TU Delft

Delft University of Technology

Document Version
Final published version

Licence
CC BY

Citation (APA)
Goes, K., Thépaut, L., Mouter, N., Chappin, E., & Kluiving, S. (2027). Participation builds trust, not framing: insights from a national energy-transition experiment. *Ecological Economics*, *251*, Article 109232. https://doi.org/10.1016/j.ecolecon.2026.109232

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
In case the licence states “Dutch Copyright Act (Article 25fa)”, this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.

Ecological Economics 251 (2027) 109232

Contents lists available at ScienceDirect

Ecological Economics

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

Analysis

Participation builds trust, not framing: insights from a national energy-transition experiment

Kevin Goes [a,*], Louison Thépaut [a], Niek Mouter [b], Emile Chappin [b], Sjoerd Kluiving [c]

[a] Vrije Universiteit Amsterdam, Institute for Environmental Studies (IVM), Netherlands
[b] Delft University of Technology, Faculty of Technology, Policy and Management, Netherlands
[c] Vrije Universiteit Amsterdam, Faculty of Social Sciences and Humanities, Netherlands

ARTICLE INFO

Keywords:
Institutional trust
Value-based framing
Citizen engagement
Participatory value evaluation
Energy transition

ABSTRACT

Institutional trust is a critical determinant of public cooperation in the energy transition, yet little is known about how different framings of government communication influence trust in practice. This study reports results from a large-scale national experiment (N ≈ 3000) embedded in a Participatory Value Evaluation (PVE) on heat-transition policy. Participants were randomly assigned to one of six value-based framing conditions, or a control group: environmental, financial, or stability messaging, each presented by either local or national government. Trust in government was measured before and after the intervention, enabling a Difference-in-Differences analysis complemented by Latent Profile Analysis and Honest Causal Forests.

Across all specifications, value-based framing failed to increase trust and, in some cases, reduced it, with the local-stability frame showing the most consistent negative effect. Trust responses varied strongly across latent trust profiles but not across demographic groups, indicating that predispositions toward government, rather than socio-demographic characteristics, drive communication sensitivity. In contrast, participation in the PVE itself produced a modest and robust increase in trust independent of the framing treatment. These findings suggest that participatory decision tools may strengthen trust, whereas value-laden communication risks unintended negative effects.

1. Introduction

The energy transition requires substantial public engagement and behavioral change. Trust plays a critical role in ensuring the success of energy transition policies, as individuals and communities must believe in the credibility of the information provided and the intentions of policymakers (Otto et al., 2023). Mistrust can lead to skepticism, resistance, and lower support of energy-saving initiatives. Understanding how trust is influenced and by whom information is communicated is therefore essential for the design of legitimate and effective energy policies. This study examines how value-based framing and the level of government communicating a message influence citizens' institutional trust, specifically in the context of heat grid policies.

Although there is debate among scientists about how to define trust, the work by Otto et al. (2023) has identified some alignment in definitions. They follow the alignment from Earle and Siegrist (2008) and Siegrist et al. (2011), where trust is determined by competence, integrity and later also with benevolence. Competence indicates the trust in the experience and expertise of an organization, whereas integrity indicates the perceived honesty, openness and concern (Otto et al., 2023). Grimmelikhuijsen and Knies (2017) have further operationalized trust with a third dimension, which is benevolence, based on the work of McKnight et al. (2002). Benevolence indicates the perceived kind-heartedness of an institution. In this study, we adopt the operationalization of trust proposed by Grimmelikhuijsen and Knies (2017).

Trust is a cornerstone of effective communication in governance and public policy (Grimmelikhuijsen, 2012). As societies transition from fossil fuels to renewable energy sources, trust becomes even more crucial in ensuring public acceptance and engagement with energy policies. Institutions such as local and national governments rely on public trust to communicate policies, promote initiatives, and foster compliance. However, institutional trust is not only shaped by the content of the communicated message, but also by the perceived credibility and salience of its source (Metzger and Flanagin, 2013).

Efforts to strengthen trust in government communication are often linked to the broader literature on nudging, following the framework

* Corresponding author.
E-mail address: k.goes@vu.nl (K. Goes).

https://doi.org/10.1016/j.ecolecon.2026.109232
Received 17 April 2026; Received in revised form 1 September 2026; Accepted 2 September 2026
Available online 10 September 2026
0921-8009/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

introduced by Thaler and Sunstein (2008). Rather than altering choices or imposing financial incentives, nudges modify the informational context in which people make decisions, drawing on heuristic decision processes to encourage certain actions. Within energy-policy settings, informational nudges have been shown to stimulate engagement in a relatively low-cost manner (Thépaut et al., 2024; Botzen et al., 2025; Gillingham and Tsvetanov, 2018). Yet these interventions depend heavily on the credibility of the messenger: if the source is not salient, even subtle informational cues may lose their persuasive power.

The role of trust in the effectiveness of informational nudges remains a topic of debate in the literature. While some studies, such as Holladay et al. (2016), suggest that both price-based and non-price interventions can shape behavior effectively, others, like Allcott and Greenstone (2017), argue that financial incentives play a more decisive role in increasing engagement. However, Bolderdijk et al. (2012) found that moral motives are more effective than price-based interventions in bolstering support for environmental campaigns. Conversely, Gillingham and Tsvetanov (2018) highlight the potential of informational nudges to drive participation, emphasizing the need for further research into the contexts in which such interventions are most effective. A key question is whether the salience of the source influences the impact of these nudges.

Existing literature indicates that the salience of information varies depending on whether it is presented by local or national governments (Bernardo and Palma-Oliveira, 2013). Place identity and place attachment, which depend on scale, influence the perceived trustworthiness of the source (Bernardo and Palma-Oliveira, 2013). Additionally, the source of information and the ownership of an issue affect information salience, either as mediators (de Vries and Pruyn, 2007) or as direct influences (Lefevere et al., 2017). The salience of the information itself also shapes how messages are received (Bettinger et al., 2020), and in the context of the energy transition this reception may be guided by the values that frames foreground (Bolderdijk et al., 2013). Yet despite the widespread use of informational nudges in environmental and energy policy, almost no studies examine whether such nudges can alter institutional trust. Trust is typically treated as a precondition for behavioral change rather than an outcome. As a result, little is known about how trust in the source interacts with the salience and value content of different types of information, or whether framing strategies can strengthen or erode trust in government communication (de Vries and Pruyn, 2007).

The study examines how value-based framing and the level of government communicating a message influence citizens' trust in public communication about the energy transition. Specifically, it investigates whether trust differs when messages originate from local versus national governments, and whether emphasizing different values affects perceived trust.

In addition to examining these relationships at the aggregate level, this paper explores whether responses to different combinations of value framing and government level vary across subgroups of the population. Such heterogeneity may reflect differences in value salience and pre-existing institutional orientations. By examining both aggregate and heterogeneous responses, the study contributes to a deeper understanding of when and how government communication fosters or erodes trust, and whether framing strategies can be tailored to sustain legitimacy in the energy transition.

The results reveal that value salience influences trust in the opposite direction to what was hypothesized. Rather than strengthening trust, value-based framing often reduced it, due to the emphasized value not being salient to respondents. In contrast, value-neutral communication maintained or even increased trust levels. This suggests that attempts to appeal to specific values in communication about energy transition policies can backfire when audiences do not share or prioritize those values, underscoring the importance of neutrality and clarity in government communication about the energy transition.

The next section outlines the methods and analysis strategy. We first describe the experimental design, survey structure, and measurement of institutional trust, followed by the empirical approach used to estimate average and heterogeneous treatment effects. We then present the results, beginning with the Difference-in-Differences (DiD) estimates and followed by the triple-difference (DDD) models that examine heterogeneity across trust profiles. Finally, we discuss the findings in relation to existing research on institutional trust, source credibility, and value-based communication, and reflect on their implications for the design of government communication in the heat transition. The paper concludes with a summary of key insights, limitations, and directions for future research.

1.1. Hypotheses

Based on theories of source credibility and value salience, we expect that highlighting a value that is intrinsically important to citizens would enhance message credibility. According to framing and motivational-salience research, value-congruent information should increase perceived sincerity and trustworthiness (Bolderdijk et al., 2013; Metzger and Flanagin, 2013). Accordingly, we formulate the following hypothesis:

H1. : Value-based framing increases institutional trust relative to a value-neutral message by making relevant concerns more salient.

Building on research showing that different governance levels are perceived as authoritative for different issue domains, we anticipate that source–value alignment matters (Bernardo and Palma-Oliveira, 2013; de Vries and Pruyn, 2007; Lefevere et al., 2017). Prior work demonstrates that trust varies by scale, with local governments being perceived as more credible on practical, household-level issues and national governments on strategic or environmental issues (Bernardo and Palma-Oliveira, 2013; de Vries and Pruyn, 2007; Lefevere et al., 2017). We therefore formulate the following hypothesis:

H2. : Messages from local governments generate higher institutional trust when they emphasize practical or financial considerations.

Conversely, environmental and stability-related considerations concern broader strategic issues that extend beyond the local level. Given the national scope and perceived ownership of these issues (Renn and Levine, 1991; Terpstra, 2011), we formulate the following hypothesis:

H3. : Messages from national government generate higher institutional trust when they emphasize environmental or stability-related considerations.

2. Methods

2.1. Experimental design

To address the research objectives, the study employed a 3 × 2 factorial design with an additional control group. Participants were randomly assigned to one of seven conditions. The two manipulated factors were:

1. Source of communication – *local government* vs *national government*;
2. Value-based framing – *environmental*, *financial*, or *stability*-related content.

The resulting six treatment combinations, together with the control group, are presented in Table 1. In the experimental texts, the local government condition was explicitly identified using the Dutch term “gemeente” (municipality), while the national government condition referred to the “Rijksoverheid” (national government). Thus, “local” in

Table 1
3 × 2 Factorial research design.

	Environmental framing	Financial framing	Stability framing
Local Government	T1: LocalEnviron	T3: LocalFinan	T5: LocalStab
National Government	T2: NatEnviron	T4: NatFinan	T6: NatStab

this study refers specifically to the municipal level of government rather than to respondents' subjective perceptions of what constitutes their local area.

2.2. Measures

Before exposure to the treatment, participants completed a short questionnaire measuring baseline institutional trust. Trust was conceptualized as a multidimensional construct comprising competence, integrity, and benevolence, following the operationalization by Grimmelikhuijsen and Knies (2017), itself based on McKnight et al. (2002). This operationalization should not be interpreted as treating trust as a purely instrumental or static construct. Institutional trust is conceptualized here as fundamentally relational, reflecting the relationship between citizens and governmental institutions, consistent with Otto et al. (2023). The quantitative operationalization serves a methodological purpose by allowing trust to be measured at a particular point in time and changes in this relationship to be examined. The relational scope of the study is deliberately restricted to citizens' trust in the government. Although trust may also operate in the opposite direction, from the government toward citizens, this dimension falls outside the empirical scope of the study. The experiment focuses specifically on whether participation and value-based government communication affect citizens' institutional trust.

The items were translated and adapted to the context of energy policy, using the generic term "de overheid" (the government) to maintain neutrality across governance levels. The items used are summarized in Table 2. The original English is presented in the left column, the Dutch translations are presented in the right column. The items are measured on a 5-point scale, and are then aggregated with a factor analysis. The factor analysis will be discussed in chapter 4.1. In addition, respondents provided demographic information, including age, education level, and political orientation.

Table 2
Questionnaire used by Grimmelikhuijsen and Knies (2017).

Item by Grimmelikhuisen and Knies (2017)	Translated and applied
ADDED PREFIX: When it concerns energy policy (…)	Als het gaat om energiebeleid, (…)
COMP1: [The municipality of XX] is capable.	(…) is de overheid bekwaam
COMP2: [The municipality of XX] is effective.	(…) is de overheid effectief
COMP3: [The municipality of XX] is skilful.	(…) is de overheid vaardig
COMP4: [The municipality of XX] is expert.	(…) is de overheid een expert
COMP5: [The municipality of XX] carries out its duty very well.	(…) voert de overheid zijn taak goed uit
BEN1: If citizens need help, [the municipality of XX] will do its best to help them.	(…) zal de overheid zijn best doen burgers te helpen, als zij hulp nodig hebben
BEN2: [The municipality of XX] acts in the interest of citizens.	(…) handelt de overheid in het belang van inwoners
BEN3: [The municipality of XX] is genuinely interested in the well-being of citizens.	(…) is de overheid werkelijk geïnteresseerd in het welzijn van inwoners
INT1: [The municipality of XX] approaches citizens in a sincere way.	(…) benadert de overheid inwoners op een oprechte manier
INT2: [The municipality of XX] is sincere.*	(…) is de overheid oprecht
INT3: [The municipality of XX] keeps its commitments.	(…) houdt de overheid zich aan zijn afspraken
INT4: [The municipality of XX] is honest.*	(…) is de overheid eerlijk

2.3. Participatory value evaluation

After completing the baseline questions, participants were introduced to the Participatory Value Evaluation (PVE), which served as the experimental context. The introductory text to the PVE constituted the treatment and was presented according to the participant's randomly assigned condition. Each text varied in both governance level and value emphasis, creating the six experimental combinations described above. The control group received a neutral introduction without reference to a governance level or specific value frame. The full treatment texts and their rationale are provided in Appendix A.

We use a Participatory Value Evaluation (PVE) to elicit citizens' preferences in policy decisions. The PVE is a choice-based method in which participants allocate a fixed budget of points across multiple policy objectives, thereby revealing the relative importance they assign to different societal goals. The method is widely applied in the Netherlands, particularly in the context of the energy transition, and has been shown to generate richer insights than conventional survey approaches (de Vries et al., 2025; Mouter et al., 2021).

In this study, participants distributed 40 points across eight policy aspects related to heat-grid affordability in response to the question: *"The government wants sustainable heat to be affordable. What aspects are important in your opinion?"* The policy aspects were developed in collaboration with Energie Beheer Nederland and administrators from sixteen Dutch municipalities, ensuring that the included dimensions reflect realistic and policy-relevant trade-offs.

Grounding the PVE in an existing participatory process strengthens the ecological validity of the study, as respondents engage with decision contexts that closely mirror actual policy dilemmas. While the PVE design is informed by practice, the analyses in this paper rely on primary data from a randomized controlled experiment conducted on a broadly representative national sample (N ≈ 3000).

After completing the PVE, participants again answered the same set of trust questions as in the baseline survey. This pre- and post-measurement design enabled estimation of within-subject changes in trust and the use of a difference-in-differences (DiD) framework to isolate treatment effects.

2.4. Data collection

Data were collected through a large-scale online survey administered by Dynata in the Netherlands (N ≈ 3000). Respondents were recruited from an online panel using quota-based sampling to approximate the Dutch population based on four characteristics: age, gender, region, and education, following benchmarks from Statistics Netherlands (CBS). Participants were randomly assigned to experimental conditions. The survey consisted of five sections: (1) a baseline measurement of institutional trust, (2) socio-demographic questions, (3) the experimental treatment embedded in a Participatory Value Evaluation (PVE), (4) additional background questions on housing and energy use, and (5) a post-treatment trust measurement. The average completion time was approximately 18 minutes. Table 3 provides descriptive statistics for the sample across the seven experimental conditions, including age, gender, educational attainment, and financial situation. The full balance tables can be found in Appendix F.

3. Analysis strategy

The inclusion of both pre- and post-treatment measures enabled estimation of treatment effects using a difference-in-differences (DiD) framework. This approach captures the change in trust before and after exposure to the experimental framing, relative to the control group,

Table 3
Descriptive statistics by experimental condition.

Characteristic	Control	Local Env.	National Env.	Local Financial	National Financial	Local Stability	National Stability
N	417	438	415	439	459	432	396
Age, mean (SD)	47.88 (17.22)	49.67 (17.79)	50.74 (17.28)	48.28 (16.92)	49.59 (17.82)	48.79 (17.31)	49.13 (17.22)
Gender							
Man	50.1%	52.3%	50.4%	51.7%	49.2%	47.7%	44.9%
Woman	49.6%	47.3%	49.2%	48.3%	50.5%	52.1%	55.1%
Other / non-binary / prefer not to say	0.2%	0.4%	0.4%	0.2%	0.2%	0.2%	0.0%
Education							
Low	20.1%	19.9%	20.1%	22.3%	20.0%	20.4%	20.0%
Middle	35.3%	39.3%	38.3%	37.2%	41.2%	41.0%	42.7%
High	44.4%	40.9%	41.5%	40.5%	38.5%	38.4%	36.9%
Financial situation ("getting by")							
Easy / very easy	49.9%	44.5%	47.5%	46.9%	44.9%	44.9%	48.2%
Neither easy nor difficult	34.1%	40.6%	36.9%	39.4%	38.3%	39.8%	34.3%
Difficult / very difficult	15.1%	14.1%	14.9%	13.2%	15.9%	14.8%	16.5%
Don't know / prefer not to say	1.0%	0.7%	0.7%	0.5%	0.9%	0.5%	1.0%

thereby accounting for individual-level unobserved heterogeneity. This allows us to test H1, expecting value-based framing to be more effective in increasing trust than value-neutral communication. Additionally, we can start to address H2 and H3, as the DiD presents clear treatment effects for each separate treatment. The basic specification is:

$$Y_{it} = \alpha + \beta_1 T_i + \gamma t_t + \delta(T_i{*}t_t) + \beta X_{it} + \varepsilon_{it}$$

where Y_{it} denotes trust for individual i at time t, T_i indicates treatment assignment, t_t is a time dummy (0 = pre, 1 = post), and the interaction term $(T_i{*}t_t)$ represents the DiD estimator (δ), which captures the causal effect of treatment on trust. βX_{it} includes additional covariates such as age, education, and occupation.

