵Mislavsky et al. (2020) use three samples: Amazon’s Mechanical Turk (MTurk), a sample of nonacademic university staff, and a representative sample of regular Amazon customers. Meyer et al. (2019a) use MTurk, Pollfish, and a sample of healthcare professionals, whereas Heck et al. (2020) use MTurk only. Mazar et al. (2023) use MTurk, Prolific, and Lucid. Two other related studies are Fischer, Grewenig, Lergetporer, and Werner (2019) and Cardon (2022). Fischer et al. (2019) study whether describing an evaluation as an experiment makes evaluation less popular. Using a representative sample of the German population, they find no evidence for this. Cardon (2022) uses a representative sample of the US population as well as a representative sample of public administrators to study people’s choice between a randomized policy experiment and a quasi-experiment in the context of a municipal lead abatement program. She finds a strong preference for the quasi-experiment, but less so among public administrators than among the public at large.

mation about the public opinion (Fox & Shotts, 2009; Butler, Nickerson, et al., 2011; Kalla & Porter, 2021; Sevenans, 2021; McAndrews, Lucas, Merkley, Rubenson, & Loewen, 2022; Soontjens & Sevenans, 2022; García-Sánchez, Matanock, & Garbiras-Díaz, 2022). We are the first to study how politicians’ willingness to use policy experiments is affected by learning about the public opinion about policy experiments.⁶

Lastly, we also contribute to the discussion on the ethics of experiments (List, 2008; Glennerster, 2017; McDermott & Hatemi, 2020; Asiedu, Karlan, Lambon-Quayefio, & Udry, 2021; Phillips, 2021; Charness, Samek, & Van de Ven, 2022). Experiments oftentimes raise ethical concerns that need to be weighed against the potential benefits of the knowledge that experiments produce. Examples of ethical concerns are the inequality arising from exposing people to different treatments for some time and a lack of informed consent, which is sometimes hard to avoid. In weighing these costs and benefits, it can be useful for committees (such as Institutional Review Boards) and researchers to know what the public at large thinks about these matters. That is what our study offers in the context of policy experiments. See Desposato (2018), Yokoo (2020), Naurin and Öhberg (2021), and John, Kim, and Soto-Tamayo (2023) for similar contributions in the context of field experiments in political science and economics.

3 Theoretical predictions

Under what conditions are voters in favor of a policy experiment? In this section we offer a simple model producing a few basic predictions about when voters may find a policy experiment appealing.⁷ We will confront these predictions with the data from the survey among voters in the next section.

⁶Our paper also relates to the recent literature on the use of policy experiments and experimental findings by politicians (Callander & Harstad, 2015; Bernecker, Boyer, & Gathmann, 2021; Corduneanu-Huci, Dorsch, & Maarek, 2021; Hjort, Moreira, Rao, & Santini, 2021; Mehmood, Naseer, & Chen, 2021; Wang & Yang, 2021; DellaVigna, Kim, & Linos, 2022; Vivalt & Coville, 2022). Starting with Campbell (1969), there is also an extensive literature in public administration and psychology on this topic.

⁷The model is loosely based on Dur (2001).

Suppose the future consists of two periods. There exists a policy of which the effects are uncertain. With probability p , the policy is a ‘success’ and yields per capita benefit B per period. With the remaining probability, the policy is a ‘failure’ and yields per capita cost C per period. In addition to the potential benefits B and costs C , voters derive value V_i per period from implementation of the policy. The value V_i is not surrounded by uncertainty, differs across voters, and can take any value (positive and negative). It is included to allow for voter i ’s taste for the policy arising from individual differences in ideological views or personal stakes in the policy. Not implementing the policy means that the status quo is maintained, which yields a net expected benefit normalized to zero.

The effect of the policy can only be learned by running an experiment that exposes a proportion x of the population to the policy during one period. Experimentation produces additional joy or suffering equal to E_i for voter i during the period of experimentation. The value E_i reflects voter i ’s heterogeneous (dis)taste for experimentation, including the concerns that voters may have about experimentation discussed in the previous section (unequal treatment for some period of time, possible lack of informed consent, etcetera).

Given these assumptions, voter i ’s expected net benefits from implementing the policy in both period 1 and 2 equal:

$$2[V_i + pB - (1 - p)C]. \tag{1}$$

Hence, voters who have sufficiently strong taste for the policy (i.e., sufficiently high V_i) will favor implementation to no implementation. Higher potential benefits B and a higher likelihood of them being realized p also work in favor of implementation, whereas the reverse holds for the potential costs C . Voters with a strong distaste of the policy (i.e., negative and sizeable V_i) will always favor no implementation to implementation.

The benefit from policy experimentation in the first period depends on whether the results of the experiment are consequential for optimal policy choice in the second period. Assume that they are, i.e. that upon learning that the policy is a ‘success’, the policy

is fully implemented in period 2, whereas upon learning that the policy is a ‘failure’, the policy experiment is discontinued and the policy is not implemented in period 2.⁸ Then, the expected net benefits from experimenting in period 1 and choosing policy in period 2 in line with the experimental results are equal to:

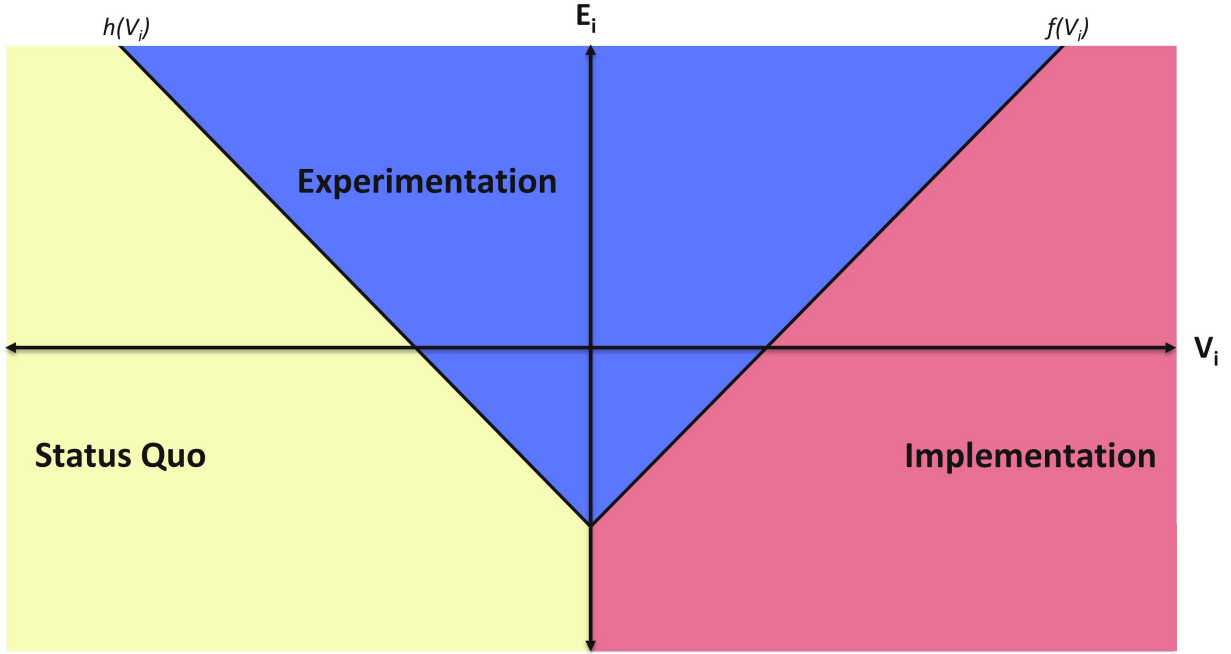
$$x[V_i + pB - (1 - p)C] + E_i + p(V_i + B). \quad (2)$$

Comparing (1) and (2), it follows that an advantage of policy experimentation is that it helps to avoid the cost C in period 2 in case the policy is a ‘failure’. To realize this advantage, however, the voter needs to give up the potentially positive expected net benefits from full implementation in period 1 as well as bear costs of (or reap benefits from) experimentation denoted by E_i .

Figure 1 plots the combinations of V_i and E_i for which implementation, no implementation, and experimentation are the best choice, for parameter values $x = 0.5$, $p = 0.5$, $B = 1$, and $C = 1$. As the figure clearly shows, experimentation is never the best choice for voters with a very strong opinion about the policy (large absolute value of V_i). Similarly, voters who bear high cost of experimentation (negative and sizeable E_i) are generally worse off from experimentation. Voters who hold no strong opinion about the policy and have no strong concerns about experimentation favor experimentation. We will test these basic predictions in the next section.

⁸It is straightforward to verify that if the results of the experiment are not consequential for the optimal policy in the second period, then running an experiment in period 1 can only be optimal if the taste for experimentation E_i is positive. In a richer model, Strulovici (2010) shows that experimenting may change the identity of the pivotal voter, which in turn may keep the pivotal voter from experimentation.