To account for heterogeneity in treatment effects, we explicitly recognize that the effectiveness of value-based messages depends on their alignment with individuals' prior predispositions toward governmental actors. Hypotheses H2 and H3 imply that message effectiveness is contingent on the perceived credibility and issue ownership of the source, which in turn varies across individuals. In other words, different respondents are likely to attach different salience to financial, environmental, or stability-related considerations depending on their baseline trust in local versus national government.

To capture this variation, we conduct a Latent Profile Analysis (LPA) on participants' baseline trust scores, identifying distinct trust profiles that reflect differences in perceived governmental competence, integrity, and benevolence. These profiles are then used to estimate Conditional Average Treatment Effects (CATEs), allowing us to assess whether the effectiveness of source–value combinations varies systematically across groups with different predispositions.

The DiD model was extended to incorporate these trust clusters as both main effects and interaction terms with the treatment effect, as follows:

$$Y_{it} = \alpha + \beta_1 T_i + \gamma t_t + \delta(T_i{*}t_t) + \beta_2 C_i + \beta_3(C_{it}{*}(T_i{*}t_t)) + \beta X_{it} + \varepsilon_{it}$$

where C_i denotes cluster membership derived from the LPA, and the term $(C_i{*}(T_i{*}t_i))$ captures variation in treatment effects across clusters. This extended model allows for the estimation of both average treatment effects (ATEs) and conditional effects (CATEs), revealing whether specific subgroups respond differently to particular value frames and governance levels.

The standard Difference-in-Differences (DiD) specification relies on the assumption that, absent treatment, treated and control units would have followed parallel trends over time (Angrist and Pischke, 2009; Wooldridge, 2014). When treatment effects are allowed to vary across subgroups, this identifying assumption becomes more restrictive, as it additionally requires that all subgroups would have experienced identical pre–post changes in the absence of treatment.

To relax this assumption for the models including the LPA clusters, we subsequently adopt a canonical triple-difference (DDD) specification by augmenting the DiD model with subgroup–time $\beta_4(C_i{*}t_t)$ and subgroup–treatment $\beta_3(C_i{*}T_i)$ interaction terms (Olden and Møen, 2022). This fully interacted model allows each subgroup to exhibit its own baseline level, untreated time trend, and treatment–control gap, such that identification relies only on parallel trends within subgroups. The resulting triple-interaction term therefore isolates differential treatment effects across trust classes, net of subgroup-specific time shocks and baseline differences. The model specification of our DDD estimate is as follows:

$$Y_{it} = \alpha + \beta_1 T_i + \gamma t_t + \delta(T_i{*}t_t) + \beta_2 C_i + \beta_3(C_i{*}T_i) + \beta_4(C_i{*}t_t) + \beta_5(C_i{*}(T_i{*}t_t)) + \beta X_{it} + \varepsilon_{it}$$

To complement the regression-based analyses, we implement an Honest Causal Forest (HCF) as an exploratory tool to better understand patterns in the data. Rather than focusing on treatment effect estimation, we pool all treatment conditions and use the model to examine how socio-demographic characteristics relate to variation in outcomes. This approach is particularly useful given that standard regression models may obscure relationships when effects are non-linear or involve higher-order interactions. The HCF framework (Athey and Imbens, 2016) recursively partitions the data, allowing for flexible detection of such non-linearities without imposing a parametric structure. In this application, it provides a more nuanced view of how respondent characteristics relate to observed outcomes, helping to interpret patterns that are

Table 4
Overview of analytical approaches.

Analytical approach	Purpose	Main contribution to the analysis
Difference-in-Differences (DiD)	Estimate average changes in trust associated with each experimental treatment relative to the control condition	Tests whether the value–source treatments affect institutional trust on average
Latent Profile Analysis (LPA)	Identify groups of respondents with distinct baseline trust profiles	Provides theoretically meaningful trust profiles for examining heterogeneity in treatment responses
Triple-Difference (DDD)	Estimate whether treatment effects differ across baseline trust profiles	Extends the DiD analysis by accounting for subgroup-specific baseline differences and time trends
Honest Causal Forest (HCF)	Explore treatment-effect heterogeneity without imposing a predefined functional form	Tests whether additional systematic heterogeneity emerges from respondent characteristics beyond the predefined trust profiles

less visible in the regression results. To provide an overview of the analytical strategy, Table 4 summarizes the purpose of each model and its specific contribution to the analysis.

4. Results

4.1. Factor analysis and trust-index

We validated the dimensionality of the twelve pre-treatment trust items using exploratory factor analysis and confirmed a three-factor structure; all statistical details are reported in Appendix B. Because the subsequent analyses require a single continuous outcome, we constructed an aggregate trust index by weighting each item by its factor loading:

$$\text{pre trust}_i = \frac{\Sigma_j w_j x_{ij}}{\Sigma_j w_j . 1(x_{ij}\text{observed})}$$

where $w_j$is the loading of item $j$on the general trust factor and $x_{ij}$is respondent i's Likert response. Scores were standardized and used as the main trust outcome. Full item loadings and index-construction details appear in Appendix C.

Our factor analysis also provides strong empirical confirmation of the multidimensional trust framework established in prior research. Consistent with Grimmelikhuijsen and Knies (2017), the twelve trust items loaded cleanly onto the three theoretically expected dimensions, competence, benevolence, and integrity. Each item clustered exactly within the category proposed in the literature, indicating that respondents interpreted the trust questions in a manner fully aligned with the conceptual model. This congruence between theory and measurement strengthens the internal validity of the trust index and supports the use of a composite measure in the subsequent analyses.

4.2. Latent profile analysis

We applied a latent profile analysis (LPA) to the standardized baseline trust index. Comparing models with one to six components showed that the six-class equal-variance solution provided the best overall fit. We therefore retained six latent trust profiles, ordered from lowest to highest mean trust.

Descriptive information on class sizes, mean standardized trust, and posterior classification quality is provided in Appendix C. To give an overview of how trust levels differ across profiles, Table 5 presents the pre- and post-treatment trust scores for each class.

Trust levels across the six latent profiles differ substantially at baseline, increasing as expected from the lowest- to highest-trust groups. The lower-trust classes (1 and 2) show the largest gains in trust between the pre- and post-measurement, whereas the highest-trust classes (5 and 6) exhibit small declines, with class 4 showing only a minimal increase (Fig. 2). This pattern suggests that trust orientations do not shift uniformly after exposure to the treatments or the PVE. These profiles enable us to examine whether treatment effects vary systematically across distinct trust dispositions.

Table 5
Pre- and Post-Treatment Trust by Latent Class.

Trust class	n (pre)	Mean pre	95% CI pre	n (post)	Mean post	95% CI post
1	428	1.16	[1.14; 1.17]	428	1.26	[1.22; 1.30]
2	1019	2.04	[2.03; 2.06]	1014	2.12	[2.09; 2.15]
3	86	2.44	[2.44; 2.44]	86	2.55	[2.45; 2.66]
4	895	2.90	[2.88; 2.91]	887	2.91	[2.88; 2.95]
5	465	3.73	[3.71; 3.76]	463	3.61	[3.57; 3.66]
6	103	4.62	[4.58; 4.67]	103	4.34	[4.21; 4.46]

4.3. Framing effects on institutional trust

We estimate both Difference-in-Differences (DiD) and Triple-Difference (DDD) models to examine framing effects on trust. The DiD models capture average treatment effects, while the DDD specifications extend this framework to account for heterogeneity across trust classes. For both approaches, we estimate models with and without socio-demographic controls. DiD results are used to assess average treatment effects across conditions, whereas the DDD models capture heterogeneous effects across trust profiles, relative to a reference trust cluster (C4).

Given the large number of coefficients, the full set of models is reported in the appendix. In the main text, we present the results in two parts to guide interpretation: Table 6 with the DiD estimates and Table 7 with the triple interaction terms that capture heterogeneous effects across trust classes. The additional interaction terms can be found in Appendix D.1.

4.3.1. Average framing effects: difference-in-differences

Table 6 presents the multi-arm Difference-in-Differences (DiD) estimates across four model specifications. Models 1 and 3 report the baseline DiD results, while Models 2 and 4 embed the same DiD structure within the triple-difference framework.

Across specifications, the coefficient on *Post* is small and positive (Model 1: 0.054, $p < .05$; Model 3: 0.049, $p < .10$; Models 2 and 4: 0.048, n.s.), indicating a modest overall increase in trust between pre- and post-measurement, averaged across treatment conditions.

Turning to the treatment-specific DiD estimators (Post × Treatment), average framing effects are limited in magnitude. The most consistent

Table 6
Diff-in-Diff estimates from the full multi-arm triple-diff specification.

Variable	DiD	DDD	DiD + Demo	DDD + Demo
Post	0.054**	0,048	0.049*	0,048
	(0.025)	(0.048)	(0.025)	(0.048)
LocalEnviron	−0,019	0,000	−0,013	0,001
	(0.014)	(0.029)	(0.015)	(0.030)
NatEnviron	−0,007	0,024	−0,002	0,024
	(0.015)	(0.029)	(0.016)	(0.030)
LocalFinan	−0,018	0,006	−0,019	−0.000
	(0.015)	(0.030)	(0.015)	(0.030)
NatFinan	0,008	0,04	0,019	0,044
	(0.014)	(0.029)	(0.015)	(0.030)
LocalStab	0,006	0.052*	0,014	0.054*
	(0.015)	(0.030)	(0.015)	(0.031)
NatStab	−0.010	0,02	−0,001	0,026
	(0.015)	(0.029)	(0.016)	(0.030)
Post × LocalEnviron	−0,054	−0,072	−0,048	−0,072
	(0.034)	(0.065)	(0.035)	(0.066)
Post × NatEnviron	−0,001	0,052	0,004	0,054
	(0.034)	(0.062)	(0.035)	(0.062)
Post × LocalFinan	−0.066**	−0.113*	−0.064*	−0.113*
	(0.033)	(0.061)	(0.033)	(0.061)
Post × NatFinan	−0,022	0,032	−0,017	0,036
	(0.034)	(0.063)	(0.034)	(0.064)
Post × LocalStab	−0.059*	−0,065	−0.055*	−0,065
	(0.033)	(0.062)	(0.033)	(0.063)
Post × NatStab	−0.020	−0,045	−0,015	−0,045
	(0.033)	(0.061)	(0.033)	(0.061)
Controls (Demo)	NO	NO	YES	YES
N	5962	5962	5936	5936
R^2	0.824	0.828	0.829	0.833

Note: All models estimated using ordinary least squares with robust standard errors in parentheses. Models 2 and 4 embed the DiD structure within a fully interacted triple-difference specification. Trust cluster C4 serves as the reference category in the DDD models.

Table 7
Heterogeneous framing effects across trust clusters: triple-difference estimators.

Variable	DDD	DDD + Demo
Post × LocalEnviron × Trust C1	0,136	0,148
	(0.108)	(0.109)
Post × NatEnviron × Trust C1	−0,058	−0,048
	(0.094)	(0.094)
Post × LocalFinan × Trust C1	0,061	0,074
	(0.090)	(0.090)
Post × NatFinan × Trust C1	−0.080	−0,072
	(0.097)	(0.098)
Post × LocalStab × Trust C1	0,089	0,099
	(0.098)	(0.099)
Post × NatStab × Trust C1	0,023	0,035
	(0.094)	(0.094)
Post × LocalEnviron × Trust C2	−0,038	−0,026
	(0.087)	(0.088)
Post × NatEnviron × Trust C2	−0,067	−0,063
	(0.087)	(0.088)
Post × LocalFinan × Trust C2	0,083	0,086
	(0.084)	(0.084)
Post × NatFinan × Trust C2	−0,054	−0,051
	(0.086)	(0.087)
Post × LocalStab × Trust C2	0,018	0,025
	(0.085)	(0.085)
Post × NatStab × Trust C2	0,031	0,037
	(0.088)	(0.088)
Post × LocalEnviron × Trust C3	−0,147	−0,147
	(0.209)	(0.210)
Post × NatEnviron × Trust C3	−0.348**	−0.350**
	(0.171)	(0.172)
Post × LocalFinan × Trust C3	0,043	0,043
	(0.212)	(0.212)
Post × NatFinan × Trust C3	−0.320*	−0.324*
	(0.187)	(0.188)
Post × LocalStab × Trust C3	−0.410***	−0.410***
	(0.143)	(0.143)
Post × NatStab × Trust C3	−0,001	−0,001
	(0.151)	(0.152)
Post × LocalEnviron × Trust C5	0,028	0,028
	(0.107)	(0.108)
Post × NatEnviron × Trust C5	−0,037	−0,039
	(0.095)	(0.095)
Post × LocalFinan × Trust C5	−0,087	−0,093
	(0.109)	(0.110)
Post × NatFinan × Trust C5	−0,063	−0,067
	(0.102)	(0.103)
Post × LocalStab × Trust C5	−0,032	−0,032
	(0.105)	(0.105)
Post × NatStab × Trust C5	0,067	0,066
	(0.093)	(0.093)
Post × LocalEnviron × Trust C6	0,239	0,239
	(0.177)	(0.178)
Post × NatEnviron × Trust C6	−0,138	−0.140
	(0.306)	(0.307)
Post × LocalFinan × Trust C6	0.357**	0.357**
	(0.175)	(0.176)
Post × NatFinan × Trust C6	−0,075	−0.080
	(0.208)	(0.209)
Post × LocalStab × Trust C6	0,12	0,12
	(0.172)	(0.173)
Post × NatStab × Trust C6	0,081	0,081
	(0.223)	(0.224)
N	5962	5936
R^2	0,828	0,833

Note: This table isolates the triple interaction terms (Post × Treatment × Trust) from the fully interacted triple-difference specification. These coefficients represent differential treatment effects relative to the reference trust cluster (C4). The corresponding baseline DiD effects for C4 are reported in Table 4. Standard errors in parentheses.

pattern concerns the local financial frame, which yields a negative and statistically significant average effect in the baseline DiD models (Model 1: −0.066, $p < .05$; Model 3: −0.064, $p < .10$). This negative effect persists when embedded in the DDD framework (Models 2 and 4: −0.113, $p < .10$), indicating that, for the reference trust cluster (C4), emphasizing local financial implications slightly reduces trust relative to the control condition.

The local stability frame produces a small negative effect in the simpler DiD models (Model 1: −0.059, $p < .10$; Model 3: −0.055, $p < .10$), although this coefficient becomes statistically insignificant once cluster-specific dynamics are modeled explicitly (Models 2 and 4: −0.065, n.s.). The remaining treatment arms—local environmental, national environmental, national financial, and national stability framing—do not generate statistically significant average effects. Their coefficients remain small and close to zero across specifications.

Taken together, the DiD models suggest that aggregate framing effects on trust are modest. The only consistent average pattern is a negative response to local financial framing. At the same time, trust generally increases between the pre- and post-measurement, with the largest increase observed in the control condition without value-based framing (Fig. 1). However, these aggregate patterns may mask heterogeneity across baseline trust profiles. We therefore turn to the triple-difference specification.

4.3.2. *Heterogeneous effects: triple differences*

Models 2 and 4 estimate a canonical triple-difference (DDD) specification in which treatment effects are allowed to vary across trust clusters. Trust cluster C4 serves as the reference category, implying that the Post × Treatment coefficients reported in these models represent the DiD treatment effect for C4.

Appendix D.1 reports the additional interaction terms that extend the DiD into a DDD model, including cluster main effects and cluster-specific time trends. Several clusters exhibit statistically significant time trends independent of treatment assignment. Relative to C4, Trust C3 shows a significantly larger pre–post increase in trust (Post × Trust C3 = 0.250, $p < .05$), whereas Trust C5 (−0.125, $p < .10$) and Trust C6 (−0.373, $p < .01$) display relative declines over time. These patterns indicate that trust evolves differently across baseline trust segments, underscoring the importance of relaxing the parallel-trends assumption across subgroups.

Table 7 then isolates the triple interaction terms, which constitute the triple-difference estimators.

The triple-interaction terms reveal heterogeneous framing effects. The most consistent pattern concerns Trust C3, which responds significantly less positively, or more negatively, than C4 to several frames. Specifically, the interaction terms for Post × NatEnviron × Trust C3 (−0.348 / −0.350, $p < .05$), Post × NatFinan × Trust C3 (−0.320 / −0.324, $p < .10$), and Post × LocalStab × Trust C3 (−0.410, $p < .01$) are negative and statistically significant in both DDD models. Substantively, these coefficients indicate that, relative to the reference cluster, C3 exhibits a downward shift in the treatment effect under these frames.

In contrast, Trust C6 shows a significantly more positive response than C4 to the local financial frame (Post × LocalFinan × Trust C6 = 0.357, $p < .05$ in both specifications). Although the average DiD effect of the local financial frame is negative for C4, this positive differential effect implies that C6 reacts in the opposite direction relative to the reference group. The aggregate negative effect therefore reflects partially offsetting responses across trust segments.

For the remaining cluster-by-treatment combinations, the triple-interaction terms are not statistically distinguishable from zero. The pattern of significant interactions is virtually identical between Models 2 and 4, indicating that the inclusion of demographic controls does not materially alter the results.

Overall, the DDD results show that differential framing effects are concentrated in particular trust clusters, most notably the negative differential responses of Trust C3 and the positive response of Trust C6 under local financial framing. The interaction between the post period and trust classes further reveals differences in pre–post trust trajectories across the trust profiles. Trust increases for classes C1, C2, and C3, whereas trust decreases for the higher-trust groups C5 and C6. These changes are statistically significant only for a subset of the trust classes.

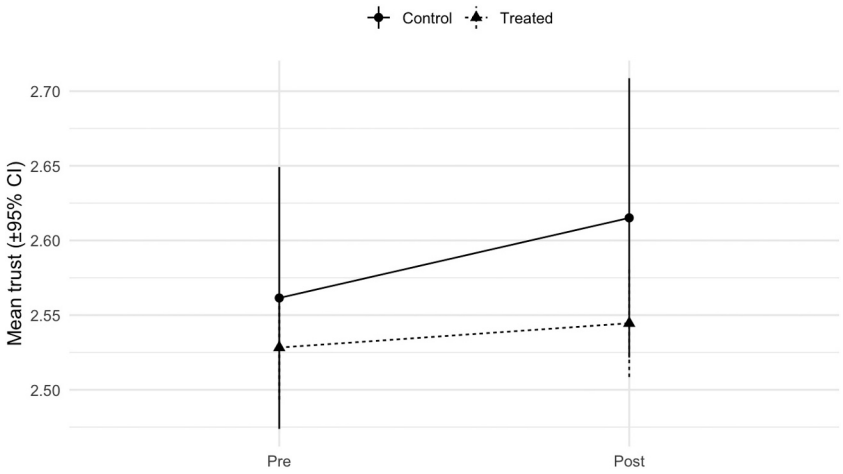


Fig. 1. Mean trust by time and treatment.

4.4. Honest causal forest

To complement the parametric Difference-in-Differences models, an Honest Causal Forest (HCF) was estimated to examine whether the intervention generated treatment heterogeneity not captured by the linear specifications. Separate forests were trained for post-treatment trust and for within-subject trust change. Both models included the complete set of demographic variables, affordability indicators, and baseline trust.

Across both outcomes, the forests reproduce the central finding from the DiD models: the intervention does not generate a meaningful average treatment effect. The mean Conditional Average Treatment Effects (CATEs) center around zero, indicating that the null effect is not due to linearity assumptions.

Although the average treatment effect estimated by the causal forests is effectively zero, the models do uncover a few small and inconsistent pockets of heterogeneity, which illustrate the limited extent to which certain subgroups might respond differently to the intervention. Appendix E provides examples of these irregular patterns. We find that household composition produces largely flat CATE profiles, yet respondents living alone with children display a modest negative deviation, though this effect is not mirrored in adjacent categories and does not reappear in the alternative forest outcome. Similarly there exists an isolated pattern at the regional level: residents of Noord-Holland exhibit a slightly positive effect for both trust_post and trust change, while all other provinces center around zero, and the DiD models show no corresponding geographic structure. Housing quality provides a third example. We additionally find evidence that respondents living in the best-insulated dwellings experience a small negative shift in trust, whereas all other insulation categories align closely with zero. In each of these cases, the deviations are small, statistically fragile, and not theoretically coherent, and they fail to reoccur across related subgroups or outcomes. Taken together, these examples demonstrate that although the forests technically detect some heterogeneity, it is sparse, non-systematic, and unlikely to reflect meaningful or policy-relevant subgroup differences.

To explore the structure of heterogeneity more systematically, we computed permutation-based variable importance scores from the HCF. These scores are presented in Fig. 2.

Across both models, age and baseline trust dominate the importance rankings. Their influence is substantially larger than that of any demographic or affordability variable. This suggests that the most meaningful heterogeneity operates along these two axes, rather than along standard demographic categories.

Both moderators exhibit distinctly non-linear patterns, as visualized in Fig. 3.

Fig. 3 plots the CATE–age relationship, showing that younger respondents experience slightly more positive effects, middle-aged respondents more negative effects, and older respondents a partial return toward zero. The pattern is smooth rather than clustered, indicating that age acts as a continuous moderator rather than creating categorical subgroups. Importantly, the absence of sharp discontinuities suggests no

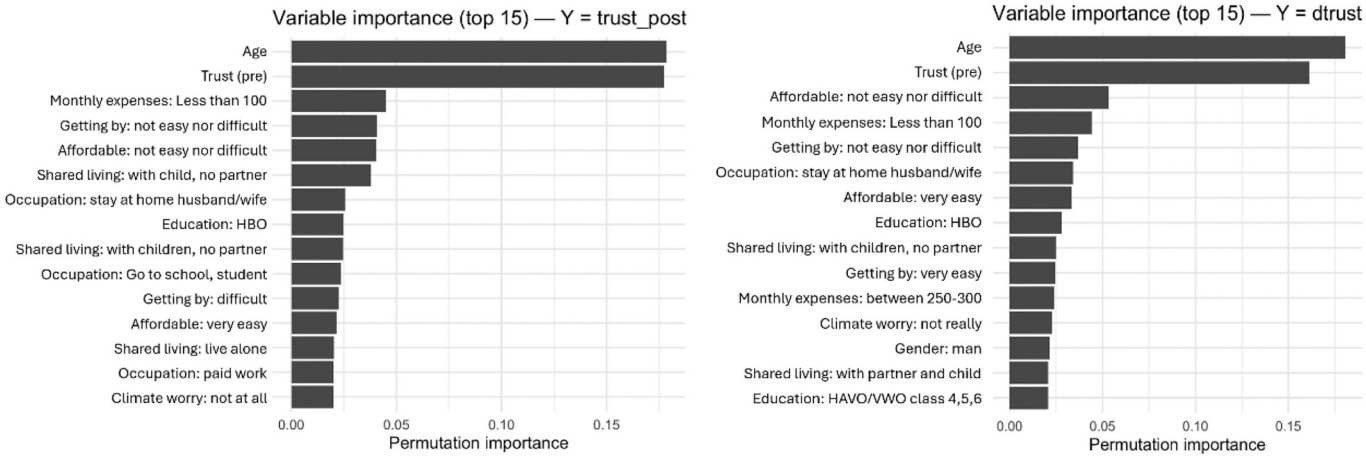


Fig. 2. Variable importance for delta trust and post trust.

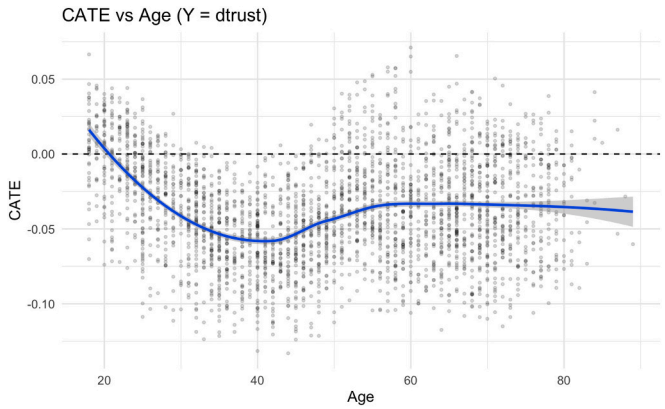


Fig. 3. CATE vs age on change in trust.

clear age thresholds at which treatment effects change, but rather a gradual evolution over the life course. This supports modelling age flexibly and cautions against interpreting heterogeneity in terms of discrete age-based segments.

Fig. 4 shows CATE heterogeneity by baseline trust, revealing a similar structure: individuals with very high initial trust experience more negative treatment effects, while those with low or moderate trust cluster around zero. This confirms that the most important axes of heterogeneity are continuous and non-linear.

Taken together, the HCF results reinforce the conclusions from the DiD analysis. The intervention does not shift trust on average. The heterogeneity in treatment effects that does exist is sparse, weak, and highly non-systematic. The most meaningful variation is non-linear, continuous, and does not produce distinct policy-relevant subgroups. Even under a flexible non-parametric modelling framework, the value-based framing interventions fail to generate consistent or predictable changes in trust.

5. Discussion

The results offer a cautious but consistent set of insights into how value-based government communication shapes public trust in the context of sustainable heat policy. Across specifications, value-based framing does not meaningfully increase trust on average and, in specific cases, decreases it. However, these average effects conceal substantial heterogeneity across trust clusters. In particular, the local financial frame reduces trust for the reference trust group (C4), while the fully interacted DDD models reveal pronounced differential responses across other trust profiles. The most consistent heterogeneous pattern concerns Trust C3, which reacts significantly more negatively than C4 to several frames, including national environmental, national financial,

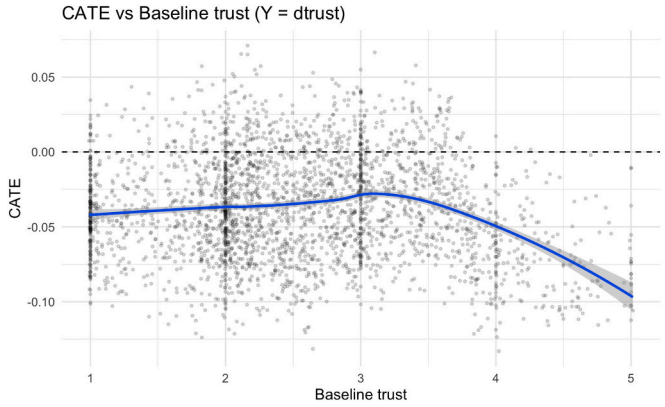


Fig. 4. CATE vs baseline trust on change in trust.

and local stability messaging. By contrast, Trust C6 shows a more positive response to the local financial frame, partially offsetting the negative average effect observed for C4. These findings underscore that framing effects are not uniform but conditional on pre-existing trust orientations.

Several mechanisms from the institutional trust and political communication literature help explain why local financial framing and, to a lesser extent, stability framing produce differentiated responses across trust clusters. Financial messaging directly engages distributive fairness and institutional integrity, two core dimensions of trust (Mayer et al., 1995; Grimmelikhuijsen and Knies, 2017). When local governments emphasize costs or financial implications, the message does more than provide information; it signals how responsibilities and burdens are allocated. For the reference trust cluster (C4), local financial framing is associated with a modest decline in trust, suggesting sensitivity to how financial accountability is communicated at the local level. At the same time, Trust C6 responds more positively than C4 to the same frame, indicating that financial transparency may signal competence and responsibility for some segments while activating concerns about cost shifting or administrative capacity for others. Rather than uniformly increasing trust, financial framing appears to redistribute trust evaluations across existing orientations.

Stability framing exhibits a related but distinct pattern. Messages emphasizing system continuity and reliability touch upon expectations of competence and risk management, domains that are closely tied to institutional credibility. Prior research shows that trust is particularly sensitive when communication activates expectations about institutional performance and mandate boundaries (Earle and Siegrist, 2008; Tyler, 2006). When such cues are perceived as congruent and appropriate, they may reassure; when perceived as strategic or misaligned, they can heighten skepticism. The heterogeneous responses observed in the richer specifications suggest that stability cues, like financial ones, operate primarily as relational signals. They interact with pre-existing trust orientations rather than shifting attitudes uniformly.

Taken together, these findings align with expectancy and motivated reasoning accounts of institutional trust (Druckman and McGrath, 2019; Grimmelikhuijsen and Meijer, 2022). Citizens evaluate government messages through the lens of prior experiences and relational expectations, rather than as neutral informational inputs. Value-based framing therefore functions less as persuasion and more as a diagnostic cue about fairness, competence, and credibility. Where such cues resonate with existing expectations, trust can be maintained or modestly reinforced; where they appear incongruent, trust declines. In this sense, both financial and stability frames intensify pre-existing trust gradients rather than shifting attitudes uniformly.

From the broader trust literature, this pattern reinforces the view that perceptions of competence, integrity, and benevolence (Grimmelikhuijsen and Knies, 2017; Siegrist et al., 2011) are not easily altered by short textual interventions. Value-heavy messages may heighten the salience of contested trade-offs or unintentionally signal strategic intent, thereby provoking skepticism among those predisposed to distrust institutions. Moreover, citizens with higher baseline trust tend to hold stronger expectations regarding procedural fairness and institutional neutrality; when communication violates these norms, they may update trust downward more sharply (Otto et al., 2023; Grimmelikhuijsen and Meijer, 2022). Importantly, this interpretation does not imply that government itself is a neutral actor or that its communication is without policy objectives. Rather, it concerns the normative content of the message: whereas the treatment conditions explicitly foreground particular values as reasons for supporting the policy, the control condition does not privilege a particular value. In this sense, the control condition is more appropriately characterized as non-normative rather than value-free or neutral. By contrast, non-normative information appears safest for maintaining or modestly increasing trust, potentially because it allows citizens to interpret policy proposals through their own value frameworks without perceiving manipulation.

This interpretation is consistent with findings that transparency and straightforward communication can foster trust when perceived as sincere and non-strategic (Grimmelikhuijsen, 2012).

This interpretation also invites reflection on our original theoretical expectation. We hypothesized that emphasizing a salient value cue in government communication would enhance trust by aligning messages with citizens' primary concerns. However, the results indicate that the probability of selecting a cue that is not perceived as salient for a given trust segment is substantial. Rather than reliably generating reassurance, value-based framing appears to carry a meaningful risk of activating perceptions of incongruence or strategic intent when the emphasized cue does not align with recipients' relational expectations. In practice, the likelihood of deploying a misaligned cue may exceed the likelihood of selecting one that meaningfully strengthens trust. This asymmetry helps explain why average treatment effects are muted and why heterogeneous responses dominate the empirical pattern.

The contrast with the control group is striking. Participants exposed to a non-normative introduction exhibited the most favorable trust outcomes, with trust increasing modestly after completing the PVE. This pre–post increase should be interpreted cautiously, as we cannot disentangle the effect of participating in the PVE from a repeated-measurement effect. Nevertheless, the increase observed in the control condition is consistent with the possibility that participation itself contributes positively to trust. Importantly, this increase is generally diminished among respondents exposed to the value-based treatments, suggesting that a non-normative introduction may be more conducive to maintaining or increasing trust than explicitly emphasizing particular values.

These findings qualify common assumptions in the nudging literature that informational interventions are low-risk or benign (Thaler and Sunstein, 2008; Gillingham and Tsvetanov, 2018; Bettinger et al., 2020). In the context of institutional trust, value-laden informational cues can do more than simply fail to persuade; they can actively undermine trust by signaling strategic intent or misaligned priorities. This aligns with broader evidence that value misalignment and perceived persuasion attempts can trigger reactance or boomerang effects in public communication (Bolderdijk et al., 2012, 2013; Hart and Nisbet, 2012; Rains, 2013). Evidence from PVEs conducted during the COVID-19 crisis suggests similar dynamics. In a large-scale Dutch PVE on relaxing lockdown measures, a non-trivial minority of participants reported that participation did not make them better able to accept the government's eventual decisions, indicating that informational and participatory interventions do not uniformly enhance trust or legitimacy (Mouter et al., 2021).

A central contribution of this study lies in showing that these effects are strongly conditioned by pre-existing trust orientations. The six latent profiles reveal pronounced differences in baseline trust, and both treatment responses and pre–post trust trajectories vary across these profiles. Individuals with high baseline trust tend to react more negatively to value-loaded messages, whereas those with low trust occasionally show small positive responses, particularly to stability framing. Differences in pre–post trajectories further indicate that trust does not develop uniformly across these profiles over the course of the participatory exercise. Although these changes cannot be causally attributed to the PVE itself, they are consistent with respondents engaging with the exercise differently depending on their pre-existing relationship with government.

This aligns with the Trust–Confidence–Cooperation perspective (Earle and Siegrist, 2008; Siegrist et al., 2011), which emphasizes that acceptance of information depends on prior confidence in the institution and perceived procedural fairness. It also resonates with Otto et al.'s (2023) argument for a relational concept of trust in energy system transformations: communication is interpreted against the backdrop of existing relationships and experiences, not as a neutral input. Our results extend this literature by demonstrating empirically that these trust orientations operate as continuous gradients rather than discrete audience segments, and that they outweigh socio-demographic differences in explaining communication sensitivity.

The Honest Causal Forest analysis sharpens this picture by confirming that the absence of an average treatment effect is not an artefact of linear modelling, but a genuine consequence of heterogeneous and largely offsetting responses. Age and baseline trust emerge as the dominant moderators, operating in non-linear ways, while standard demographic variables such as education, income, and household composition contribute little to explaining treatment-effect variation. This finding challenges common assumptions in both communication and policy practice that segmenting audiences along socio-demographic lines will improve the effectiveness of value-based messaging (Metzger and Flanagin, 2013; Bernardo and Palma-Oliveira, 2013). It also qualifies source-salience arguments from work on local versus national communicators (de Vries and Pruyn, 2007; Lefevere et al., 2017): in our case, the level of government and the value emphasized matter less for trust than who the audience is in relational terms. The small, irregular pockets of heterogeneity detected by the forest (e.g. specific provinces or insulation categories) do not align with the theoretically expected patterns from place-attachment or issue-ownership literatures and are therefore unlikely to reflect robust mechanisms.

Finally, our results speak directly to the emerging literature on the trust–participation nexus in the energy transition. Otto et al. (2023) argue that trust and participation are mutually constitutive, and that relational forms of engagement are key for sustaining legitimacy in net-zero transformations. The modest but consistent increase in trust we observe across all conditions after the PVE supports this view: even when communication framing fails or backfires, participation itself appears to enhance institutional trust. Being asked for input, reflecting on trade-offs, and engaging with a structured decision tool resonates with the dimensions of procedural fairness and benevolence highlighted in the trust literature. This suggests that, in the context of sustainable heat policy, the most promising route for strengthening trust is not to refine value-based messages but to invest in participatory processes that make government responsiveness tangible. In that sense, our study bridges the literatures on nudging, trust, and participation by showing that informational framing alone is a fragile instrument for building trust, whereas well-designed participatory tools can contribute to the relational foundations upon which trust in energy-transition governance ultimately rests.