Figure 1: Voter's preferences for experimentation, implementation, and status quo



Note: The figure illustrates how voter's taste for the policy V_i and her taste for experimentation E_i affect her most preferred policy choice (experimentation, implementation, or status quo) according to the stylized model presented in this section. The parameter values used to construct the figure are $x = 0.5$, $p = 0.5$, $B = 1$, and $C = 1$. These values imply that the voter prefers implementation to status quo whenever $V_i > 0$. The lines labeled $h(V_i)$ and $f(V_i)$ plot combinations of V_i and E_i for which the voter is indifferent between experimentation and status quo and between experimentation and implementation, respectively.

4 Survey among voters

The voter survey data were collected by Centerdata’s LISS Panel. The LISS Panel consists of about 5,000 households that have been randomly selected from the population register of Statistics Netherlands. They complete an online survey about once a month. LISS Panel has made a serious effort to make the sample representative of the Dutch population.⁹ For instance, when people who have been selected indicate that they cannot take part because they do not have a laptop or internet connection, LISS Panel makes it available to them. Another attractive feature of the LISS Panel is the rich data that have already been collected on the sample, which can be easily linked to the newly collected data.

The ODISSEI LISS Grant that was awarded to us in 2020 allowed us to collect data on about 2,000 people. For that purpose, a random selection of 2,814 LISS Panel participants were approached in August 2021 via e-mail by LISS Panel, inviting them to take part in the survey. The invitation mentioned that the survey questions would be about issues related to public policy. Nonrespondents received a reminder twice. In the end, 2,114 people completed the survey, which means that the response rate was 75%.¹⁰ The full text of the survey (the original in Dutch and the translation in English) as well as the complete dataset can be downloaded from: https://www.dataarchive.lissdata.nl/study_units/view/1247.

How representative is our sample? Figure 2 reports some basic demographic variables for three groups: 1) the random selection of 2,814 LISS Panel participants who were invited to take part in our survey; 2) the group of 2,114 respondents to our survey; and 3) the 4,424 LISS Panel participants who report in a later wave (Wave 14) to have voted in the Dutch national parliamentary elections of March 2021.¹¹ Inspection of Figure 2 shows that, overall, the demographic characteristics of our respondents are close to those invited (compare the green and blue bars). However, we do observe some selective nonresponse among the young,

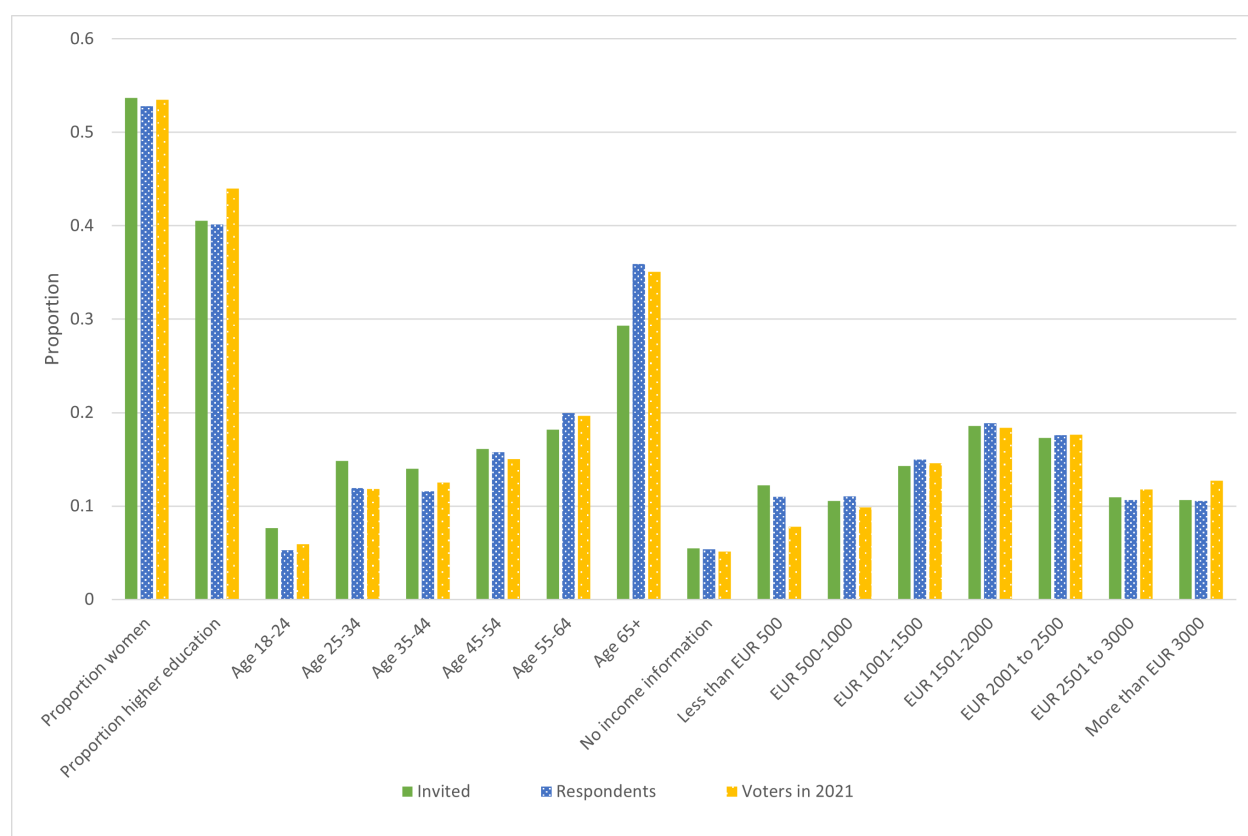
⁹See for a detailed description of the sampling and an analysis of the representativeness of the LISS Panel: <https://www.lissdata.nl/about-panel>.

¹⁰The LISS data archive reports that there are 2,112 respondents. We included two additional respondents for whom only the variable “response time” was missing.

¹¹Wave 14 took place from December 2021 to March 2022 and had a response rate of 86%.

whereas the older people are somewhat overrepresented in our sample. Interestingly, the age distribution of our respondents is very close to those who report in Wave 14 to have voted (compare the blue and yellow bars). Hence, our respondents seem to represent the kind of people who show up at the voting boot. In line with this, the election outcome based on our respondents' stated votes in Wave 14 is very close to the official election outcome, except for an underestimation of the size of radical right parties, see Figure A.1 in Appendix A. According to Wave 14, the turnout at the 2021 national elections among our respondents was 90%, whereas it was 79% in the general population. We therefore refer to our respondents as 'voters' in most of the text.

Figure 2: Demographic characteristics



Note: The figure shows demographic characteristics of people invited to take part in the survey, the respondents to the survey, and people who in a later wave report to have voted in the Dutch national parliamentary elections in 2021. Details are described in the main text.

4.1 Scenarios

To make clear to the respondents what we mean by a policy experiment, we started the survey by sketching a couple of scenarios. We created a total of six scenarios and showed a random four in randomized order to each respondent. Each scenario described a policy proposal and mentioned explicitly that the effects of the proposed policy are uncertain. Right after the description of each scenario, we asked the respondent to rate on a seven-point scale the desirability of three possible policy decisions: implementation, no implementation, and experimentation. We made sure to clearly describe what experimentation would entail (a treatment and control group, randomization, a comparison of outcomes of both groups after some time to assess the policy effect and to decide whether to implement the policy or not). Inspired by the theoretical model in the previous section, we constructed scenarios about a variety of policies so that respondents would not have a strong opinion about *each* of the four policies they evaluated. The scenarios included the following policies: 1) spending more money to combat tax evasion by large companies; 2) a program to help the unemployed find vacancies; 3) anonymizing applications for job vacancies at a municipality to reduce hiring discrimination; 4) spending more money on detecting benefit fraud; 5) a new policing approach (‘Very Irritating Police’) to reduce nuisance by loitering youth and asylum seekers;¹² 6) softer job application requirements for welfare recipients. As an example, the scenario about combating tax evasion reads as follows (translated from Dutch; the full text of the other scenarios can be found in Appendix B):

In order to combat tax evasion by big companies, the Minister of Finance is considering to spend more money on increased monitoring. It is not certain what the effects of this will be. The Minister can do three things:

1. **Implement:** The Minister decides to spend more money on increased monitoring of all big companies.
2. **Not implement:** The Minister decides *not* to spend more money on increased monitoring of big companies.

¹²This policy, including the explicit mention of youth and asylum seekers, was taken from the policy program of the largest political party in the Netherlands in 2021, the liberal party VVD.

3. Experiment: The Minister decides to conduct a policy experiment, in which a random selection of the big companies will be monitored more closely, while the others continue to be monitored the same as before. After the policy experiment, the two groups will be compared to see whether increased monitoring was successful and should be introduced for all companies.

What do you think about:

Implement? [1=Very bad; 2; 3; 4; 5; 6; 7=Very good]

Not implement? [1=Very bad; 2; 3; 4; 5; 6; 7=Very good]

Experiment? [1=Very bad; 2; 3; 4; 5; 6; 7=Very good]

Table 1 reports the percentages of respondents showing appreciation for implementation, no implementation, and experimentation for each of the six scenarios. Showing appreciation is defined as rating a policy decision with 5, 6, or 7 on the seven-point scale. Experimentation turns out to be quite popular. For all but one of the policies, experimentation is appreciated by a majority of respondents. Moreover, for four out of the six policies, experimentation is appreciated by a higher share of respondents than both implementation and no implementation.

Table 1: Appreciation for policy experimentation, implementation, and no implementation

Scenario	Experiment	Implement	Not implement
1: Combating corporate tax evasion	54%	63%	10%
2: Helping the unemployed find vacancies	64%	45%	14%
3: Anonymizing job applications	54%	47%	26%
4: Detecting benefit fraud	57%	52%	16%
5: Very Irritating Police approach	58%	49%	17%
6: Softening job application requirements	47%	19%	51%

Note: The table shows for each scenario the percentage of respondents that rates the policy decision with at least 5 on the seven-point scale. Bold numbers indicate for each scenario the policy decision with most support. As respondents evaluated a random four of the six scenarios, each scenario was evaluated by approximately 1,400 respondents.

Another way to assess the popularity of policy experiments is to ask: how often do respondents rate experimentation highest (alone or together with implementation and/or no implementation). Figure 3 shows the results. The blue bars indicate that only one out of seven never puts experimentation on top. Hence, 86% of the respondents rates experimentation highest at least once. A quarter of respondents even rates experimentation highest for all four scenarios.¹³ Conversely, we can ask: how often do people rate experimentation lowest (alone or together with implementation and/or no implementation)? The orange bars show that this does not happen often. Half of our respondents never give experimentation the lowest rating and only 15% does so more than two times.¹⁴

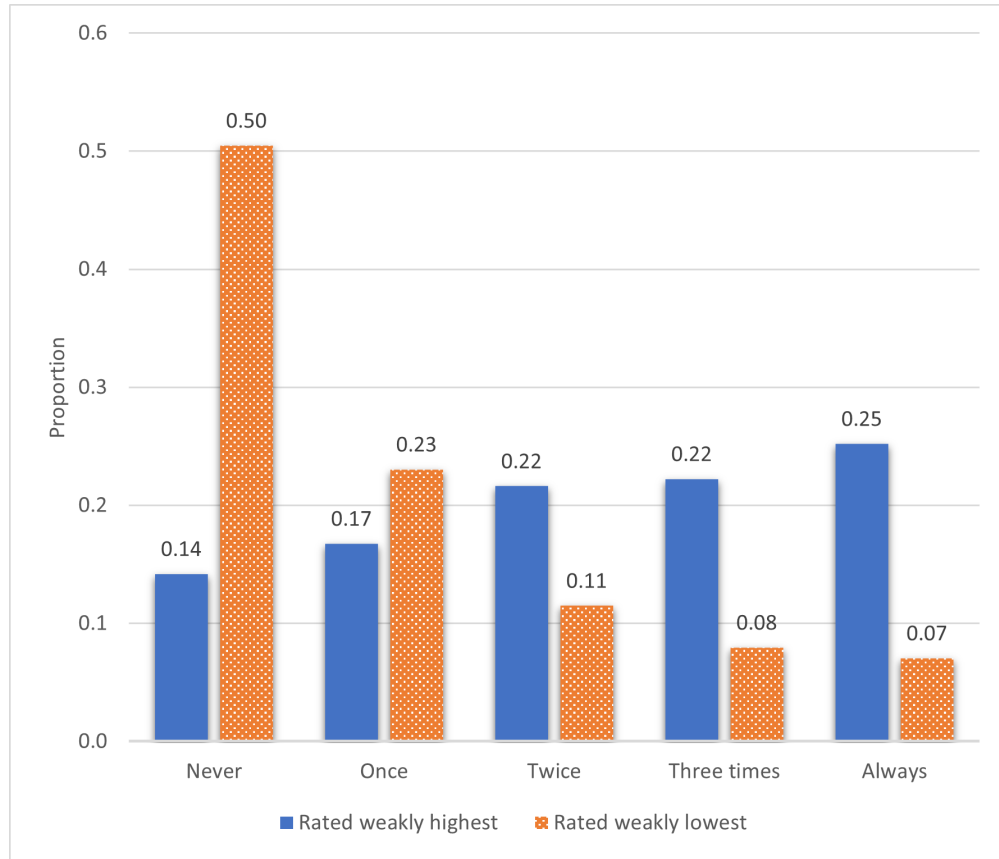
A more stringent measure of experimentation aversion considers how often experimentation is rated *strictly* lower than the alternative options (Mislavsky et al., 2020). Only 0.2% of the respondents does so in all four scenarios, 0.5% in three of them, 1.9% twice, and 9.9% once. In sum, in a mere 4% of all evaluations do respondents rate experimentation strictly lower than both implementation and no implementation. Hence, in 96% of all evaluations people find that experiments “are at least as acceptable as the worst policies they contain”, corroborating Mislavsky et al. (2020)’s findings for corporate experiments (p.1101).¹⁵

¹³Using the stricter criterion of strict rather than weak preference for experimentation, we find that 66% rates experimentation highest at least once and 10% always does so.

¹⁴In 6% of all cases, rating 4 (“neither good nor bad”) is given to all three options (experimentation, implementation, and no implementation). This may reflect a lack of knowledge, misunderstanding, or refusal to answer. From the answers to the questions at the end of the questionnaire, we see that indeed some respondents expressed confusion or misunderstanding. 10% finds the questions unclear. 30% states that the questions were difficult to answer. At the same time, 60% found the topic interesting and 53% says that the questionnaire made them think.

¹⁵A possible concern about these comparisons is that when respondents care strongly about a policy, they may be restricted in expressing their experimentation aversion. For instance, when they evaluate a policy that they really dislike, they likely rate implementation with the lowest score, so that they cannot give an even lower score to experimentation. Likewise, when they really like a policy, they likely rate no implementation with the lowest score, so that they cannot give an even lower score to experimentation. As a result, some of the above results may underestimate the true experimentation aversion. However, the data show that this underestimation is modest. In only 8% of the evaluated scenarios, respondents rate both implementation and experimentation as “very bad” or rate both not implementation and experimentation as “very bad”.

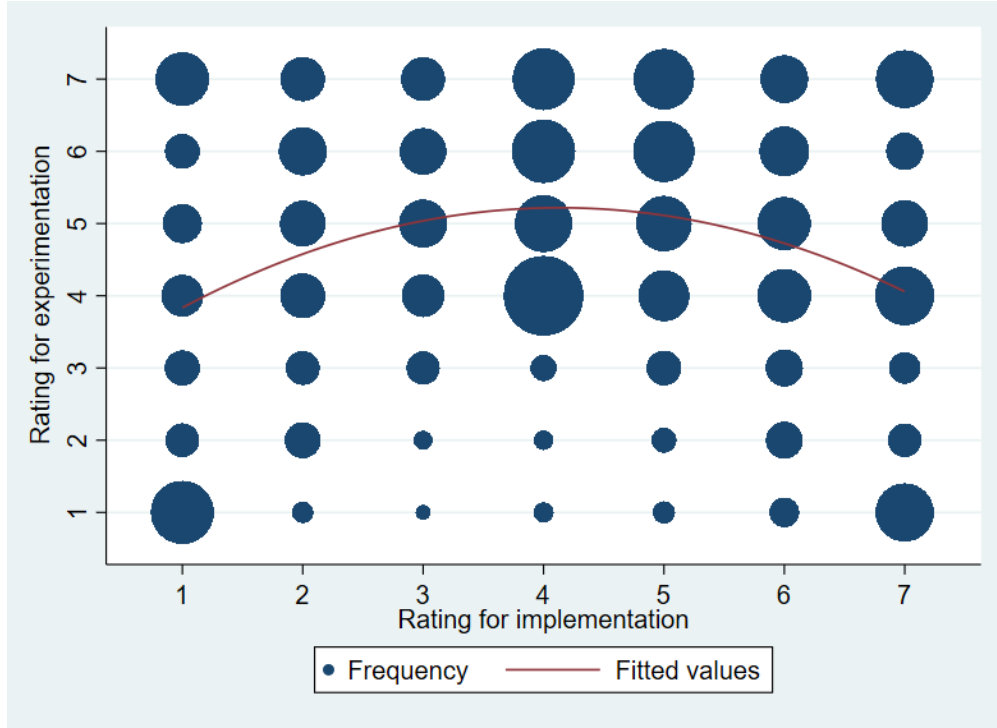
Figure 3: Preferences regarding experimentation



Note: The figure shows the percentage of respondents who rate experimentation highest (blue) and lowest (orange) – alone or together with implementation and/or no implementation – in none of the four scenarios, in one, two, three, or all of them. Details are described in the main text.