6. Policy recommendations

The results of this study point to a clear message for policymakers: when the objective is to maintain or strengthen institutional trust during the heat transition, communication strategies should prioritize neutrality, clarity, and procedural transparency rather than value-based persuasion. The experiment demonstrates that attempts to guide public interpretation by emphasizing environmental, financial, or stability-related values do not generate higher trust. In several cases such approaches even reduced trust. For policy practice, this means that communicators should resist the temptation to "sell" heat-transition policies through targeted value narratives. Instead, providing concise, factual information about what is being decided, why certain options are being considered, and how these decisions will affect residents is likely to be far more effective in sustaining trust. In this context, non-normative communication does not imply that government communication can be entirely neutral or without policy objectives. Rather, it refers to communication that does not explicitly privilege a particular value or motivation. For example, an invitation to participate may simply communicate that a heat grid is being considered and invite citizens to provide their views, rather than presenting environmental, financial, or stability-related considerations as the reason for supporting the policy. Our findings suggest that the former approach is more conducive to maintaining trust.

In designing information campaigns, policymakers should be aware

that the public does not respond to communication in homogeneous or demographically predictable ways. The heterogeneity uncovered in this study is driven almost entirely by citizens' underlying trust orientations rather than by age, education, income, or other common segmentation criteria. This finding challenges the conventional assumption that different demographic groups require different types of framing. Rather than crafting specialized narratives for specific population segments, governments should focus on consistency and transparency across their communication. Efforts to tailor messages to presumed demographic sensitivities are unlikely to generate meaningful gains in trust and may even risk undermining credibility when those assumptions misalign with citizens' concerns.

A central implication of the study is that trust is more effectively strengthened through participation than through communication alone. The modest but consistent increase in trust observed among respondents suggests that structured participatory processes like the PVE can play a crucial role in cultivating institutional legitimacy. Policymakers should therefore treat participation not as an add-on but as a core component of trust-sensitive policy design. This entails integrating participatory tools early in policy development, providing citizens with opportunities to engage with trade-offs, and ensuring that the outcomes of participatory processes are visibly connected to policy decisions. When citizens see that their input is taken seriously, trust grows. Not because they are persuaded by a message, but because they experience government as attentive, transparent, and fair.

7. Conclusion

The findings of this study provide a clear and consistent picture of how citizens respond to government communication about the heat transition in the Netherlands. Across specifications, value-based framing does not enhance institutional trust, contrary to our initial hypothesis that emphasizing salient values would increase trust (H1). In particular, emphasizing local financial implications reduces trust for the reference trust cluster, while other frames yield largely muted average effects. Rather than uniformly increasing trust, value-based messages generate heterogeneous responses across trust profiles. These results suggest that the likelihood of selecting a cue that meaningfully enhances trust is limited, while the risk of activating perceptions of incongruence or strategic intent is substantial. In this context, backfiring refers specifically to communication producing an effect opposite to that intended: rather than strengthening institutional trust, the value-based framing results in a reduction in trust. In contrast, the value-neutral introduction used for the control group is associated with modest increases in trust, indicating that straightforward communication is less likely to provoke skepticism and more likely to sustain credibility.

The results also challenge our expectations about source–value alignment. We hypothesized that local governments would be more trusted when addressing financial or practical concerns (H2), and that national governments would be more trusted when emphasizing environmental or stability-related values (H3). The empirical evidence does not support these assumptions. Trust does not increase systematically when value frames appear to match presumed governance responsibilities. Instead, responses depend more on citizens' prior trust orientations than on the formal alignment between source and value emphasis. This suggests that perceived relational credibility outweighs institutional role congruence in shaping trust reactions.

Citizens do not passively receive policy messages; they interpret them through existing relational expectations. The six latent trust profiles reveal that baseline trust levels strongly shape communication sensitivity. Individuals in certain clusters respond negatively to specific value frames, while others respond positively to the same cues. Importantly, this heterogeneity does not map cleanly onto socio-demographic characteristics. Age and baseline trust are more relevant moderators than education, income, or household composition. Communication strategies based on demographic segmentation are therefore unlikely to be effective. What matters is the relational position citizens already occupy vis-à-vis government.

A further conclusion concerns the comparative effect of communication versus participation. While value-based framing does not increase trust, participation in the PVE is associated with small but consistent increases in trust across conditions. These effects differ across trust classes, with lower-trust groups showing increases in trust and higher-trust groups showing decreases. Regardless of treatment assignment, respondents report slightly higher levels of trust after engaging with the participatory exercise. Reflecting on trade-offs and being invited to provide input appears to foster perceptions of procedural fairness and responsiveness that simple framing cannot achieve. Participation thus emerges as a more promising mechanism for sustaining or strengthening institutional trust than attempts to optimize rhetorical cues. Finally, the empirical patterns observed across the survey items align with the three-dimensional conceptualization of institutional trust proposed by Grimmelikhuijsen and Knies (2017), providing empirical support for the competence, integrity, and benevolence dimensions in the context of energy transition communication.

Taken together, these findings indicate that trust in the heat transition is shaped less by the strategic selection of value frames and more by pre-existing institutional relationships and participatory experience. Value-based messages function as relational signals that may intensify existing trust gradients rather than shift them. The challenge for policymakers is therefore not to identify a universally persuasive value cue, but to communicate in ways that are transparent, proportionate, and embedded within genuine participatory processes. Recognizing the limits of framing and the relational nature of trust can help governments design communication strategies that support legitimacy during the transition away from natural gas.

CRediT authorship contribution statement

Kevin Goes: Writing – review & editing, Writing – original draft, Visualization, Validation, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Louison Thépaut:** Writing – original draft, Visualization, Validation, Methodology, Formal analysis, Conceptualization. **Niek Mouter:** Writing – original draft, Supervision, Software, Methodology. **Emile Chappin:** Writing – original draft, Supervision, Methodology. **Sjoerd Kluiving:** Writing – original draft, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This publication is part of the project ALIGN4energy (with project number NWA.1389.20.251) of the research programme NWA ORC 2020 which is (partly) financed by the Dutch Research Council (NWO).

We would like to thank Daan van Soest for his thoughtful and constructive feedback on an earlier version of this paper, which provided valuable input on both the empirical approach and the broader framing and interpretation of our findings.

Appendix A

A.1. Control

The control group gets presented the topic of the PVE, without any additional information. It is neither from the perspective of a local or national government. Furthermore, none of the value frames are present in the short message.

- *"De overheid zoekt manieren om energiesystemen te verbeteren. Warmtenetten spelen hierbij een grote rol, omdat ze een manier bieden om huizen en bedrijven te verwarmen met centraal opgewekte warmte. Uw mening is belangrijk om te begrijpen hoe deze verandering past bij de wensen en behoeften van de samenleving."*

A.2. T1: Environmental Values - Local Level

This treatment focuses on the local environmental benefits of heat grids, such as reducing emissions and improving air quality. The emphasis is on how these projects can help protect the natural environment in the municipality. Participants in this group will be encouraged to consider how local actions contribute to a cleaner, healthier community.

- *"Onze gemeente zoekt manieren om energiesystemen te verbeteren, zodat de energievoorziening duurzamer en milieuvriendelijker wordt. Warmtenetten spelen hierbij een grote rol, omdat ze een duurzame manier bieden om huizen en bedrijven te verwarmen en uitstoot van broeikasgassen kunnen verminderen.*
  Door lokaal actie te ondernemen, zorgen we voor een schonere en gezondere omgeving, nu en in de toekomst. Uw mening is belangrijk om te begrijpen hoe deze verandering past bij de wensen en behoeften van de inwoners van onze gemeente."

A.3. T2: Environmental Values - National Level

Participants in this group will be introduced to the national-level environmental benefits of heat grids, with a focus on reducing nationwide greenhouse gas emissions and promoting cleaner energy use. The treatment highlights how the municipality's actions align with national sustainability goals and the broader fight against climate change.

- *"De rijksoverheid zoekt manieren om energiesystemen te verbeteren, zodat de energievoorziening duurzamer en milieuvriendelijker wordt. Warmtenetten spelen hierbij een grote rol, omdat ze een duurzame manier bieden om huizen en bedrijven te verwarmen en mogelijk uitstoot van broeikasgassen te verminderen.*
  Door actie te ondernemen, zorgen we voor een schonere en gezondere samenleving, nu en in de toekomst. Uw mening is belangrijk om te begrijpen hoe deze verandering past bij de wensen en behoeften van de maatschappij."

A.4. T3: Financial Values - Local Level

This treatment emphasizes the individual financial benefits of heat grids, such as cost savings and improved energy efficiency. The focus is on how local residents can benefit directly from these systems through lower energy bills and greater comfort. Participants will be asked to consider how these changes might impact their personal financial well-being.

- *"Onze gemeente zoekt manieren om energiesystemen te verbeteren, zodat de energievoorziening goedkoper en comfortabeler wordt. Warmtenetten spelen hierbij een grote rol, omdat ze een efficiënte manier bieden om huizen en bedrijven te verwarmen en mogelijk stookkosten verminderen.*
  Door lokaal actie te ondernemen, zorgen we voor een sterke en weerbare lokale economie, nu en in de toekomst. Uw mening is belangrijk om te begrijpen hoe deze verandering past bij de wensen en behoeften van inwoners van onze gemeente."

A.5. T4: Financial Values - National Level

Participants in this group will be presented with the financial implications of heat grids from a national perspective. The treatment discusses how these systems could help reduce overall energy costs and improve energy affordability at the national level, while also benefiting individual households. The aim is to emphasize both personal and national financial advantages.

- *"De rijksoverheid zoekt manieren om energiesystemen te verbeteren, zodat de energievoorziening goedkoper en comfortabeler wordt. Warmtenetten spelen hierbij een grote rol, omdat ze een efficiënte manier bieden om huizen en bedrijven te verwarmen en mogelijk stookkosten verminderen.*
  Door actie te ondernemen, zorgen we voor een sterke en weerbare economie, nu en in de toekomst. Uw mening is belangrijk om te begrijpen hoe deze verandering past bij de wensen en behoeften van de samenleving."

A.6. T5: Stability - Local Level

This group highlights the role of heat grids in enhancing local energy independence. The focus is on how local municipalities can reduce their reliance on external energy suppliers and create a more resilient energy system, that has economic benefits in the long term. Participants are asked to consider how these changes can contribute to the security and self-sufficiency of their own community.

- *"Onze gemeente zoekt manieren om energiesystemen te verbeteren, zodat de energievoorziening veerkrachtiger en zekerder wordt. Warmtenetten spelen hierbij een grote rol, omdat ze een doelmatige manier bieden om huizen en bedrijven te verwarmen en mogelijk afhankelijkheid van buitenlands gas te verminderen.*
  Door lokaal actie te ondernemen, zorgen we voor een onafhankelijke en zekere energievoorziening, nu en in de toekomst. Uw mening is belangrijk om te begrijpen hoe deze verandering past bij de wensen en behoeften van inwoners van onze gemeente."

A.7. T6: Stability - National Level

In this treatment, participants will learn about the national-level implications of heat grids for energy independence, with particular attention to reducing reliance on foreign energy sources, especially in light of geopolitical challenges like the war in Ukraine. The treatment underscores how local municipalities' efforts contribute to national energy security and resilience. Which would provide economic and financial stability.

- *"De rijksoverheid zoekt manieren om energiesystemen te verbeteren, zodat de energievoorziening veerkrachtiger en zekerder wordt. Warmtenetten spelen hierbij een grote rol, omdat ze een doelmatige manier bieden om huizen en bedrijven te verwarmen en mogelijk afhankelijkheid van buitenlands gas te verminderen.*

  Door actie te ondernemen, zorgen we voor een onafhankelijke en zekere energievoorziening, nu en in de toekomst. Uw mening is belangrijk om te begrijpen hoe deze verandering past bij de wensen en behoeften van de samenleving."

Appendix B. Factor analysis

B.1. Procedure

To validate the dimensional structure of the trust questions used, we conducted an exploratory factor analysis on the 12 pre-treatment items. Because all variables were measured on 5-point Likert scales, a polychoric correlation matrix was computed. Sampling adequacy was evaluated using the Kaiser–Meyer–Olkin (KMO) statistic and Bartlett's test of sphericity.

The KMO value was excellent (KMO = 0.967), and Bartlett's test was highly significant ($\chi^2 = 40{,}966.8$, df = 66, $p < .001$), indicating that the data were suitable for factor analysis.

A parallel analysis suggested extraction of three factors, consistent with prior work by Grimmelikhuijsen and Knies (2017). Maximum likelihood (ML) extraction with promax rotation was applied.

B.2. Parallel analysis

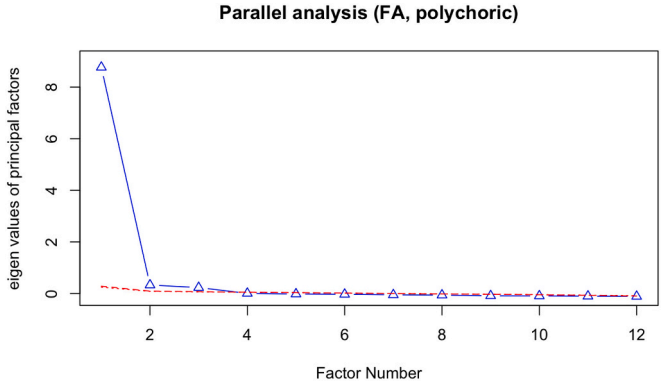


Fig. B.1. Parallel analysis.

B.3. Item-factor assignment

Table B.1
Factor scores.

Item	Factor 1 (ML1)	Factor 2 (ML2)	Factor 3 (ML3)
pre_sincere_num	0.120	–	0.756
pre_commitment_num	0.169	–	0.723
pre_honest_num	0.130	–	0.748
pre_capable_num	0.639		0.147
pre_effective_num	0.766		0.104
pre_skillful_num	0.881		
pre_expert_num	0.698	0.124	
pre_duty_num	0.470	0.173	0.273
pre_best_num	–	0.709	
pre_interest_num	–	0.808	
pre_wellbeing_num	–	0.855	
pre_approach_num	0.122	0.665	0.114

Note: Factor 2 (ML2) loads strongly on competence indicators (capable, effective, skilful, expert). Factor 1 (ML1) loads strongly on benevolence items (best, interest, wellbeing, approach). Factor 3 (ML3) loads strongly on integrity items (sincere, honest, commitment).
Reliability: Cronbach's α = 0.96, Omega Total = 0.968, Omega Hierarchical = 0.910

Appendix C. Latent trust profiles

C.1. Construction of the trust index

To identify latent profiles of institutional trust, we first constructed a single trust score at before the treatment and PVE. The score was derived from the 12 pre-treatment trust items using the factor loadings from the exploratory factor analysis (Appendix B). Specifically, each respondent's score is the weighted mean of observed items:

$$\text{pre trust}_i = \frac{\Sigma_j w_j x_{ij}}{\Sigma_j w_j . 1\left(x_{ij}\text{observed}\right)}$$

where w_j denotes the factor loading of item j on the general trust factor, and x_{ij} the corresponding Likert response (1–5). Scores were only computed for respondents with at least six answered items. The same weights were used to construct the post-treatment trust score. These scores were standardized and served as input for latent profiles.

C.2. Latent profile analysis of baseline trust

We conducted a latent profile analysis on the standardized baseline trust scores. Comparing models with 1 to 6 components based on the Bayesian Information Criterion (BIC), showed us that 6 components with equal variance generated the best fit. We labelled these profiles in ascending order of mean trust.

Table C.1
Latent trust profiles.

Trust class	n	% of sample	Mean baseline trust (z)	SD	Mean posterior probability	Mean classification uncertainty
1	428	14.29	−1.53	0.22	0.94	0.06
2	1019	34.01	−0.54	0.21	0.50	0.50
3	86	2.87	−0.10	0.02	0.38	0.62
4	895	29.87	0.40	0.26	0.87	0.13
5	465	15.52	1.33	0.27	0.89	0.11
6	103	3.44	2.32	0.27	0.93	0.07

Note: By construction, mean standardized trust increases monotonically from Class 1 (lowest trust) to Class 6 (highest trust). Average maximum posterior probabilities are high and average uncertainties low, indicating well-separated latent classes.

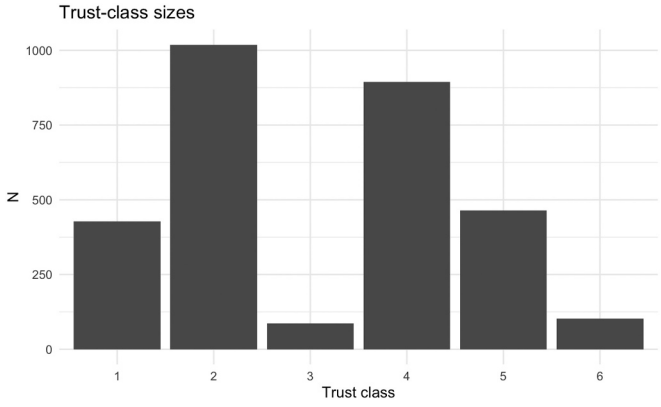


Fig. C.1. Trust-class sizes.

Table C.2
Trust-class Pre vs Post.