In Section 3 we hypothesized that voters may find a policy experiment particularly attractive when they do not hold a strong opinion about the policy. Figure 4 provides evidence supporting this hypothesis. The fitted regression line shows that experimentation is rated highest when respondents find implementation neither good nor bad. Experimentation is rated lowest when respondents find implementation either very bad or very good.¹⁶

Figure 4: Relation between rating for implementation and experimentation



Note: The figure shows the joint distribution of ratings for implementation and experimentation for all 8,456 evaluations made. The size of each dot indicates the frequency. The curve plots fitted values based on a regression of rating for experimentation on a constant (coefficient: 2.8; t-value: 22.5), the rating for implementation (coefficient: 1.2; t-value: 18.2), and the rating for implementation squared (coefficient: -0.14 ; t-value: -18.2). We also estimated a regression with individual fixed effects and find similar results. An interactive 3-D version of Figure 4 including the respondents' assessment of 'no implementation' is available online: https://osf.io/mnjv3/?view_only=bef9186a07414de489cc71e37c0e49c7.

¹⁶Note that a non-negligible fraction of respondents with a strong opinion about implementation (rating 1 or 7) rate experimentation highly. In the spirit of Callander and Hummel (2014) and Millner, Ollivier, and Simon (2014), one possible reason for why these respondents appreciate policy experiments is that they want to prove the (lack of) effectiveness of the policy to others.

4.2 General questions

The questions about the scenarios were followed by a number of general questions about respondents' attitudes towards political parties that propose policy experiments, their beliefs about others' attitudes (voters and politicians), and the concerns they may have regarding policy experiments. Table 2 shows that a majority of voters (60%) would approve it if one of their favorite political parties would be in favor of conducting policy experiments much more often. Only one out of six voters would consider it as bad. Appreciation is particularly high among the highly educated. Many voters also state that they would more likely vote for the party in the next elections because the party is in favor of more policy experiments, see the lower part of Table 2. In the full sample, the modal voter agrees with this; among the highly educated, almost half does so. Only one out of five voters say they would less likely vote for the political party because it favors more policy experiments.

What do voters believe about others' appreciation of policy experiments? Table C.1 in Appendix C shows that voters tend to be somewhat too pessimistic. The mean, median, and modal respondent believes that appreciation of others is 4 on the seven-point scale, whereas it is actually 5 (or, more specifically, 4.7). Moreover, almost two thirds of respondents believe that the party will not get more votes because it is in favor of more policy experiments. We also asked respondents to indicate why they think politicians are not often in favor of policy experiments. Many respondents think that politicians believe that voters will not appreciate policy experiments and less likely vote for their party if they would call for policy experiments, see the bottom part of Table C.1 in Appendix C.

Table 2: Attitudes towards political parties that propose policy experiments

<i>Suppose that at the next elections one of your favorite political parties is in favor of conducting policy experiments much more often. . . .</i>	Bad	Neither bad nor good	Good	Average score
<i>What do you think about the fact that this political party is in favor of more policy experiments?</i>				
Full sample:	16%	24%	60%	4.7
Highly educated:	12%	17%	70%	5.0
Not highly educated:	19%	29%	52%	4.5
	Less likely	Neither less nor more likely	More likely	Average score
<i>Has it become more likely or less likely that you vote for this party in the next elections because they are in favor of more policy experiments?</i>				
Full sample:	20%	40%	40%	4.3
Highly educated:	15%	39%	46%	4.5
Not highly educated:	23%	40%	37%	4.2

Note: Answers to both questions were given on a seven-point scale, where score 4 indicated “neither bad nor good” and “neither less nor more likely”, respectively.

Lastly, respondents assessed four often-raised concerns against policy experiments: fairness concerns, lack of informed consent, lack of external validity, and the time it takes to conduct them. We started the question by writing that “Some people are concerned about policy experiments.” and then asked to indicate in how far the respondent agrees or disagrees with the following concerns:

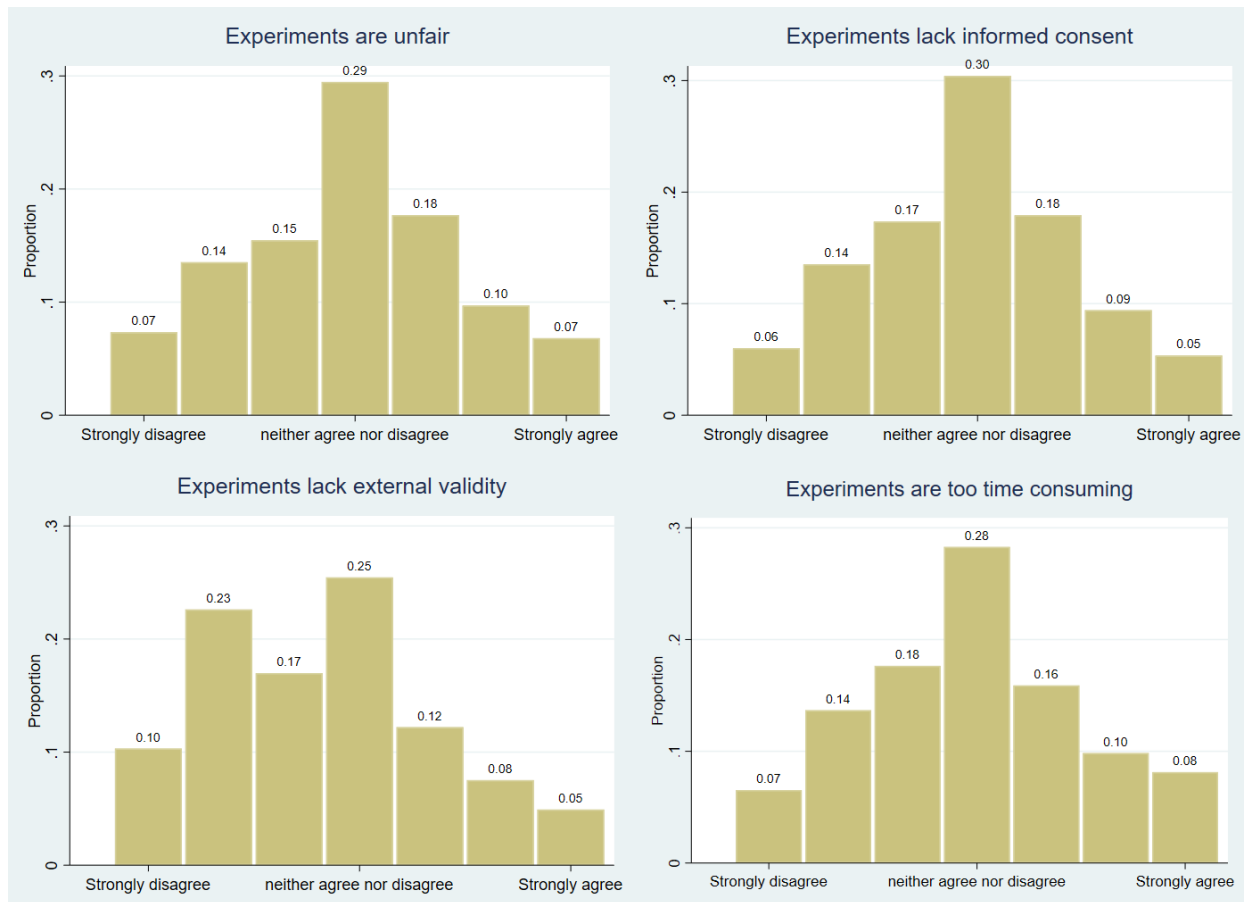
1. “Policy experiments are unfair, because people or companies are treated unequally for a period of time.”
2. “Policy experiments are not good, because people or companies are often not informed and have not always agreed to take part in a policy experiment.”
3. “You can learn little from a policy experiment, because the situation is different everywhere and the world continuously changes.”
4. “A policy experiment takes too much time, it is better to take action immediately.”

Figure 5 shows the distributions of responses. External validity is considered a concern by only one quarter. For the other three concerns, about one third of the respondents agrees at least to some extent.¹⁷ How important are these concerns for voters’ overall appreciation of policy experiments? Appendix D shows the results of regressing respondent’s general judgement of policy experiments (as measured by the variables presented in Table 2) on the respondent’s agreement with the four concerns. The coefficient for each concern is negative, sizeable, and highly statistically significant. Together the four concerns explain 25% of the variation in voters’ appreciation of policy experiments, suggesting that they are important drivers of attitudes towards policy experiments.¹⁸

¹⁷Highly educated respondents are less concerned than not highly educated regarding all four issues. These results are available upon request.

¹⁸The bivariate correlations between the concerns range between 0.42 and 0.62.

Figure 5: Concerns about policy experiments



Note: See details in the main text.

5 Survey experiment among politicians

5.1 Set-up of the experiment

Nine months after the data about voters had been collected, we ran a survey experiment among Dutch politicians¹⁹. The purpose of the experiment – which we preregistered in the AEA RCT Registry (AEARCTR-0009142) – is to find out whether politicians’ opinions about policy experiments change when they learn what voters think about policy experimentation. Politicians may be receptive to voters’ opinions either for electoral reasons or because they act as ‘delegates’ (as opposed to ‘trustees’) in the sense of Fox and Shotts (2009).