Trust class	n (pre)	Mean pre	95% CI pre	n (post)	Mean post	95% CI post
1	428	1.16	[1.14; 1.17]	428	1.26	[1.22; 1.30]
2	1019	2.04	[2.03; 2.06]	1014	2.12	[2.09; 2.15]
3	86	2.44	[2.44; 2.44]	86	2.55	[2.45; 2.66]
4	895	2.90	[2.88; 2.91]	887	2.91	[2.88; 2.95]
5	465	3.73	[3.71; 3.76]	463	3.61	[3.57; 3.66]
6	103	4.62	[4.58; 4.67]	103	4.34	[4.21; 4.46]

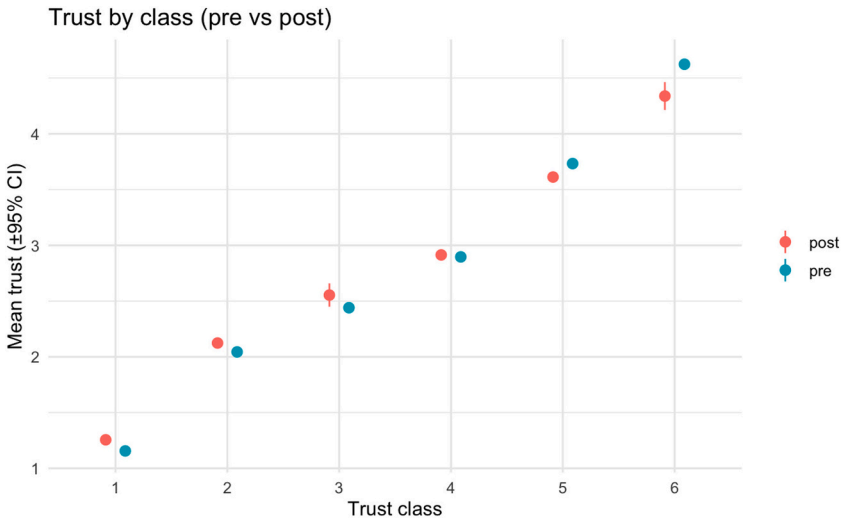


Fig. C.2. Trust-class Pre vs Post.

Appendix D. Difference in difference estimation

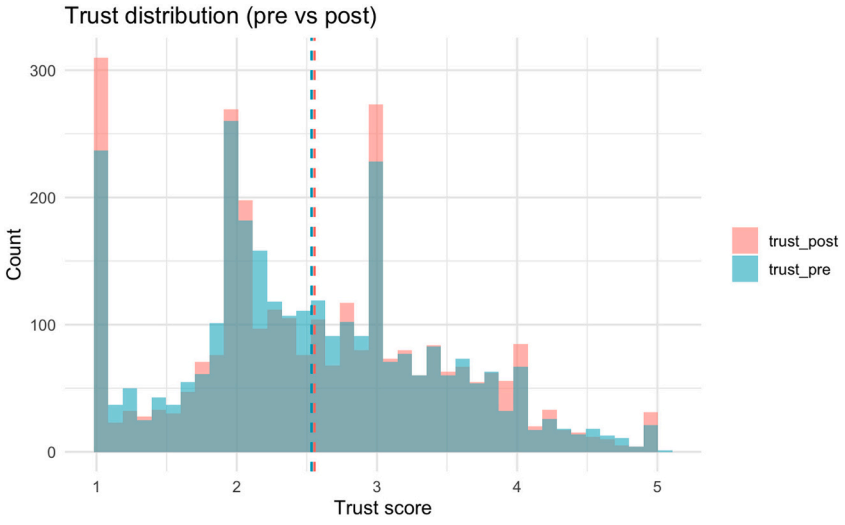


Fig. D.1. Pre vs Post levels of trust.

Table D.1
Multiple treatment arms.

Variable	DiD	DDD	DiD + Demo	DDD + Demo
Post	0.054**	0,048	0.049*	0,048
	(0.025)	(0.048)	(0.025)	(0.048)
LocalEnviron	−0,019	0	−0,013	0,001
	(0.014)	(0.029)	(0.015)	(0.030)
NatEnviron	−0,007	0,024	−0,002	0,024
	(0.015)	(0.029)	(0.016)	(0.030)
LocalFinan	−0,018	0,006	−0,019	−0.000
	(0.015)	(0.030)	(0.015)	(0.030)
NatFinan	0,008	0,04	0,019	0,044
	(0.014)	(0.029)	(0.015)	(0.030)
LocalStab	0,006	0.052*	0,014	0.054*
	(0.015)	(0.030)	(0.015)	(0.031)
NatStab	−0.010	0,02	−0,001	0,026
	(0.015)	(0.029)	(0.016)	(0.030)
Post × LocalEnviron	−0,054	−0,072	−0,048	−0,072
	(0.034)	(0.065)	(0.035)	(0.066)

(*continued on next page*)

Table D.1 (*continued*)

Variable	DiD	DDD	DiD + Demo	DDD + Demo
Post × NatEnviron	−0,001	0,052	0,004	0,054
	(0.034)	(0.062)	(0.035)	(0.062)
Post × LocalFinan	−0.066**	−0.113*	−0.064*	−0.113*
	(0.033)	(0.061)	(0.033)	(0.061)
Post × NatFinan	−0,022	0,032	−0,017	0,036
	(0.034)	(0.063)	(0.034)	(0.064)
Post × LocalStab	−0.059*	−0,065	−0.055*	−0,065
	(0.033)	(0.062)	(0.033)	(0.063)
Post × NatStab	−0.020	−0,045	−0,015	−0,045
	(0.033)	(0.061)	(0.033)	(0.061)
Post × LocalEnviron × Trust C1		0,136		0,148
		(0.108)		(0.109)
Post × NatEnviron × Trust C1		−0,058		−0,048
		(0.094)		(0.094)
Post × LocalFinan × Trust C1		0,061		0,074
		(0.090)		(0.090)
Post × NatFinan × Trust C1		−0.080		−0,072
		(0.097)		(0.098)
Post × LocalStab × Trust C1		0,089		0,099
		(0.098)		(0.099)
Post × NatStab × Trust C1		0,023		0,035
		(0.094)		(0.094)
Post × LocalEnviron × Trust C2		−0,038		−0,026
		(0.087)		(0.088)
Post × NatEnviron × Trust C2		−0,067		−0,063
		(0.087)		(0.088)
Post × LocalFinan × Trust C2		0,083		0,086
		(0.084)		(0.084)
Post × NatFinan × Trust C2		−0,054		−0,051
		(0.086)		(0.087)
Post × LocalStab × Trust C2		0,018		0,025
		(0.085)		(0.085)
Post × NatStab × Trust C2		0,031		0,037
		(0.088)		(0.088)
Post × LocalEnviron × Trust C3		−0,147		−0,147
		(0.209)		(0.210)
Post × NatEnviron × Trust C3		−0.348**		−0.350**
		(0.171)		(0.172)
Post × LocalFinan × Trust C3		0,043		0,043
		(0.212)		(0.212)
Post × NatFinan × Trust C3		−0.320*		−0.324*
		(0.187)		(0.188)
Post × LocalStab × Trust C3		−0.410***		−0.410***
		(0.143)		(0.143)
Post × NatStab × Trust C3		−0,001		−0,001
		(0.151)		(0.152)
Post × LocalEnviron × Trust C5		0,028		0,028
		(0.107)		(0.108)
Post × NatEnviron × Trust C5		−0,037		−0,039
		(0.095)		(0.095)
Post × LocalFinan × Trust C5		−0,087		−0,093
		(0.109)		(0.110)
Post × NatFinan × Trust C5		−0,063		−0,067
		(0.102)		(0.103)
Post × LocalStab × Trust C5		−0,032		−0,032
		(0.105)		(0.105)
Post × NatStab × Trust C5		0,067		0,066
		(0.093)		(0.093)
Post × LocalEnviron × Trust C6		0,239		0,239
		(0.177)		(0.178)
Post × NatEnviron × Trust C6		−0,138		−0.140
		(0.306)		(0.307)
Post × LocalFinan × Trust C6		0.357**		0.357**
		(0.175)		(0.176)
Post × NatFinan × Trust C6		−0,075		−0.080
		(0.208)		(0.209)
Post × LocalStab × Trust C6		0,12		0,12
		(0.172)		(0.173)
Post × NatStab × Trust C6		0,081		0,081
		(0.223)		(0.224)
Trust C1	−1.698***	−1.753***	−1.655***	−1.712***
	(0.017)	(0.031)	(0.018)	(0.033)
Trust C2	−0.821***	−0.791***	−0.804***	−0.784***
	(0.014)	(0.026)	(0.015)	(0.027)
Trust C3	−0.410***	−0.439***	−0.404***	−0.457***
	(0.029)	(0.021)	(0.029)	(0.026)

(*continued on next page*)

Table D.1 (*continued*)

Variable	DiD	DDD	DiD + Demo	DDD + Demo
Trust C5	0.767***	0.871***	0.743***	0.844***
	(0.018)	(0.035)	(0.019)	(0.037)
Trust C6	1.573***	1.755***	1.531***	1.739***
	(0.039)	(0.071)	(0.039)	(0.072)
Post × Trust C1		0,058		0,046
		(0.074)		(0.074)
Post × Trust C2		0,069		0,062
		(0.066)		(0.067)
Post × Trust C3		0.250**		0.250**
		(0.123)		(0.124)
Post × Trust C5		−0.125*		−0.125*
		(0.074)		(0.074)
Post × Trust C6		−0.373***		−0.373***
		(0.144)		(0.145)
LocalEnviron × Trust C1		0,024		0,027
		(0.043)		(0.045)
NatEnviron × Trust C1		0,02		0,028
		(0.046)		(0.049)
LocalFinan × Trust C1		0,014		0,009
		(0.044)		(0.046)
NatFinan × Trust C1		0,031		0,034
		(0.044)		(0.046)
LocalStab × Trust C1		0,002		0,001
		(0.046)		(0.046)
NatStab × Trust C1		0,003		0,016
		(0.047)		(0.049)
LocalEnviron × Trust C2		−0,043		−0,027
		(0.036)		(0.037)
NatEnviron × Trust C2		−0.078**		−0.066*
		(0.037)		(0.038)
LocalFinan × Trust C2		−0,051		−0.040
		(0.037)		(0.038)
NatFinan × Trust C2		−0.090**		−0.076**
		(0.036)		(0.038)
LocalStab × Trust C2		−0.095**		−0.083**
		(0.037)		(0.039)
NatStab × Trust C2		−0.070*		−0.065*
		(0.036)		(0.038)
LocalEnviron × Trust C3		0,013		0,046
		(0.030)		(0.046)
NatEnviron × Trust C3		−0,018		−0.020
		(0.030)		(0.042)
LocalFinan × Trust C3		0,004		0,058
		(0.031)		(0.038)
NatFinan × Trust C3		−0,031		−0,009
		(0.030)		(0.042)
LocalStab × Trust C3		−0.057*		0,006
		(0.031)		(0.043)
NatStab × Trust C3		−0,026		−0,028
		(0.029)		(0.037)
LocalEnviron × Trust C5		−0,023		−0,031
		(0.051)		(0.052)
NatEnviron × Trust C5		−0,059		−0.060
		(0.049)		(0.050)
LocalFinan × Trust C5		−0,019		−0,005
		(0.052)		(0.054)
NatFinan × Trust C5		−0,023		−0,013
		(0.049)		(0.051)
LocalStab × Trust C5		−0,056		−0,048
		(0.051)		(0.052)
NatStab × Trust C5		−0,059		−0.050
		(0.053)		(0.055)
LocalEnviron × Trust C6		−0,069		−0,098
		(0.107)		(0.107)
NatEnviron × Trust C6		−0,001		−0,028
		(0.096)		(0.096)
LocalFinan × Trust C6		−0,059		−0,089
		(0.098)		(0.101)
NatFinan × Trust C6		−0,015		−0,048
		(0.089)		(0.090)
LocalStab × Trust C6		−0.110		−0,144
		(0.091)		(0.092)
NatStab × Trust C6		0,042		0,023
		(0.101)		(0.102)
Secondary education (HAVO/VWO years 1–3)			−0,024	−0,016
			(0.075)	(0.075)

(*continued on next page*)

Table D.1 (*continued*)

Variable	DiD	DDD	DiD + Demo	DDD + Demo
Secondary education (HAVO/VWO years 4–6)			0,016	0,02
			(0.066)	(0.067)
Applied sciences / professional bachelor (HBO)			−0,025	−0,021
			(0.065)	(0.065)
Vocational education (MBO level 1)			−0,008	−0,003
			(0.073)	(0.073)
Vocational Education (MBO levels 2, 3 or 4)				
University (WO)			−0,017	−0.010
			(0.067)	(0.067)
Vocational secondary education (VMBO)			−0,015	−0.010
			(0.065)	(0.065)
Education: Don't know / Prefer not to say			0,013	0,01
			(0.117)	(0.120)
Making ends meet: Difficult			−0.055**	−0.054**
			(0.024)	(0.024)
Making ends meet: Neither easy nor difficult			−0,004	−0,002
			(0.015)	(0.015)
Making ends meet: Very easy			−0,006	−0,004
			(0.023)	(0.023)
Making ends meet: Very difficult			−0,062	−0,061
			(0.040)	(0.039)
Making ends meet: Don't know / Prefer not to say			−0,045	−0,066
			(0.079)	(0.079)
Additional monthly expenses: < €100			−0,048	−0,043
			(0.032)	(0.033)
Additional monthly expenses: €100–150			−0,012	−0,011
			(0.030)	(0.030)
Additional monthly expenses: €150–200			−0,014	−0.010
			(0.030)	(0.030)
Additional monthly expenses: €200–250			−0,011	−0,008
			(0.031)	(0.031)
Additional monthly expenses: €250–300			−0,003	0
			(0.031)	(0.031)
Additional monthly expenses: No / Don't know			0,016	0,018
			(0.039)	(0.039)
Additional monthly expenses: Prefer not to say			0,011	0,014
			(0.046)	(0.046)
Affordability: Difficult to afford			−0.046**	−0.047**
			(0.023)	(0.023)
Affordability: Neither easy nor difficult			−0.037**	−0.039**
			(0.015)	(0.015)
Affordability: Don't know / Prefer not to say			−0.155***	−0.150***
			(0.045)	(0.046)
Affordability: Very easy to afford			−0,002	−0,005
			(0.025)	(0.025)
Affordability: Very difficult to afford			−0.089**	−0.086**
			(0.042)	(0.042)
Household: Living alone			0,013	0,009
			(0.032)	(0.032)
Household: Partner with children				
Household: Partner without children			0,021	0,021
			(0.032)	(0.032)
Household: Shared housing			0,074	0,072
			(0.048)	(0.049)
Household: Single parent				
Household: Prefer not to say			0,002	0,001
			(0.144)	(0.144)
Occupation: Homemaker			−0,009	−0,011
			(0.034)	(0.034)
Occupation: Retired			0,021	0,021
			(0.027)	(0.027)
Occupation: Paid employment			0,015	0,015
			(0.024)	(0.024)
Occupation: Volunteer work			0,062	0,057
			(0.042)	(0.043)
Occupation: Student			−0,016	−0,023
			(0.040)	(0.040)
Occupation: Don't know / Prefer not to say			−0,058	−0,056
			(0.114)	(0.115)
Climate concern: Very concerned			0,011	0,011
			(0.015)	(0.015)
Climate concern: Not at all			−0.083***	−0.083***
			(0.021)	(0.021)

(*continued on next page*)

Table D.1 (*continued*)

Variable	DiD	DDD	DiD + Demo	DDD + Demo
Climate concern: Not really			−0,022	−0,022
			(0.015)	(0.015)
Climate concern: Don't know / Prefer not to say			0,036	0,032
			(0.078)	(0.080)
Gender: Male			−0,029	0,003
			(0.058)	(0.098)
Gender: Female			−0,049	−0,015
			(0.057)	(0.097)
Gender: Non-binary			−0,189	−0,143
			(0.181)	(0.197)
Gender: Prefer not to say			−0.241***	−0,184
			(0.083)	(0.117)
Age			−0.046***	−0.046***
			(0.009)	(0.009)
Controls (demographics)	NO	NO	YES	YES
N	5962	5962	5936	5936
R^2	0,824	0,828	0,829	0,833

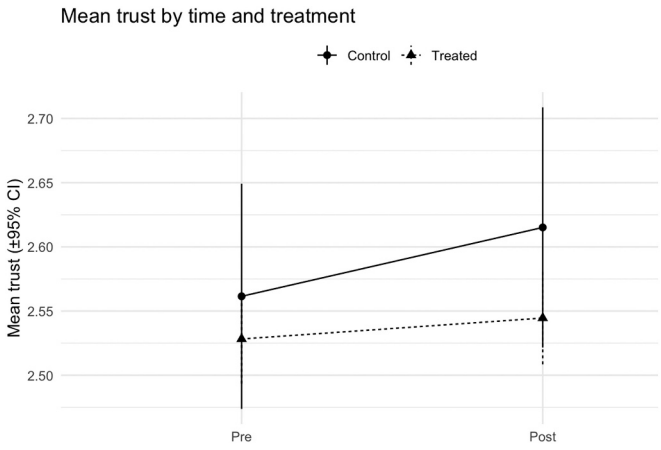


Fig. D.2. DiD Treated vs Control.

D.1. Multi-arm treatment interpretation

The multi-arm DiD models paint the same broad picture as the pooled: no overall treatment effect, but meaningful and structured heterogeneity across treatment variants and trust clusters.

The Post effect remains positive across all four specifications, indicating a general increase in trust over time. None of the treatment arms (TreatmentGroup k) differ significantly from the control group at baseline, confirming the success of randomization in balancing trust across arms.

The arm-specific DiD interactions (post × TreatmentGroup k) show that only a small number of treatment groups exhibit meaningful effects. Treatment Group 5 stands out, displaying consistent and statistically significant negative DiD estimates in the CATE and Demo+CATE specifications (-0.43 to -0.49, $t > 3$). Treatment Group 3 also shows a modest negative DiD in the baseline CATE model. All other arms produce small, noisy, and statistically insignificant effects.