We invited all members of Dutch national parliament (both first and second chamber) as well as all members of the regional parliaments (‘Provinciale Staten’) to take part in our survey by sending an e-mail from the university e-mail address of the first author.²⁰ The e-mail – which can be found in Appendix E – mentioned that the survey consisted of only four short questions. We also wrote that during the survey we would provide some results of the representative survey we held among Dutch voters about “a few policy issues”, including results specific for their party’s voters. We further made clear that answers would be fully anonymous.

The questionnaire started with a brief description of the survey that we conducted among voters (see Appendix F for the entire questionnaire). To give politicians a good impression of what we asked the voters, we included the full text of the scenario about combating tax evasion, as shown in Section 4.1. We also mentioned the topics of the other scenarios and we included the exact text of the first general question that we asked to the voters, the one described in the top panel of Table 2, measuring the voter’s general appreciation for policy experiments. We then asked the politicians: What do you think that the average Dutch

¹⁹Prior to the experiment, we received approval for running the experiment from the Erasmus School of Economics IRB-Experimental (ETH2122-0500).

²⁰We excluded members of the regional parliaments who are affiliated to political parties that are not active in the national parliament because we have no information about their voters and, hence, are unable to treat them. The political parties included in our sample are shown in Figure 6. We erroneously did not include the province of Noord-Brabant in the data collection.

person responded to this question? And: What do you think that those who voted for your party last year on average responded to this question? We asked these questions to learn about politicians' prior beliefs about voters' opinion.

Next, the politicians who were assigned to the treatment group were informed about the actual responses of voters in general and of voters for their party. Specifically, we informed them about the voters' average appreciation on the seven-point scale, the percentage of voters that rate more policy experimentation as good, and the percentage that rate it as bad. After that, we asked politicians what they think about policy experiments themselves. Politicians' appreciation is measured in two ways: their assessment of a proposal by a fellow party member to pursue policy experiments much more often, and their willingness to actively support the proposal. Both are measured on a seven-point scale. Politicians who were assigned to the control group answered the exact same questions, but were informed about actual voters' responses at the end of the survey (that is, after they completed all four questions). We made sure that politicians could not return to an earlier screen and revise their answers after the information was provided. Assignment to the treatment and control group was random with stratification for party affiliation.

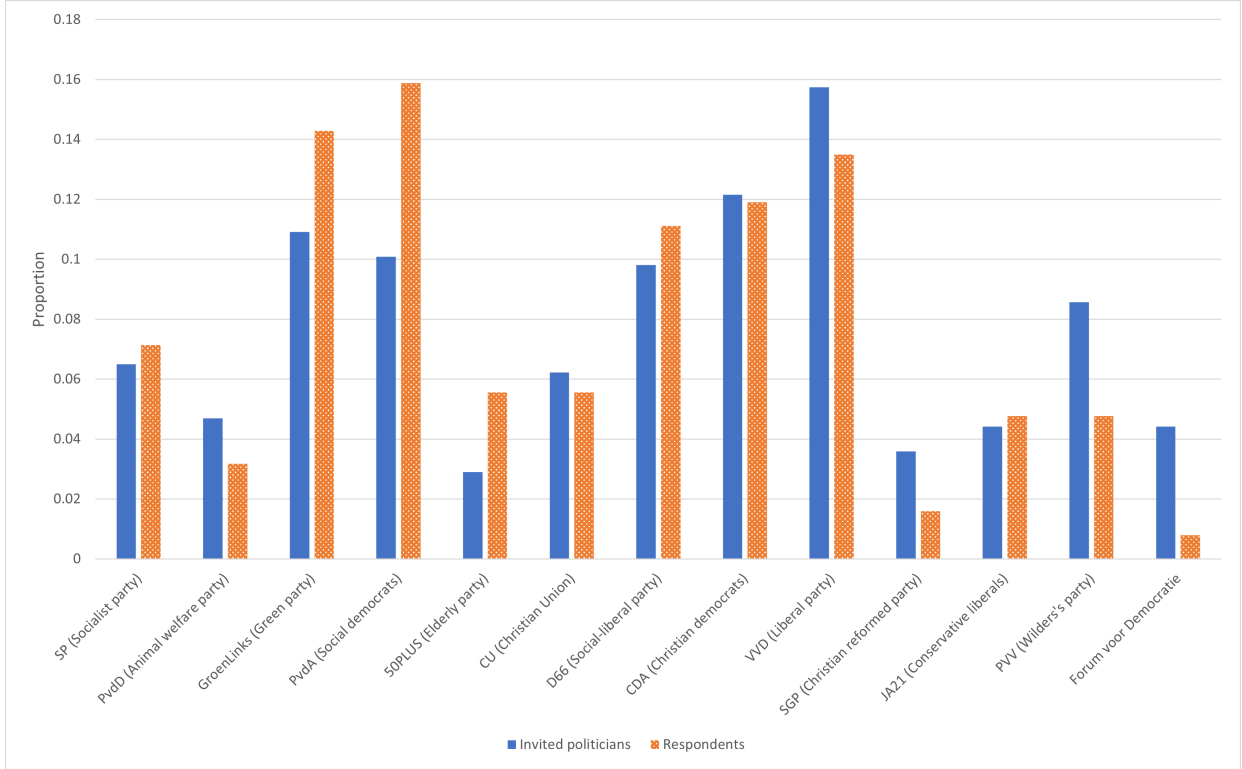
Our key prediction is that politicians who are too pessimistic about voters' appreciation of policy experimentation *ex ante* and are informed about actual voters' appreciation will report higher appreciation of policy experimentation themselves as compared to similar others in the control group. Conversely, politicians who are too optimistic *ex ante* and are informed about actual voters' appreciation will report lower appreciation of policy experimentation themselves as compared to similar others in the control group. We do not expect any treatment effect for politicians whose beliefs are already accurate, since the treatment does not provide any news to them.

5.2 Results of the experiment

We sent the invitation to take part in the survey on 2 and 3 June 2022 to 725 politicians. Unfortunately, 46 invitations were returned because the e-mail addresses were invalid. In total 126 politicians completed the survey, which implies a response rate of 19%.²¹ While we received responses from delegates of all political parties, we did not achieve full representativeness in this dimension, see Figure 6. Several left-wing parties are overrepresented, while some right-wing parties are underrepresented – especially the radical right-wing parties. Moreover, we obviously cannot rule out selectivity in other dimensions. In particular, we suspect that the stated aim of the survey (academic research) as well as the identity of the sender (a professor at a university, the first author of this study) may have led politicians who hold a more favorable view of academic research to respond in larger numbers. We should keep this in mind, particularly when looking at the descriptive results.

²¹We closed the survey after six days, on 8 June 2022. We did not send any reminder to non-respondents.

Figure 6: Party affiliation of politicians in the survey experiment



Note: The blue bars display the distribution of the party affiliation of invited politicians. The orange bars display the distribution of the party affiliation of the politicians that took part in our survey experiment. Parties are loosely ordered from the political left to right. See further details in the main text.

The data show that there is quite some variation in politicians' beliefs about what the average voter thinks about policy experiments. Specifically, 23% believe that voters have a negative view of policy experiments, 12% believe that voters are neutral (i.e. score 4 on the seven-point scale), and 65% believe that voters have a positive view. However, all in all, politicians in our sample are neither too pessimistic nor too optimistic about voters' views: on average they think voters give policy experiments a rating of 4.73, which is exactly right. Regarding their own parties' voters, politicians are on average only 0.04 points too pessimistic. Asked for their own opinion, politicians in the control group express even slightly higher appreciation than voters: their average appreciation is 5.15. However, as argued above, this estimate may be biased upward due to the non-representativeness of our sample

of politicians.

To investigate the effect of the information treatment and how it is affected by politicians' prior beliefs, we regress politician's appreciation for policy experiments on a treatment dummy, the difference in the politician's belief about voters' appreciation and voters' actual appreciation, and the interaction between the treatment dummy and the latter variable. The regression results are shown in the first column of Table 3. Three results are noteworthy. First, in line with our prediction, the estimated treatment effect for politicians whose belief about voters' appreciation is equal to the actual voters' appreciation is close to zero (0.03). Second, there is a strong positive correlation between politicians' opinion about policy experiments and their belief about voters' opinion. Among politicians in the control group, one point increase in politicians' optimism is associated with 0.69 points increase in their appreciation for policy experiments. Hence, politicians tend to think that voters share their opinion on policy experiments. Third, politicians become much more positive about policy experiments when they learn that their beliefs about voters' appreciation are too pessimistic. For each point difference, they increase their appreciation by 0.34 points. Conversely, politicians who are too optimistic adjust their opinion in downward direction.

We also measured politicians' appreciation for policy experiments by their willingness to actively support a proposal to conduct policy experiments much more often. Regression results on this alternative measure are reported in column 2 of Table 3. The results are very similar, except that providing information to politicians with erroneous beliefs no longer has a statistically significant effect on their appreciation of policy experiments. In columns 3 and 4, we run the same regressions as in columns 1 and 2 but now including the politicians' beliefs about their own party's voters rather than voters on average.²² Results are quite similar. Naturally, the correlation between politicians' opinion about policy experiments and their belief about voters' opinion is even stronger here.

²²Beliefs about own party's voters and about voters in general are strongly correlated. The correlation coefficient is 0.60. Including both in a single regression gives rise to multicollinearity.

Table 3: Regression of politicians' appreciation and active support for policy experiments

	(1)	(2)	(3)	(4)
	Appreciation	Active support	Appreciation	Active support
Treatment	0.03 (0.20)	0.26 (0.22)	0.12 (0.17)	0.36* (0.20)
Belief difference	0.69*** (0.09)	0.65*** (0.10)	0.85*** (0.08)	0.80*** (0.10)
Treatment \times Belief difference	-0.34** (0.17)	-0.17 (0.18)	-0.22* (0.13)	-0.10 (0.15)
Constant	5.10*** (0.13)	4.89*** (0.15)	5.09*** (0.11)	4.88*** (0.13)
Observations	126	126	126	126
R-squared	0.32	0.28	0.52	0.44

Note: The dependent variable in columns (1) and (3) is politician's appreciation of a call for doing more policy experiments, while the dependent variable in columns (2) and (4) is the politician's willingness to actively support such a call. "Treatment" is a dummy indicating whether the politician was informed about voters' opinion on policy experiments before expressing his/her opinion. In columns (1) and (2), "Belief difference" is the difference between the politician's belief about voters' appreciation and voters' average actual appreciation (average = 0.00, standard deviation = 1.35). In columns (3) and (4), "Belief difference" is the difference between the politician's belief about his/her own party's voters' appreciation and their actual appreciation (average = -0.04, standard deviation = 1.29). Robust standard errors in parentheses. Stars indicate statistical significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

6 Concluding remarks

Why don't we see many more policy experiments? Our paper provides convincing evidence that the reason lies not with unpopularity of policy experiments among voters. On the contrary, policy experiments get broad support from a vast majority of voters. Many voters also indicate that they are more likely to vote for a politician who proposes to run policy experiments much more often. Concerns that are often raised about policy experiments, such as fairness concerns and lack of external validity, are seen as problematic by only about a third of voters.

We informed national and regional politicians about these results. Some of them were informed by us before we asked their opinion about policy experiments, while randomly selected others were informed only after they expressed their own opinion. We find that politicians are quite responsive to the information and conform to voters' opinion to a large extent. These results make us optimistic that when the results of our study become more widely known, politicians will propose and conduct policy experiments more often.

One weakness of our study is that our sample of politicians is relatively small and not representative. While we succeeded in creating a sample that includes politicians from all parties, the radical right is strongly underrepresented and the social democrats and green party are overrepresented. Moreover, we cannot rule out selective non-response within political parties. In particular, it might be that politicians who have a more positive view of the social sciences responded in larger numbers. As a result, our data may not provide an unbiased estimate of what politicians think about policy experiments. Also, the external validity of our information experiment is uncertain. Future research should make progress in achieving larger and more representative samples of politicians.

While our study has focused on voters' experimentation aversion as a possible reason for the limited use of policy experiments, we can think of several other important reasons. First, the market for policy evaluation may not always work well due to asymmetric information. For instance, it is not always clear for non-experts when a policy experiment is feasible and

when it is preferable to other, cheaper types of evaluation, such as a non-randomized pilot study, simple before-after comparison of outcomes, or asking experts for their opinion. Second, once in office, politicians may refrain from conducting a policy experiment because of reputational concerns. In particular, politicians may fear the consequences for their reputation if the policy experiment reveals that the policy they proposed turns out to be ineffective. Risk averse politicians may therefore favor a less powerful evaluation method, as it enables them to refute unwelcome evidence and thus protect their reputation as a competent policy maker. We leave it for future research to shed light on the relevance of these considerations.

References

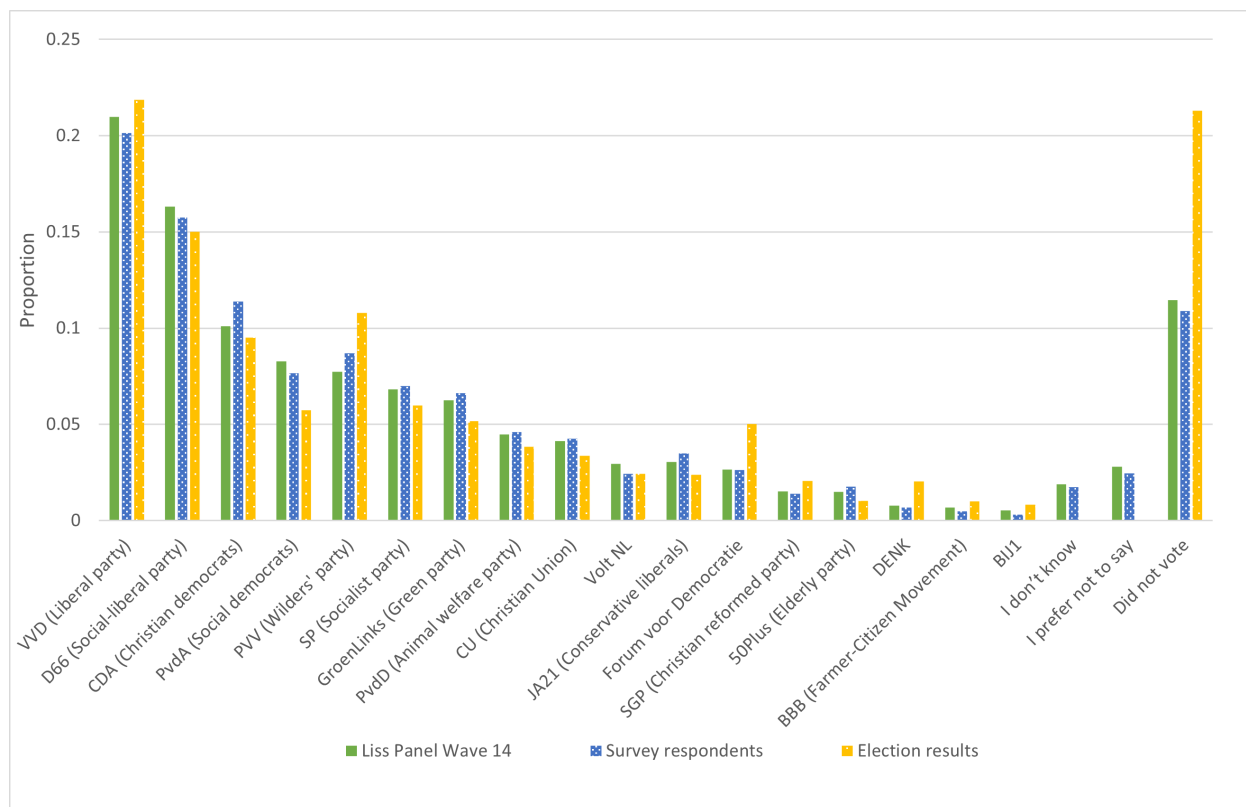
- Asiedu, E., Karlan, D., Lambon-Quayefio, M., & Udry, C. (2021). A call for structured ethics appendices in social science papers. *Proceedings of the National Academy of Sciences*, 118(29), e2024570118.
- Bernecker, A., Boyer, P. C., & Gathmann, C. (2021). The role of electoral incentives for policy innovation: Evidence from the us welfare reform. *American Economic Journal: Economic Policy*, 13(2), 26–57.
- Broockman, D. E., & Skovron, C. (2018). Bias in perceptions of public opinion among political elites. *American Political Science Review*, 112(3), 542–563.
- Butler, D. M., Nickerson, D. W., et al. (2011). Can learning constituency opinion affect how legislators vote? results from a field experiment. *Quarterly Journal of Political Science*, 6(1), 55–83.
- Callander, S., & Harstad, B. (2015). Experimentation in federal systems. *Quarterly Journal of Economics*, 130(2), 951–1002.
- Callander, S., & Hummel, P. (2014). Preemptive policy experimentation. *Econometrica*, 82(4), 1509–1528.
- Campbell, D. T. (1969). Reforms as experiments. *American Psychologist*, 24(4), 409.
- Cardon, E. B. (2022). *Public administrator aversion to randomized controlled trials*. Syracuse University, Dissertations Publishing.
- Charness, G., Samek, A., & Van de Ven, J. (2022). What is considered deception in experimental economics? *Experimental Economics*, 25(2), 385–412.
- Corduneanu-Huci, C., Dorsch, M. T., & Maarek, P. (2021). The politics of experimentation: Political competition and randomized controlled trials. *Journal of Comparative Economics*, 49(1), 1–21.
- DellaVigna, S., Kim, W., & Linos, E. (2022). Bottlenecks for evidence adoption. *NBER Working Paper 30144*.
- Desposato, S. (2018). Subjects and scholars’ views on the ethics of political science field experiments. *Perspectives on Politics*, 16(3), 739–750.
- Dur, R. (2001). Why do policy makers stick to inefficient decisions? *Public Choice*, 107(3), 221–234.
- Fischer, M., Grewenig, E., Lergetporer, P., & Werner, K. (2019). The e-word: On the public acceptance of experiments. *Mimeo*.
- Fox, J., & Shotts, K. W. (2009). Delegates or trustees? a theory of political accountability. *Journal of Politics*, 71(4), 1225–1237.
- García-Sánchez, M., Matanock, A. M., & Garbiras-Díaz, N. (2022). Do citizens’ preferences matter? shaping legislator attitudes towards peace agreements. *Journal of Conflict Resolution*, forthcoming.
- Glennerster, R. (2017). The practicalities of running randomized evaluations: Partnerships, measurement, ethics, and transparency. In *Handbook of economic field experiments* (Vol. 1, pp. 175–243). Elsevier.
- Glennerster, R., & Takavarasha, K. (2013). *Running randomized evaluations: A practical guide*. Princeton University Press.
- Heck, P. R., Chabris, C. F., Watts, D. J., & Meyer, M. N. (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, O., & List, J. A. (2021). Who’s afraid of evidence-based policymaking? *Project Syndicate*.
- Hjort, J., Moreira, D., Rao, G., & Santini, J. F. (2021). How research affects policy: Experimental evidence from 2,150 brazilian municipalities. *American Economic Review*, 111(5), 1442–80.
- John, P., Kim, K., & Soto-Tamayo, L. (2023). Using citizen voice to evaluate experiments on politicians: A uk survey experiment. *Journal of Experimental Political Science*, 1–10.
- Kalla, J. L., & Porter, E. (2021). Correcting bias in perceptions of public opinion among american elected officials: Results from two field experiments. *British Journal of Political Science*, 51(4), 1792–1800.
- List, J. A. (2008). Informed consent in social science. *Science*, 322(5902), 672.
- Mazar, N., Elbaek, C. T., & Mitkidis, P. (2023). Experiment aversion does not appear to generalize. *Proceedings of the National Academy of Sciences*, 120(16), e2217551120.
- McAndrews, J., Lucas, J., Merkley, E., Rubenson, D., & Loewen, P. J. (2022). Do politicians and citizens agree on how to combat covid-19? evidence from three studies. *OSF Preprint*.
- McDermott, R., & Hatemi, P. K. (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, S., Naseer, S., & Chen, D. L. (2021). Training policymakers in econometrics. *Mimeo*.
- Meyer, M. N., Heck, P. R., Holtzman, G. S., Anderson, S. M., Cai, W., Watts, D. J., & Chabris, C. F. (2019a). Objecting to experiments that compare two unobjectionable policies or treatments. *Proceedings of the National Academy of Sciences*, 116(22), 10723–10728.
- Meyer, M. N., Heck, P. R., Holtzman, G. S., Anderson, S. M., Cai, W., Watts, D. J., & Chabris, C. F. (2019b). Reply to mislavsky et al.: Sometimes people really are averse to experiments. *Proceedings of the National Academy of Sciences*, 116(48), 23885–23886.
- Millner, A., Ollivier, H., & Simon, L. (2014). Policy experimentation, political competition, and heterogeneous beliefs. *Journal of Public Economics*, 120, 84–96.
- Mislavsky, R., Dietvorst, B., & Simonsohn, U. (2020). Critical condition: People don’t dislike a corporate experiment more than they dislike its worst condition. *Marketing Science*, 39(6), 1092–1104.
- Mislavsky, R., Dietvorst, B. J., & Simonsohn, U. (2019). The minimum mean paradox: A mechanical explanation for apparent experiment aversion. *Proceedings of the National Academy of Sciences*, 116(48), 23883–23884.
- Naurin, E., & Öhberg, P. (2021). Ethics in elite experiments: a perspective of officials and voters. *British Journal of Political Science*, 51(2), 890–898.
- Phillips, T. (2021). Ethics of field experiments. *Annual Review of Political Science*, 24, 277–300.
- Sevenans, J. (2021). How public opinion information changes politicians’ opinions and behavior. *Political Behavior*, 43(4), 1801–1823.