When we allow for heterogeneity in the DiD (post × TreatmentGroup k × trust_class j), a richer pattern emerges:

- For lower-trust clusters (1 and 2), treatment 5 produces positive significant effects, indicating that this subgroup reacts differently from the rest of the population.
- For higher-trust clusters (4, 5, and 6), several arms, again especially Group 5, produce significantly negative interaction effects, suggesting a trust decline post-treatment for these groups.
- The direction of effects differs strongly between clusters.

  These opposing signs explain the absence of an overall treatment effect.

The trust class coefficients are nearly identical to the pooled models: classes 1–2 show substantially lower trust; classes 4–6 show substantially higher trust. These large and stable effects confirm that trust clusters are the dominant explanatory factor in the dataset ($R^2 \sim 0.83$).

The demographic effects again show the same structure as the pooled models: financial hardship and low climate concern consistently predict lower trust, and age carries a small but significant negative effect. Only significant demographic coefficients appear in the tables (all models include the full set; non-significant coefficients are suppressed for readability).

Taken together, the multi-arm models confirm that the treatment does not produce a systematic increase in trust, but that specific treatment variants can lead to positive or negative effects within certain trust clusters. These heterogeneous responses cancel each other out when averaged, explaining the null overall result.

Appendix E. Honest causal forest estimation

We complement the DiD models with honest causal forests, using demographics and baseline trust as covariates. Two forests are estimated: one with post-treatment trust score as the outcome, and one with the pre–post change in trust (dtrust). In both cases, the average treatment effect is close to zero, confirming the null results from the parametric DiD models. However, the forests reveal substantial heterogeneity in treatment effects across subgroups.

For the forest on trust change (Y = dtrust), the Conditional Average Treatment Effect (CATE) for a given subgroup measures how much more (or less) trust changes over time for treated individuals. Positive CATEs indicate that the treatment leads to a larger increase (or smaller decline) in trust between pre and post, while negative CATEs indicate that the treatment dampens trust gains or accelerates trust losses.

E.1. Subgroup bar plots (CATE by category, with 95% CI)

For each demographic, we estimate subgroup CATEs on both post-treatment trust and the change in trust (pre–post). Points with confidence intervals not crossing zero indicate subgroups for which the treatment significantly shifts trust relative to comparable controls. We highlight the variables for which significant heterogeneous effects emerge.

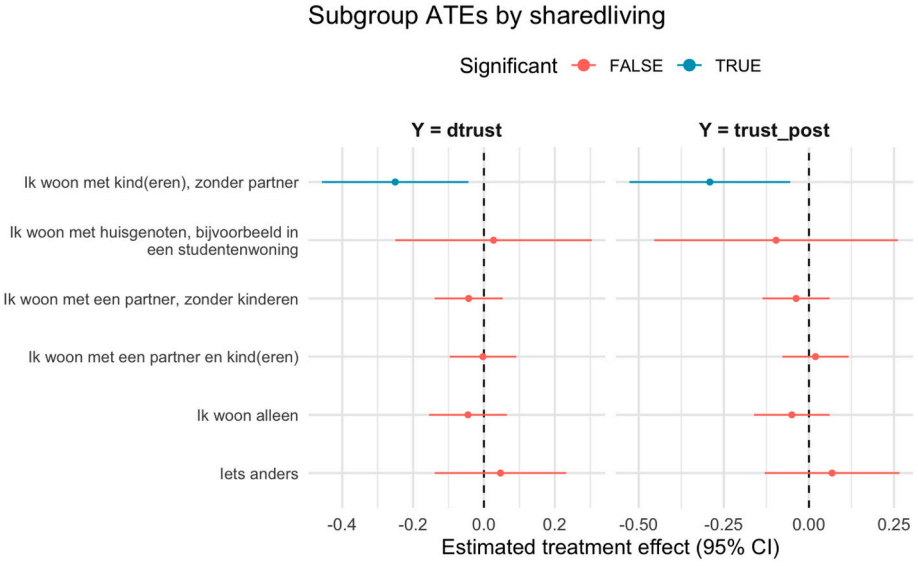


Fig. E.1. Shared living CATEs.

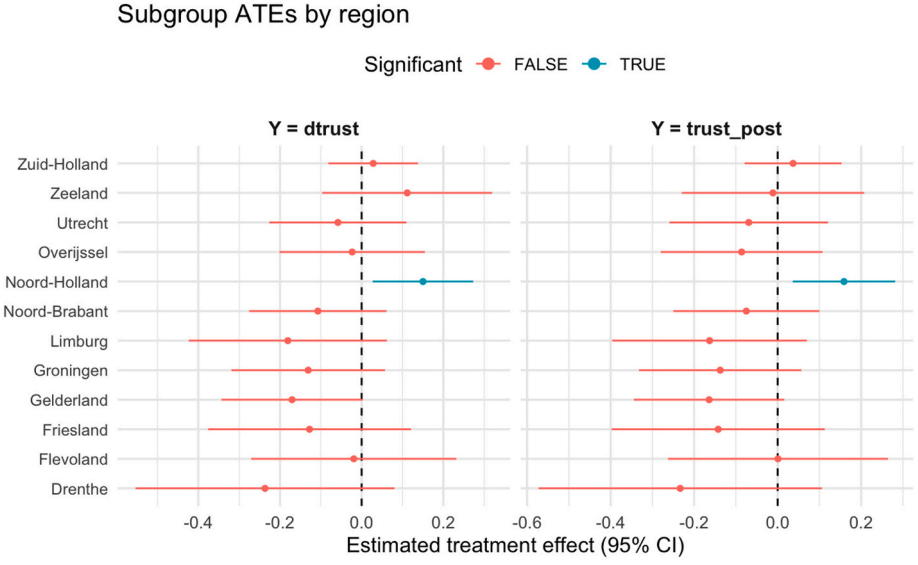


Fig. E.2. Regional CATEs.

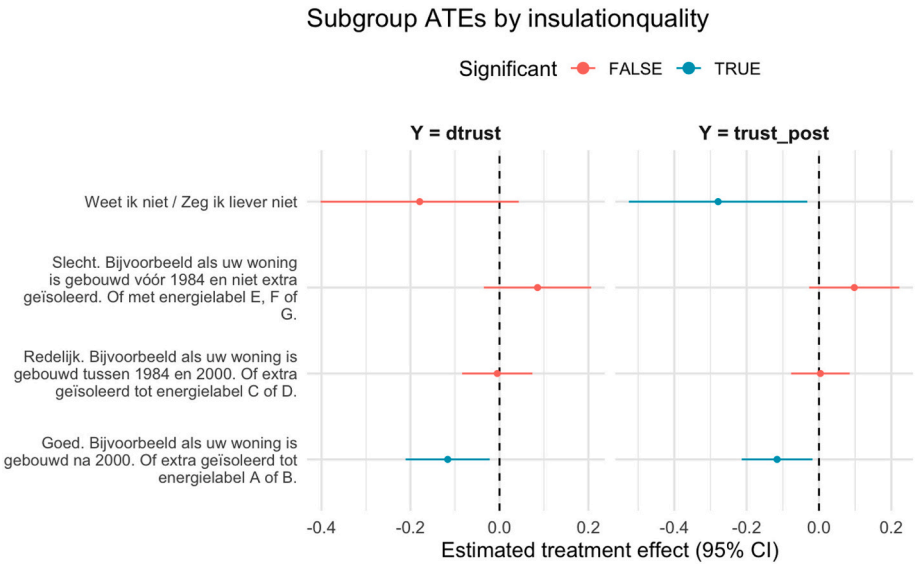


Fig. E.3. Insulation quality CATEs.

Across all demographic dimensions examined, the honest causal forest reveals limited evidence of meaningful heterogeneous treatment effects on trust. Most subgroup estimates center closely around zero, and the majority of confidence intervals include zero, indicating that for most types of households the treatment neither increases nor decreases trust in a statistically detectable way.

A small number of subgroups exhibit statistically significant effects, but these are idiosyncratic rather than systematic. For example, in the regional analysis, Noord-Holland stands out with a positive and statistically significant CATE for both *dtrust* and *trust_post*, implying a modestly stronger response to treatment among respondents from this province.

For housing-related characteristics, respondents living alone with children (sharedliving category) show a small negative effect on both *dtrust* and *trust_post*, while other household compositions exhibit null effects. In the insulation-quality analysis, households with the highest insulation grade show a small negative effect on post-treatment trust and a slightly negative effect on the pre–post change measure (*dtrust*), but these findings are again isolated and do not map onto a clear or theoretically predictable pattern of responsiveness.

Taken together, the subgroup plots indicate that heterogeneous effects exist but are sparse, unstable across outcomes, and not aligned with a coherent demographic or housing-related mechanism. The causal forest therefore reinforces the main DiD finding: the intervention does not shift trust on average, and while certain demographic subgroups exhibit departures from zero, these effects are weak, isolated, and unlikely to represent policy-relevant patterns of heterogeneity.

E.2. Variable importance calculations

To better understand which covariates drive heterogeneity in treatment effects, we compute permutation-based variable importance from the honest causal forest models. This approach measures how much the model's predictive accuracy deteriorates when the values of a given feature are randomly shuffled. Larger drops in predictive performance correspond to higher importance. Importantly, this method is agnostic to functional form, allowing for flexible non-linear and interaction effects, an essential property given that our CATE patterns already suggest strongly non-linear age and baseline-trust effects.

We report results separately for models predicting post-treatment trust and the change in trust (dtrust), using the top 15 predictors in each model.

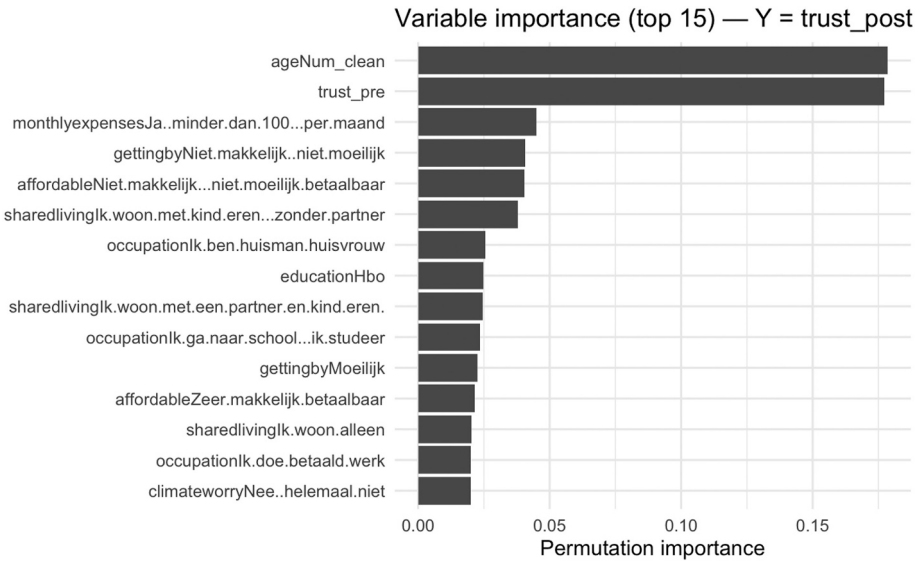


Fig. E.4. Variable Importance for post-treatment trust.

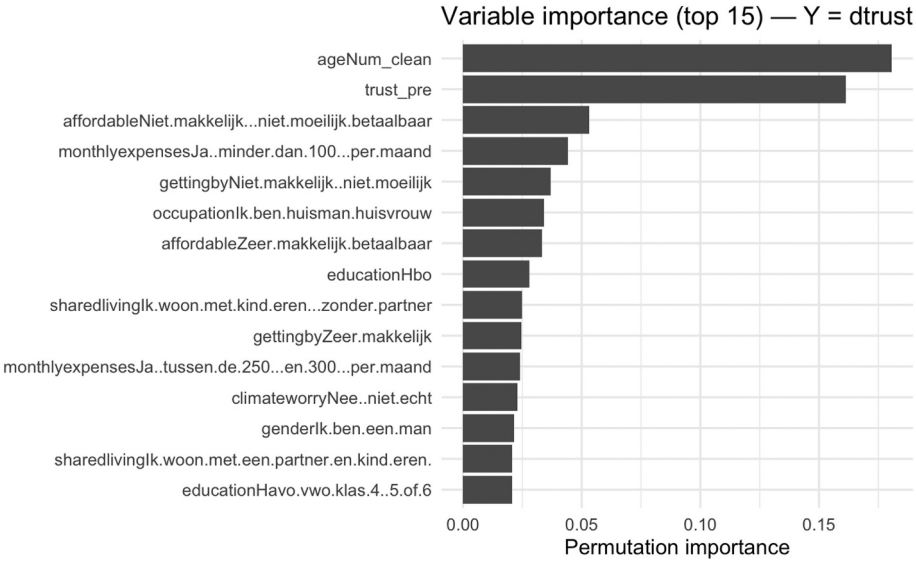


Fig. E.5. Variable Importance for change in trust.

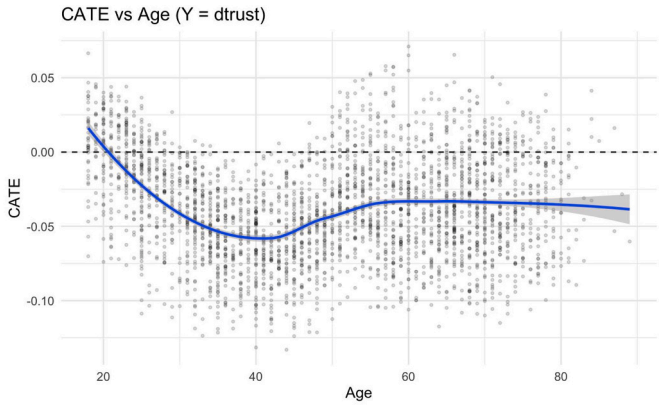


Fig. E.6. CATE vs age on change in trust.

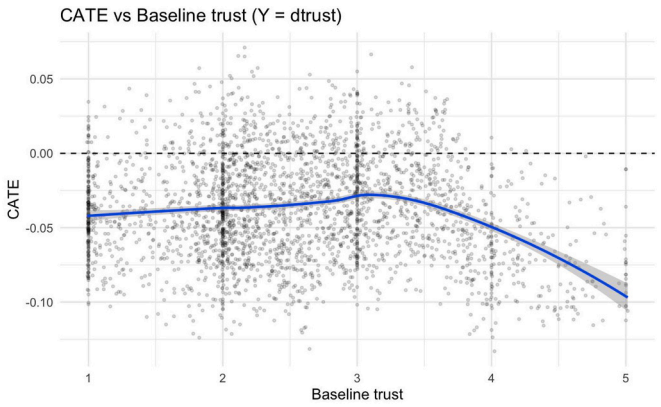


Fig. E.7. CATE vs baseline trust on change in trust.

Across both outcomes (trust_post and dtrust), the two most influential predictors are Age and Baseline trust (trust_pre). Both variables show very large importance scores, far above any other demographic variable:

- Age has a strong, non-linear moderating effect: younger respondents show slightly positive or near-zero effects, which become increasingly negative through mid-life, then partially recover at older ages.
- Baseline trust also exhibits clear heterogeneity: individuals with very high baseline trust experience notably more negative treatment effects on both outcomes.

The variable-importance analysis therefore demonstrates that age and initial trust beliefs are the primary axes of heterogeneity in this intervention. After age and baseline trust, the next most important variables cluster around:

- monthly expenses
- difficulty making ends meet (gettingby)
- perceived affordability of renovation measures

These results indicate that financial strain and perceived economic insecurity are key moderators. Individuals who report difficulty affording energy improvements or meeting monthly expenses generally exhibit more negative CATEs.

This pattern is consistent across both trust_post and dtrust, supporting the interpretation that financially constrained groups react less positively, or more skeptically, to interventions encouraging renovation uptake.

E.3. Best linear projection analysis

To understand which observable characteristics explain the small amount of heterogeneity detected by the causal forest, we estimate Best Linear Projections (BLPs) of the CATEs on the full set of covariates. BLPs provide a transparent, low-dimensional summary of the high-dimensional heterogeneity "learned" by the forest: each coefficient identifies how the conditional treatment effect *co-varies* with a given observable, holding all others constant. Positive coefficients indicate subgroups with larger (less negative) treatment effects, while negative coefficients correspond to more negative treatment effects.

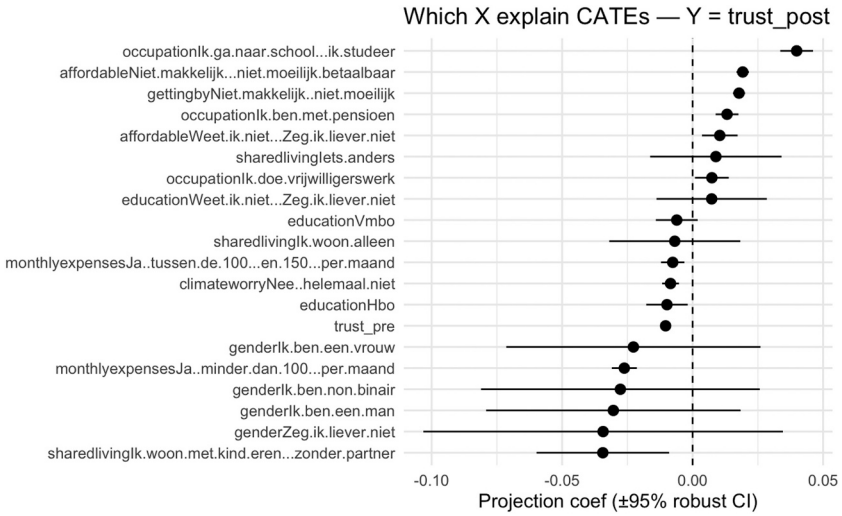


Fig. E.8. BLP post treatment trust.