- Soontjens, K., & Sevenans, J. (2022). Electoral incentives make politicians respond to voter preferences: Evidence from a survey experiment with members of parliament in belgium. *Social Science Quarterly*, 103(5), 1125–1139.
- Strulovici, B. (2010). Learning while voting: Determinants of collective experimentation. *Econometrica*, 78(3), 933–971.
- Vivalt, E., & Coville, A. (2022). How do policy-makers update their beliefs? *Mimeo*.
- Wang, S., & Yang, D. Y. (2021). Policy experimentation in china: The political economy of policy learning. *NBER Working Paper 29402*.
- Yokoo, H.-F. (2020). Ethics of randomized field experiments: Evidence from a randomized survey experiment. *Mimeo*.

A Appendix

Figure A.1: Election results: representativeness of the sample



Note: The figure shows the vote share for all parties that obtained seats in national parliament in the elections in 2021 according to responses in Wave 14 of the LISS Panel (green bars), according to responses by people who also took part in our survey (blue bars), and according to the official election results (yellow bars, source: <https://www.verkiezingsuitslagen.nl/verkiezingen/detail/TK20210317>).

B Appendix

This appendix describes all six scenarios used in the survey among voters. The original text shown to the respondents was in Dutch and can be downloaded from: https://www.dataarchive.lisssdata.nl/hosted_files/download/7215.

Scenario 1: Combating tax evasion by large companies

In order to combat tax evasion by big companies, the Minister of Finance is considering to spend more money on increased monitoring. It is uncertain what the effects of this will be. The Minister can do three things:

1. **Implement:** The Minister decides to spend more money on increased monitoring of all big companies.
2. **Not implement:** The Minister decides *not* to spend more money on increased monitoring of big companies.
3. **Experiment:** The Minister decides to conduct a policy experiment, in which a random selection of the big companies will be monitored more closely, while the others continue to be monitored the same as before. After the policy experiment, the two groups will be compared to see whether increased monitoring was successful and should be introduced for all companies.

What do you think about:

Implement? [1=Very bad; 2; 3; 4; 5; 6; 7=Very good]

Not implement? [1=Very bad; 2; 3; 4; 5; 6; 7=Very good]

Experiment? [1=Very bad; 2; 3; 4; 5; 6; 7=Very good]

Scenario 2: Helping the unemployed find vacancies

The board of the public employment services is contacted by a company that can help unemployed people find vacancies. It is uncertain what the effects of this help will be. The board can do three things:

1. **Implement:** The board decides to hire the company to help all unemployed people.
2. **Not implement:** The board decides *not* to hire the company.

3. **Experiment:** The board decides to conduct a policy experiment, in which the company helps a random selection of unemployed people to find vacancies, whereas the others do not receive help from the company. After the policy experiment, the two groups will be compared to see whether the help was successful and should be implemented for all unemployed people.

What do you think about:

Implement? [1=Very bad; 2; 3; 4; 5; 6; 7=Very good]

Not implement? [1=Very bad; 2; 3; 4; 5; 6; 7=Very good]

Experiment? [1=Very bad; 2; 3; 4; 5; 6; 7=Very good]

Scenario 3: Anonymous job applications

A municipal alderman wants to prevent discrimination and is thinking about introducing anonymous job applications for municipal job openings. This means that when choosing candidates for job interviews, the name, gender, and age of the candidates will not be visible. It is uncertain what the effects of this will be. The alderman can do three things:

1. **Implement:** The alderman decides to introduce anonymous job applications.
2. **Not implement:** The alderman decides *not* to introduce anonymous job applications.
3. **Experiment:** The alderman decides to conduct a policy experiment, in which anonymous job applications are introduced for a random selection of job openings, whereas the selection procedure for other job openings remains the same as before. After the policy experiment, the two groups will be compared to see whether anonymous applications were successful and should be implemented.

What do you think about:

Implement? [1=Very bad; 2; 3; 4; 5; 6; 7=Very good]

Not implement? [1=Very bad; 2; 3; 4; 5; 6; 7=Very good]

Experiment? [1=Very bad; 2; 3; 4; 5; 6; 7=Very good]

Scenario 4: Detecting benefit fraud

A municipal alderman is considering spending more money on detecting benefit fraud.

It is uncertain what the effects of this will be. The alderman can do three things:

1. **Implement:** The alderman decides to spend more money on detecting benefit fraud.
2. **Not implement:** The alderman decides *not* to spend more money on detecting benefit fraud.
3. **Experiment:** The alderman decides to conduct a policy experiment, in which extra money is spent on detecting benefit fraud among a random selection of benefit recipients, whereas fraud detection efforts for other benefit recipients remain the same as before. After the policy experiment, the two groups will be compared to see whether spending extra money on the detection of benefit fraud was successful and should be implemented for all benefit recipients.

What do you think about:

Implement? [1=Very bad; 2; 3; 4; 5; 6; 7=Very good]

Not implement? [1=Very bad; 2; 3; 4; 5; 6; 7=Very good]

Experiment? [1=Very bad; 2; 3; 4; 5; 6; 7=Very good]

Scenario 5: “Very Irritating Police” approach

The Minister of Justice and Security is considering introducing a “Very Irritating Police” approach for nuisance-causing loitering youths and asylum seekers. Nuisance causers will be constantly hounded and subjected to inspections. It is uncertain what the effects of this will be. The minister can do three things:

1. **Implement:** The minister decides to implement the “Very Irritating Police” approach.
2. **Not implement:** The minister decides *not* to implement the “Very Irritating Police” approach.
3. **Experiment:** The Minister decides to conduct a policy experiment, in which the “Very Irritating Police” approach is implemented in a random selection of Dutch police districts, whereas the other police districts will continue to use the same approach as before. After the policy experiment, the two groups will be compared to see whether the “Very Irritating Police” approach was successful and should be implemented everywhere.

What do you think about:

Implement? [1=Very bad; 2; 3; 4; 5; 6; 7=Very good]

Not implement? [1=Very bad; 2; 3; 4; 5; 6; 7=Very good]

Experiment? [1=Very bad; 2; 3; 4; 5; 6; 7=Very good]

Scenario 6: Softer job application requirements for welfare recipients

A municipal alderman is considering softening the job application requirements for young people on welfare. They will not have to apply for as many job openings, and there will be fewer controls to check whether they are applying for jobs. It is uncertain what the effects of this will be. The alderman can do three things:

1. **Implement:** The alderman decides to soften the job application requirements.
2. **Not implement:** The alderman decides *not* to soften the job application requirements.
3. **Experiment:** The alderman decides to conduct a policy experiment, in which the job application requirements are softened for a random selection of young welfare recipients, whereas they remain the same for the other young welfare recipients. After the policy experiment, the two groups will be compared to see whether the softening was successful and should be implemented for all young welfare recipients.

What do you think about:

Implement? [1=Very bad; 2; 3; 4; 5; 6; 7=Very good]

Not implement? [1=Very bad; 2; 3; 4; 5; 6; 7=Very good]

Experiment? [1=Very bad; 2; 3; 4; 5; 6; 7=Very good]

C Appendix

Table C.1: Respondents' beliefs about opinions of the general public and politicians

<i>Imagine that in the next general elections, one of your preferred political parties is in favor of doing policy experiments much more often:</i>	Bad (1-3)	Neither bad nor good (4)	Good (5-7)	Average score
<i>What do you think the average Dutch citizen thinks about the fact that this political party is in favor of more policy experiments?</i>	25%	37%	38%	4.16
	Fewer votes (1-3)	Neither more nor fewer (4)	More votes (5-7)	Average score
<i>Do you think this party will receive more or fewer votes in the next elections because they are in favor of more policy experiments?</i>	22%	43%	35%	4.13
<i>Politicians tend to not be in favor of doing policy experiments. How important do you think the following reasons are for this?</i>	Not important (1-3)	Neither important nor unimportant (4)	Important (5-7)	Average score
<i>Politicians think that voters will not appreciate policy experiments.</i>	22%	35%	43%	4.31
<i>Politicians think that voters will be less likely to vote for their party if their party calls for policy experiments.</i>	18%	35%	47%	4.49
<i>Politicians personally think that policy experiments will have few or no advantages, but many disadvantages.</i>	22%	37%	41%	4.37

Note: The table shows for each question the percentage of responses falling into an answer category. Classification into three answer categories is based on the answers given on the seven-point scale.

D Appendix

This Appendix shows the results of regressing respondent’s general judgement of policy experiments (as measured by the variables presented in Table 2) on the respondent’s agreement with the four concerns described in Section 4.2 and Figure 5.

Table D.1: Regression of voters’ appreciation on their concerns regarding policy experiments

	(1)	(2)
	Attitude	Vote intention
Fairness	-0.12*** (0.02)	-0.08*** (0.02)
Informed consent	-0.06** (0.03)	-0.05** (0.03)
External validity	-0.23*** (0.02)	-0.17*** (0.02)
Time consuming	-0.13*** (0.02)	-0.08*** (0.02)
Constant	6.73*** (0.10)	5.72*** (0.1)
Observations	2,114	2,114
R-squared	0.25	0.15

Note: The dependent variables in columns (1) and (2) are, respectively, the answers to the following questions: (1) Suppose that at the next elections one of your favorite political parties is in favor of conducting policy experiments much more often. What do you think about the fact that this political party is in favor of more policy experiments? and (2) Has it become more likely or less likely that you vote for this party in the next elections because they are in favor of more policy experiments? A higher score of an independent variable indicates stronger concerns about that issue. All answers are given on a seven-point scale. Stars indicate statistical significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

E Appendix

The full invitation e-mail to politicians reads as follows (translated from Dutch):

Subject: Study Erasmus University among voters and politicians

Dear [First Name] [Last Name],

We would like to invite you to take part in a brief survey consisting of 4 short questions. Taking part does not take longer than 3 minutes. All members of Provinciale Staten, the Second Chamber, and the First Chamber belonging to one of the main parties have been invited to take part.

During the survey, we will show you some results of a representative survey we conducted among Dutch voters about a few policy issues. We will also inform you about the results among the group of people that voted for your party last year.

Your answers cannot be linked to you personally. Your participation will be highly appreciated! Your answers are fully anonymous and will only be used and shared for research purposes. By clicking on the link, you give permission for this.

To the survey: [<https://erasmusuniversity.eu.qualtrics.com/jfe/form/XXXXXXXXX>]

If you have any questions, do not hesitate to contact me.

Yours sincerely,

Prof.dr. Robert Dur
Professor at Erasmus University Rotterdam

F Appendix

The original full survey text was in Dutch and is available online: https://osf.io/mnjv3/?view_only=bef9186a07414de489cc71e37c0e49c7.

The translated full survey text for politicians that were assigned to treatment reads as follows:

We are glad that you participate in this survey experiment that consists of 4 questions. Before asking the questions, let us tell you briefly about a representative survey we recently conducted among Dutch voters. We were particularly interested in the question of what voters think about policy experiments, i.e. trying out new policies on a small scale first and only introducing them if they turn out to be a success.

To show voters what we mean by policy experiments, we outlined several hypothetical scenarios. For example, one of them was about tax evasion:

In order to combat tax evasion by big companies, the Minister of Finance is considering to spend more money on increased monitoring. It is uncertain what the effects of this will be. The Minister can do three things:

1. Implement: The Minister decides to spend more money on increased monitoring of all big companies.
2. Not implement: The Minister decides not to spend more money on increased monitoring of big companies.
3. Experiment: The Minister decides to conduct a policy experiment, in which a random selection of the big companies will be monitored more closely, while the others continue to be monitored the same as before. After the policy experiment, the two groups will be compared to see whether increased monitoring was successful and should be introduced for all companies.

Voters were asked to indicate on a 7-point scale (ranging from “very bad” to “very good”) what they think of each of those three options (implement, not implement, experiment).

There were six such scenarios in total. The other five topics were a training program for the unemployed, anonymous applications to government jobs, detecting benefit fraud, the “Very Irritating Police” approach, and softening job application requirements for welfare recipients.

After the scenarios, voters were asked a number of general questions, including this one:

Suppose that at the next elections one of your favourite political parties is in favour of conducting policy experiments much more often. What do you think about the fact that this political party is in favour of more policy experiments?

Voters were asked to indicate on a 7-point scale (from “very bad” to “very good”) what they think about this.

What do you think the average Dutch person thinks about this political party being in favour of more policy experiments?

[Answer on a 7-point scale, from “very bad” to “very good”]

On average, what do you think those who voted for your political party last year think about this political party being in favour of more policy experiments?

[Answer on a 7-point scale, from “very bad” to “very good”]

New screen

Before we go to the last two questions, we would like to inform you about what the Dutch voters in general and those who voted for your party last year think about the call for more policy experiments. Dutch voters gave an average of 4.7 on the 7-point scale.

60% thinks it is good.

16% thinks it is bad.

People who voted for your party in 2021 gave an average of X on the 7-point scale.

Y% thinks it is good.

Z% thinks it is bad.

Suppose someone in your party advocates doing policy experiments much more often. What do you think about this call for more policy experiments?

[Answer on a 7-point scale, from “very bad” to “very good”]

How likely are you to actively support the call for more policy experiments?

[Answer on a 7-point scale, from “very unlikely” to “very likely”]

New screen

Thank you for answering the questions. Would you like to respond or would you like more information about policy experiments? Please contact Professor Robert Dur via email (dur@ese.eur.nl).

The politicians that were assigned to the control group completed the exact same questionnaire, except that the information on voters’ opinions was given at the end of the survey on the last screen, right before the thank you message, and that the message started with “Finally, we would like to inform you” instead of “Before we go to the last two questions, we would like to inform you”.

Respondents were required to answer all questions displayed on the screen before they could proceed to the next screen. It was not possible to return to a previous screen to rule out that respondents revise their answers after learning the information we provided them.