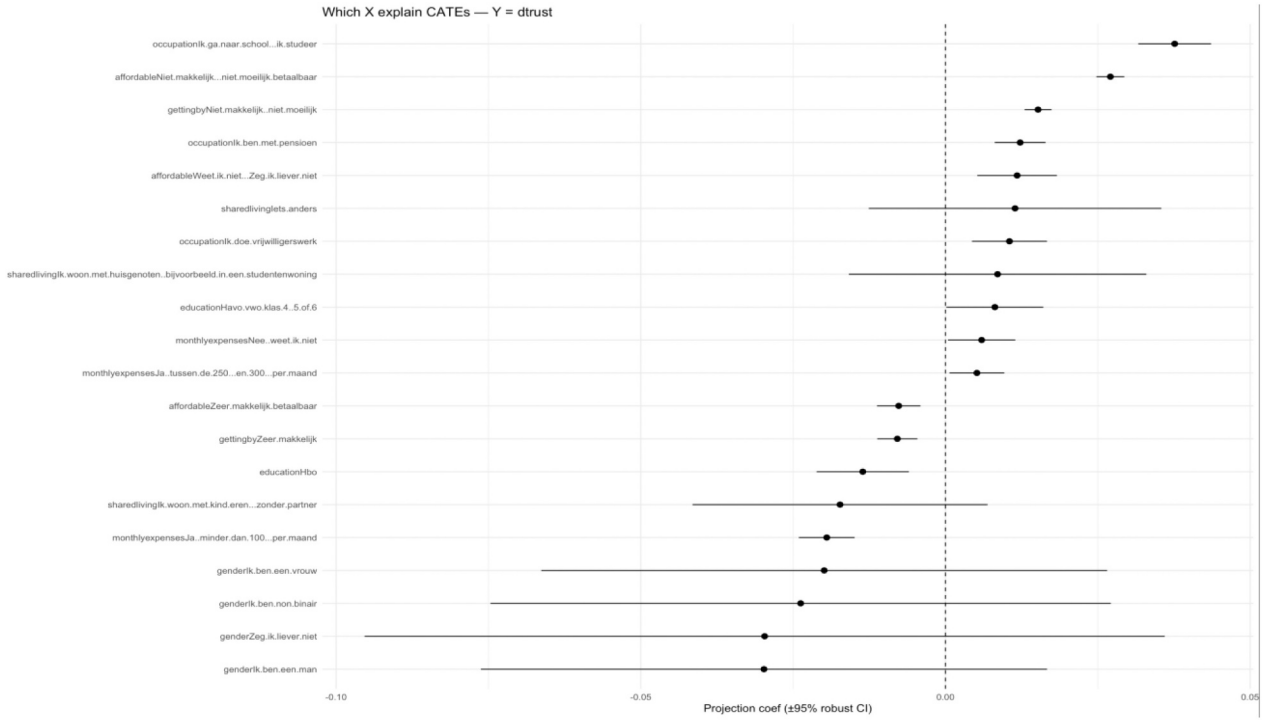


Fig. E.9. BLP change in trust.

Across both outcomes (*trust_post* and *dtrust*), the BLP coefficients confirm and sharpen the insights from the variable-importance plots: the heterogeneity learned by the forest is weak and concentrated in socio-economic and affordability-related variables, rather than in treatment-group indicators.

The majority of coefficients cluster tightly around zero, with wide confidence intervals that routinely cover zero. This pattern aligns with the null average treatment effect and the weak heterogeneity found in the subgroup plots. In other words, even when the model is forced to explain CATE variation linearly, most observable characteristics simply do not predict meaningful differences in treatment response.

A small number of coefficients emerge as statistically positive across both outcomes:

- Students (occupation: "ik ga naar school/ik studeer")
- Respondents reporting affordability as "niet makkelijk, niet moeilijk betaalbaar"
- Individuals indicating difficulties getting by financially ("Niet makkelijk, niet moeilijk")

These coefficients indicate that these groups experience slightly less negative treatment effects, i.e., they are marginally more receptive to the intervention. However, even these effects are small in magnitude and should not be overinterpreted.

At the other end of the spectrum, a few affordability categories (e.g. "Weet ik niet / zeg ik liever niet") and certain income brackets show slightly more negative CATEs. These patterns mirror the variable importance rankings, where affordability repeatedly appeared as the strongest non-age, non-baseline predictor.

Despite being one of the most important predictors of post-treatment trust in general, trust_pre has a near-zero BLP coefficient for explaining treatment heterogeneity. This suggests that people with low baseline trust were *not* systematically more or less influenced by the treatment, an important point for policy relevance.

Although individual estimates vary slightly across subgroups, the overall pattern is clear:

standard demographics do not explain the (limited) heterogeneity detected by the model. Most point estimates sit near zero with wide intervals.

Appendix F

Table F.1
Balance table across treatments.

Variable	SMD (max)	*p*-value	Stars	0	1	2	3	4	5	6
N				417	438	415	439	459	432	396
Age	0.166	0.266		47.88 (17.22)	49.67 (17.79)	50.74 (17.28)	48.28 (16.92)	49.59 (17.82)	48.79 (17.31)	49.13 (17.22)
Gender: I am a man	0.078	0.667		50.1%	52.3%	50.4%	51.7%	49.2%	47.7%	44.9%

(*continued on next page*)

Table F.1 (*continued*)

Variable	SMD (max)	*p*-value	Stars	0	1	2	3	4	5	6
Gender: I am a woman	0.078	0.667		49.6%	47.3%	49.2%	48.3%	50.5%	52.1%	55.1%
Gender: I am non-binary	0.078	0.667		0.2%	0.2%	0.2%		0.2%	0.2%	
Gender: Other	0.078	0.667			0.2%					
Gender: Prefer not to say	0.078	0.667				0.2%				
Education: Primary school	0.074	0.421		1.9%	1.1%	1.0%	1.8%	1.5%	1.4%	0.5%
Education: High school (VWO/HAVO), Years 1–3	0.074	0.421		3.4%	2.3%	2.7%	3.9%	2.8%	1.9%	1.8%
Education: High school (VWO/HAVO), Years 4–6	0.074	0.421		8.4%	6.2%	10.6%	6.4%	9.6%	6.7%	8.6%
Education: HBO	0.074	0.421		31.7%	29.9%	26.3%	27.3%	24.8%	28.0%	26.8%
Education: MBO Level 1	0.074	0.421		2.6%	1.4%	1.2%	2.5%	2.0%	2.5%	2.8%
Education: MBO Levels 2–4	0.074	0.421		26.9%	33.1%	27.7%	30.8%	31.6%	34.3%	34.1%
Education: University	0.074	0.421		12.7%	11.0%	15.2%	13.2%	13.7%	10.4%	10.1%
Education: VMBO	0.074	0.421		12.2%	15.1%	15.2%	14.1%	13.7%	14.6%	14.9%
Education: Don't know / Prefer not to say	0.074	0.421		0.2%		0.2%		0.2%	0.2%	0.5%
Getting by: Easy	0.066	0.796		40.3%	34.9%	35.9%	35.5%	34.4%	33.8%	38.1%
Getting by: Difficult	0.066	0.796		12.5%	10.0%	12.5%	10.7%	11.5%	12.5%	11.4%
Getting by: Neither easy nor difficult	0.066	0.796		34.1%	40.6%	36.9%	39.4%	38.3%	39.8%	34.3%
Getting by: Very easy	0.066	0.796		9.6%	9.6%	11.6%	11.4%	10.5%	11.1%	10.1%
Getting by: Very difficult	0.066	0.796		2.6%	4.1%	2.4%	2.5%	4.4%	2.3%	5.1%
Getting by: Prefer not to say / Don't know	0.066	0.796		1.0%	0.7%	0.7%	0.5%	0.9%	0.5%	1.0%
Monthly expenses: Yes, more than €300 per month	0.073	0.917		3.6%	5.0%	5.1%	5.2%	4.6%	5.3%	6.6%
Monthly expenses: Yes, less than €100 per month	0.073	0.917		13.2%	16.2%	13.3%	16.2%	13.9%	16.4%	14.6%
Monthly expenses: Yes, between €100 and €150 per month	0.073	0.917		21.8%	24.0%	21.7%	22.1%	23.1%	16.7%	20.5%
Monthly expenses: Yes, between €150 and €200 per month	0.073	0.917		22.8%	21.0%	24.6%	21.9%	23.1%	21.1%	23.7%
Monthly expenses: Yes, between €200 and €250 per month	0.073	0.917		17.0%	14.6%	16.9%	14.8%	15.7%	17.8%	14.1%
Monthly expenses: Yes, between €250 and €300 per month	0.073	0.917		12.5%	10.3%	10.4%	11.8%	10.0%	13.4%	12.4%
Monthly expenses: No, don't know	0.073	0.917		6.7%	6.6%	4.8%	5.5%	7.0%	6.5%	6.8%
Monthly expenses: Prefer not to say	0.073	0.917		2.4%	2.3%	3.4%	2.5%	2.6%	2.8%	1.3%
Affordability: Easy to afford	0.058	0.48		33.6%	30.4%	29.6%	31.9%	29.8%	30.1%	27.8%
Affordability: Difficult to afford	0.058	0.48		12.7%	14.6%	14.2%	14.1%	16.3%	16.2%	15.9%
Affordability: Neither easy nor difficult to afford	0.058	0.48		39.1%	40.4%	39.5%	37.6%	40.3%	36.6%	39.1%
Affordability: Don't know / Prefer not to say	0.058	0.48		2.9%	2.7%	1.2%	3.9%	3.1%	2.3%	1.5%
Affordability: Very easy to afford	0.058	0.48		7.9%	9.4%	11.6%	9.1%	6.3%	9.5%	11.1%
Affordability: Very difficult to afford	0.058	0.48		3.8%	2.5%	3.9%	3.4%	4.1%	5.3%	4.5%
Living situation: Something else	0.08	0.524		5.5%	5.0%	4.1%	4.6%	5.4%	5.1%	4.5%
Living situation: I live alone	0.08	0.524		26.4%	29.0%	23.9%	25.5%	27.9%	24.1%	26.8%
Living situation: I live with a partner and child(ren)	0.08	0.524		28.3%	26.7%	25.3%	28.2%	23.5%	23.4%	29.5%
Living situation: I live with a partner, without children	0.08	0.524		31.2%	34.2%	39.0%	33.3%	36.2%	38.7%	31.1%
Living situation: I live with housemates, for example in student housing	0.08	0.524		2.6%	0.9%	3.4%	2.7%	2.0%	1.9%	1.5%
Living situation: I live with child(ren), without a partner	0.08	0.524		5.8%	4.1%	4.1%	5.7%	5.0%	6.7%	6.3%
Living situation: Prefer not to say	0.08	0.524		0.2%		0.2%			0.2%	0.3%
Occupation: Other	0.06	0.169		6.2%	5.7%	7.2%	6.8%	4.8%	8.8%	6.6%
Occupation: I am a homemaker	0.06	0.169		5.8%	3.7%	5.3%	8.0%	5.0%	5.8%	6.6%
Occupation: I am retired	0.06	0.169		19.7%	23.7%	25.5%	19.6%	22.7%	21.3%	21.0%
Occupation: I am in paid employment	0.06	0.169		58.3%	60.3%	55.9%	59.2%	60.3%	55.1%	59.8%
Occupation: I do volunteer work	0.06	0.169		3.8%	3.4%	2.4%	2.1%	2.6%	2.3%	1.8%
Occupation: I go to school / I study	0.06	0.169		5.8%	3.0%	2.9%	4.3%	4.4%	5.6%	4.0%
Occupation: Don't know / Prefer not to say	0.06	0.169		0.5%	0.2%	0.7%		0.2%	1.2%	0.3%
Climate worry: Yes, a little	0.07	0.526		49.4%	43.6%	50.4%	48.7%	44.4%	47.7%	46.2%
Climate worry: Yes, very much	0.07	0.526		21.6%	21.5%	22.2%	21.4%	23.3%	20.1%	19.4%
Climate worry: No, not at all	0.07	0.526		9.8%	9.4%	8.4%	8.2%	11.5%	9.5%	10.4%
Climate worry: No, not really	0.07	0.526		19.2%	25.3%	18.3%	21.0%	20.3%	21.5%	23.5%
Climate worry: Don't know / Prefer not to say	0.07	0.526			0.2%	0.7%	0.7%	0.4%	1.2%	0.5%
Region: Drenthe	0.072	0.87		3.8%	3.2%	2.9%	3.2%	2.6%	4.2%	3.3%
Region: Flevoland	0.072	0.87		2.9%	2.3%	2.9%	4.3%	3.1%	2.1%	3.8%
Region: Friesland	0.072	0.87		3.6%	4.3%	4.1%	5.9%	3.7%	3.7%	4.5%
Region: Gelderland	0.072	0.87		9.8%	10.7%	9.9%	11.4%	9.2%	10.0%	9.1%
Region: Groningen	0.072	0.87		3.8%	4.6%	3.4%	4.6%	5.0%	3.5%	3.8%
Region: Limburg	0.072	0.87		9.1%	7.8%	6.7%	8.7%	7.6%	9.3%	6.8%
Region: North Brabant	0.072	0.87		12.7%	16.0%	15.2%	10.7%	13.7%	15.5%	14.6%
Region: North Holland	0.072	0.87		14.9%	15.3%	16.6%	12.5%	19.4%	15.7%	19.7%
Region: Overijssel	0.072	0.87		6.5%	5.7%	6.0%	7.3%	5.9%	8.1%	6.8%
Region: Utrecht	0.072	0.87		6.7%	3.9%	7.2%	6.4%	7.6%	6.5%	6.8%
Region: Zeeland	0.072	0.87		2.2%	3.2%	3.6%	3.4%	2.2%	2.5%	2.3%
Region: South Holland	0.072	0.87		24.0%	23.1%	21.4%	21.6%	20.0%	19.0%	18.4%
Housing type: Other	0.089	0.185		6.5%	5.3%	3.1%	4.6%	4.4%	5.6%	5.3%
Housing type: In an apartment, flat, upstairs apartment, or downstairs apartment	0.089	0.185		32.4%	34.9%	37.3%	33.0%	39.0%	30.1%	32.6%

(*continued on next page*)

Table F.1 (*continued*)

Variable	SMD (max)	*p*-value	Stars	0	1	2	3	4	5	6
Housing type: In a detached house, semi-detached house, or single-family home	0.089	0.185		60.7%	59.8%	59.0%	62.4%	56.6%	63.7%	61.6%
Housing type: Don't know / Prefer not to say	0.089	0.185		0.5%		0.5%			0.7%	0.5%
Heating method: Other	0.035	0.934		4.8%	5.3%	5.1%	3.9%	4.4%	5.6%	5.8%
Heating method: With a central heating boiler and a hybrid heat pump	0.035	0.934		6.0%	7.1%	6.5%	8.0%	7.4%	8.3%	7.1%
Heating method: With a central heating boiler. The boiler is located in your own home or elsewhere in the building.	0.035	0.934		69.1%	70.8%	67.2%	69.0%	70.6%	67.8%	68.9%
Heating method: With a fully electric heat pump	0.035	0.934		7.7%	7.8%	10.1%	8.0%	7.4%	9.5%	7.8%
Heating method: Via a heat grid, such as district heating	0.035	0.934		8.9%	8.0%	8.7%	9.1%	7.6%	6.7%	9.1%
Heating method: Don't know / Prefer not to say	0.035	0.934		3.6%	1.1%	2.4%	2.1%	2.6%	2.1%	1.3%
Insulation quality: Good. Energy label A or B.	0.049	0.868		35.3%	35.4%	39.5%	35.8%	36.6%	36.3%	35.4%
Insulation quality: Reasonable. Energy label C or D.	0.049	0.868		40.3%	43.2%	39.3%	44.2%	42.0%	40.7%	42.9%
Insulation quality: Poor. Energy label E, F, or G.	0.049	0.868		18.0%	17.6%	17.3%	15.5%	17.2%	19.2%	16.2%
Insulation quality: Don't know / Prefer not to say	0.049	0.868		6.5%	3.9%	3.9%	4.6%	4.1%	3.7%	5.6%
Housing ownership: Other	0.052	0.221		1.0%	1.1%	1.4%	2.5%	0.4%	1.2%	1.0%
Housing ownership: I live in social housing owned by a housing corporation	0.052	0.221		29.5%	30.4%	34.0%	30.3%	32.9%	32.2%	33.3%
Housing ownership: I rent a home, not from a housing corporation	0.052	0.221		8.4%	10.7%	8.4%	10.3%	8.5%	7.9%	7.3%
Housing ownership: I own my home	0.052	0.221		60.7%	57.8%	55.4%	56.7%	58.2%	58.8%	57.3%
Housing ownership: Don't know / Prefer not to say	0.052	0.221		0.5%		0.7%	0.2%			1.0%

Table F.2
Balance table across trust classes.

Variable	SMD (max)	p-value	Stars	1	2	3	4	5	6
N				428	1019	86	895	465	103
Age	1.034	0	***	53.28 (15.11)	52.96 (16.48)	46.94 (19.01)	47.75 (17.80)	42.39 (17.74)	39.15 (12.06)
Gender: I am a man	0.098	0.001	**	50.9%	50.2%	48.8%	48.3%	47.5%	57.3%
Gender: I am a woman	0.098	0.001	**	48.6%	49.5%	50.0%	51.7%	52.3%	42.7%
Gender: I am non-binary	0.098	0.001	**	0.5%	0.2%			0.2%	
Gender: Prefer not to say	0.098	0.001	**		0.1%				
Gender: Other	0.098	0.001	**			1.2%			
Education: Primary school	0.132	0	***	2.8%	0.7%	1.2%	1.7%	0.9%	1.0%
Education: High school (VWO/HAVO), Years 1–3	0.132	0	***	4.2%	3.4%	4.7%	1.8%	1.5%	
Education: High school (VWO/HAVO), Years 4–6	0.132	0	***	7.5%	8.7%	5.8%	7.8%	8.8%	3.9%
Education: HBO	0.132	0	***	27.1%	30.4%	33.7%	25.3%	25.2%	34.0%
Education: MBO Level 1	0.132	0	***	1.6%	2.4%		3.0%	1.3%	
Education: MBO Levels 2–4 (basic vocational, vocational, or middle-management and specialist programs)	0.132	0	***	29.0%	29.1%	32.6%	33.4%	33.1%	32.0%
Education: University	0.132	0	***	11.2%	11.2%	8.1%	10.9%	17.4%	21.4%
Education: VMBO	0.132	0	***	16.4%	13.8%	12.8%	16.0%	11.6%	7.8%
Education: Don't know / Prefer not to say	0.132	0	***	0.2%	0.2%	1.2%	0.1%	0.2%	
Getting by: Easy	0.167	0	***	28.7%	35.0%	41.9%	37.4%	40.6%	39.8%
Getting by: Difficult	0.167	0	***	15.4%	12.4%	14.0%	10.1%	9.2%	9.7%
Getting by: Neither easy nor difficult	0.167	0	***	33.4%	39.8%	34.9%	40.0%	35.7%	26.2%
Getting by: Very easy	0.167	0	***	12.4%	9.4%	4.7%	9.2%	12.7%	21.4%
Getting by: Very difficult	0.167	0	***	9.3%	2.8%	3.5%	2.5%	1.1%	1.0%
Getting by: Prefer not to say / Don't know	0.167	0	***	0.7%	0.5%	1.2%	0.9%	0.6%	1.9%
Monthly expenses: Yes, more than €300 per month	0.12	0.022	*	5.4%	4.5%	3.5%	4.2%	6.5%	10.7%
Monthly expenses: Yes, less than €100 per month	0.12	0.022	*	17.1%	16.9%	16.3%	14.3%	10.8%	7.8%
Monthly expenses: Yes, between €100 and €150 per month	0.12	0.022	*	20.8%	21.4%	25.6%	21.6%	21.1%	21.4%
Monthly expenses: Yes, between €150 and €200 per month	0.12	0.022	*	22.7%	23.1%	18.6%	23.1%	21.1%	22.3%
Monthly expenses: Yes, between €200 and €250 per month	0.12	0.022	*	14.0%	15.7%	12.8%	15.0%	19.1%	20.4%
Monthly expenses: Yes, between €250 and €300 per month	0.12	0.022	*	11.2%	11.3%	8.1%	11.6%	12.3%	13.6%
Monthly expenses: No, don't know	0.12	0.022	*	5.6%	4.8%	14.0%	7.6%	7.1%	1.9%
Monthly expenses: Prefer not to say	0.12	0.022	*	3.3%	2.4%	1.2%	2.6%	2.2%	1.9%
Affordability: Easy to afford	0.175	0	***	25.5%	26.7%	30.2%	32.7%	37.0%	39.8%
Affordability: Difficult to afford	0.175	0	***	20.1%	16.2%	16.3%	13.4%	11.0%	9.7%
Affordability: Neither easy nor difficult to afford	0.175	0	***	31.3%	42.8%	40.7%	39.8%	38.7%	25.2%
Affordability: Don't know / Prefer not to say	0.175	0	***	1.9%	1.6%	5.8%	3.5%	3.0%	1.9%
Affordability: Very easy to afford	0.175	0	***	10.7%	9.1%	5.8%	7.7%	8.8%	21.4%
Affordability: Very difficult to afford	0.175	0	***	10.5%	3.6%	1.2%	2.9%	1.5%	1.9%
Living situation: Something else	0.198	0	***	3.0%	3.3%	10.5%	6.6%	5.8%	4.9%
Living situation: I live alone	0.198	0	***	30.4%	25.4%	26.7%	25.7%	25.2%	26.2%
Living situation: I live with a partner and child(ren)	0.198	0	***	25.2%	24.7%	27.9%	24.9%	29.9%	42.7%

(*continued on next page*)

Table F.2 (*continued*)

Variable	SMD (max)	p-value	Stars	1	2	3	4	5	6
Living situation: I live with a partner, without children	0.198	0	***	33.9%	39.3%	27.9%	34.9%	30.8%	19.4%
Living situation: I live with housemates, for example in student housing	0.198	0	***	1.2%	1.9%	3.5%	2.7%	2.4%	1.9%
Living situation: I live with child(ren), without a partner	0.198	0	***	6.3%	5.4%	3.5%	4.9%	6.0%	3.9%
Living situation: Prefer not to say	0.198	0	***				0.3%		1.0%
Occupation: Other	0.342	0	***	12.1%	7.5%	5.8%	5.0%	3.4%	2.9%
Occupation: I am a homemaker	0.342	0	***	7.0%	5.7%	1.2%	5.7%	5.8%	3.9%
Occupation: I am retired	0.342	0	***	25.7%	26.9%	22.1%	20.7%	13.8%	4.9%
Occupation: I am in paid employment	0.342	0	***	50.2%	53.7%	64.0%	59.4%	67.7%	84.5%
Occupation: I do volunteer work	0.342	0	***	2.8%	2.5%		3.5%	2.2%	1.0%
Occupation: I go to school / I study	0.342	0	***	1.9%	3.3%	7.0%	5.1%	6.7%	2.9%
Occupation: Don't know / Prefer not to say	0.342	0	***	0.2%	0.5%		0.6%	0.4%	
Climate worry: Yes, a little	0.241	0	***	29.4%	45.7%	51.2%	53.3%	53.5%	49.5%
Climate worry: Yes, very much	0.241	0	***	21.0%	22.4%	18.6%	17.5%	24.3%	35.9%
Climate worry: No, not at all	0.241	0	***	27.1%	8.6%	5.8%	6.1%	4.3%	3.9%
Climate worry: No, not really	0.241	0	***	22.0%	23.1%	24.4%	21.8%	17.6%	10.7%
Climate worry: Don't know / Prefer not to say	0.241	0	***	0.5%	0.2%		1.2%	0.2%	
Region: Drenthe	0.21	0.001	***	4.7%	4.1%	2.3%	1.8%	3.0%	4.9%
Region: Flevoland	0.21	0.001	***	3.0%	2.7%	2.3%	3.5%	3.7%	
Region: Friesland	0.21	0.001	***	2.6%	5.0%	4.7%	4.7%	3.9%	1.9%
Region: Gelderland	0.21	0.001	***	8.6%	11.6%	8.1%	9.6%	9.9%	5.8%
Region: Groningen	0.21	0.001	***	4.4%	4.3%	2.3%	3.9%	3.7%	5.8%
Region: Limburg	0.21	0.001	***	7.7%	9.6%	7.0%	7.6%	6.0%	6.8%
Region: North Brabant	0.21	0.001	***	14.3%	13.8%	19.8%	14.9%	12.3%	11.7%
Region: North Holland	0.21	0.001	***	17.5%	13.9%	19.8%	15.4%	17.2%	35.0%
Region: Overijssel	0.21	0.001	***	5.1%	6.1%	7.0%	7.3%	7.7%	6.8%
Region: Utrecht	0.21	0.001	***	4.7%	7.4%	3.5%	7.6%	5.4%	1.9%
Region: Zeeland	0.21	0.001	***	1.9%	3.0%	2.3%	3.1%	2.6%	1.9%
Region: South Holland	0.21	0.001	***	25.5%	18.4%	20.9%	20.7%	24.7%	17.5%
Housing type: Other	0.17	0.037	*	5.1%	4.3%	2.3%	6.3%	4.3%	3.9%
Housing type: In an apartment, flat, upstairs apartment, or downstairs apartment	0.17	0.037	*	32.9%	32.3%	30.2%	34.5%	37.6%	44.7%
Housing type: In a detached house, semi-detached house, or single-family home	0.17	0.037	*	61.9%	63.2%	67.4%	58.5%	58.1%	50.5%
Housing type: Don't know / Prefer not to say	0.17	0.037	*		0.2%		0.7%		1.0%
Heating method: Other	0.259	0	***	9.6%	4.5%	5.8%	4.4%	3.7%	
Heating method: With a central heating boiler and a hybrid heat pump	0.259	0	***	6.5%	6.3%	5.8%	6.6%	9.9%	13.6%
Heating method: With a central heating boiler. The boiler is located in your own home or elsewhere in the building.	0.259	0	***	68.2%	71.9%	74.4%	70.9%	63.7%	48.5%
Heating method: With a fully electric heat pump	0.259	0	***	6.5%	7.0%	3.5%	7.5%	11.0%	28.2%
Heating method: Via a heat grid, such as district heating	0.259	0	***	7.9%	8.8%	7.0%	8.0%	8.6%	5.8%
Heating method: Don't know / Prefer not to say	0.259	0	***	1.2%	1.5%	3.5%	2.6%	3.2%	3.9%
Insulation quality: Good. Energy label A or B.	0.158	0.002	**	36.4%	37.3%	33.7%	34.4%	35.3%	49.5%
Insulation quality: Reasonable. Energy label C or D.	0.158	0.002	**	36.7%	40.2%	45.3%	43.5%	47.1%	37.9%
Insulation quality: Poor. Energy label E, F, or G.	0.158	0.002	**	22.2%	18.8%	15.1%	16.4%	13.3%	8.7%
Insulation quality: Don't know / Prefer not to say	0.158	0.002	**	4.7%	3.6%	5.8%	5.7%	4.3%	3.9%
Housing ownership: Other	0.114	0.005	**	1.9%	1.3%		1.2%	0.9%	1.0%
Housing ownership: I live in social housing owned by a housing corporation	0.114	0.005	**	36.7%	32.1%	36.0%	30.7%	29.2%	25.2%
Housing ownership: I rent a home, not from a housing corporation	0.114	0.005	**	8.4%	6.5%	5.8%	10.5%	10.5%	13.6%
Housing ownership: I own my home	0.114	0.005	**	53.0%	60.1%	57.0%	57.1%	58.9%	58.3%
Housing ownership: Don't know / Prefer not to say	0.114	0.005	**		0.1%	1.2%	0.4%	0.4%	1.9%

Data availability

Data will be made available on request.

References

Allcott, H., Greenstone, M., 2017. Measuring the Welfare Effects of Residential Energy Efficiency Programs. NBER Working Paper. https://doi.org/10.3386/w23386.

Angrist, J.D., Pischke, J.-S., 2009. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press.

Athey, S., Imbens, G.W., 2016. Recursive partitioning for heterogeneous causal effects. Proc. Natl. Acad. Sci. 113 (27), 7353–7360. https://doi.org/10.1073/pnas.1510489113.

Bernardo, F., Palma-Oliveira, J., 2013. Place identity, place attachment and the scale of place: the impact of place salience. Psyecology 4 (2), 167–193. https://doi.org/10.1080/21711976.2013.10773867.

Bettinger, E., Cunha, N., Lichand, G., Madeira, R., 2020. Are the effects of informational interventions driven by salience? SSRN Electron. J. https://doi.org/10.2139/ssrn.3633821.

Bolderdijk, J.W., Steg, L., Geller, E.S., Lehman, P.K., Postmes, T., 2012. Comparing the effectiveness of monetary versus moral motives in environmental campaigning. Nat. Clim. Chang. 3 (4), 413–416. https://doi.org/10.1038/nclimate1767.

Bolderdijk, J.W., Gorsira, M., Keizer, K., Steg, L., 2013. Values determine the (in) effectiveness of informational interventions in promoting pro-environmental behavior. PNAS 110 (2), 468–473.

Botzen, W.J.W., Thépaut, L.D., Banerjee, S., 2025. Kahneman's insights for climate risks: lessons from bounded rationality, heuristics and biases. Environ. Resour. Econ. 88, 2663–2688. https://doi.org/10.1007/s10640-025-00980-4 link.springer.com.

de Vries, P., Pruyn, A., 2007. Source salience and the persuasiveness of peer recommendations: the mediating role of social trust. Lect. Notes Comput. Sci 164–175. https://doi.org/10.1007/978-3-540-77006-0_22.

de Vries, M.O., Mouter, N., Metze, T.A.P., Spruit, S., 2025. Who is listening? Comparing 7 cases of citizen participation at different levels of government in the Dutch energy transition. Energy Policy 209, 114998. https://doi.org/10.1016/j.enpol.2025.114998.

Druckman, J.N., McGrath, M.C., 2019. The evidence for motivated reasoning in climate change preference formation. Nat. Clim. Chang. 9 (2), 111–119. https://doi.org/10.1038/s41558-018-0360-1.

Earle, T., Siegrist, M., 2008. Trust, confidence and cooperation model: a framework for understanding the relation between trust and risk perception. Int. J. Global Environ. Issues 8 (1/2), 17. https://doi.org/10.1504/ijgenvi.2008.017257.

Gillingham, K., Tsvetanov, T., 2018. Nudging energy efficiency audits: evidence from a field experiment. J. Environ. Econ. Manag. 90 (303), 316. https://doi.org/10.1016/j.jeem.2018.06.009.

Grimmelikhuijsen, S., 2012. Linking transparency, knowledge and citizen trust in government: an experiment. Int. Rev. Adm. Sci. 78 (1), 50–73. https://doi.org/10.1177/0020852311429667.

Grimmelikhuijsen, S., Knies, E., 2017. Validating a scale for citizen trust in government organizations. Int. Rev. Adm. Sci. 83 (3), 583 601. https://doi.org/10.1177/0020852315585950.

Grimmelikhuijsen, S., Meijer, A., 2022. Does transparency undermine trust? A critical review of the paradox of transparency. Public Adm. Rev. 82 (1), 116–128. https://doi.org/10.1111/puar.13409.

Hart, P.S., Nisbet, E.C., 2012. Boomerang effects in science communication. Commun. Res. 39 (6), 701–723.

Holladay, J.S., LaRiviere, J., Novgorodsky, D., Price, M., 2016. Asymmetric Effects of Non-Pecuniary Signals on Search and Purchase Behavior for Energy-Efficient Durable Goods. NBER Working Paper Series. https://doi.org/10.3386/w22939.

Lefevere, J., Sevenans, J., Walgrave, S., Lesschaeve, C., 2017. Issue reframing by parties. Party Polit. https://doi.org/10.1177/1354068817736755, 135406881773675.

Mayer, R.C., Davis, J.H., Schoorman, F.D., 1995. An integrative model of organizational trust. Acad. Manag. Rev. 20 (3), 709–734. https://doi.org/10.5465/amr.1995.9508080335.

McKnight, D.H., Choudhury, V., Kacmar, C., 2002. Developing and validating trust measures for e-commerce: an integrative typology. Inf. Syst. Res. 13 (3), 334–359.

Metzger, M.J., Flanagin, A.J., 2013. Credibility and trust of information in online environments: the use of cognitive heuristics. J. Pragmat. 59 (210), 220. https://doi.org/10.1016/j.pragma.2013.07.012.

Mouter, N., Hernandez, J.I., Itten, A.V., 2021. Public participation in crisis policymaking. How 30,000 Dutch citizens advised their government on relaxing COVID-19 lockdown measures. PLoS One 16 (5), e0250614. https://doi.org/10.1371/journal.pone.0250614.

Olden, A., Møen, J., 2022. The triple difference estimator. Econ. J. 25 (3), 531–553.

Otto, D., Chilvers, J., Trdlicova, K., 2023. A synthetic review of the trust-participation nexus: towards a relational concept of trust in energy system transformations to net zero. Energy Res. Soc. Sci. 101, 103140. https://doi.org/10.1016/j.erss.2023.103140.

Rains, S.A., 2013. The nature of psychological reactance revisited: a meta-analytic review. Hum. Commun. Res. 39 (1), 47–73.

Renn, O., Levine, D., 1991. Credibility and trust in risk communication. In: Kasperson, R. E., Stallen, P.J.M. (Eds.), Communicating Risks to the Public: International Perspectives. Springer, pp. 175–218. https://doi.org/10.1007/978-94-009-1952-5_10.

Siegrist, M., Connor, M., Keller, C., 2011. Trust, confidence, procedural fairness, outcome fairness, moral conviction, and the acceptance of GM field experiments. Risk Anal. 32 (8), 1394–1403. https://doi.org/10.1111/j.1539-6924.2011.01739.x.

Terpstra, T., 2011. Emotions, trust, and perceived risk: affective and cognitive routes to flood preparedness behavior. Risk Anal. 31 (10), 1658–1675. https://doi.org/10.1111/j.1539 6924.2011.01616.x.

Thaler, R.H., Sunstein, C.R., 2008. Nudge: Improving Decisions about Health, Wealth, and Happiness. Yale University Press.

Thépaut, L.D., Blasch, J., Khadjavi, M., Spadaro, G., 2024. Energy Transition at Home: Nudging the Uptake of Energy Audits by Dutch Homeowners [Unpublished manuscript]. SSRN. https://ssrn.com/abstract=5387432.

Tyler, T.R., 2006. Why People Obey the Law. Princeton University Press.

Wooldridge, J.M., 2014. Introduction to Econometrics (Europe, Middle East & Africa Edition). Cengage Learning EMEA, Andover, United Kingdom.