Can behavioral interventions optimize self-consumption? Evidence from a field experiment with prosumers in Germany

Pelka, Sabine; Kesselring, Anne; Preuß, Sabine; Chappin, Emile; de Vries, Laurens

DOI
10.1016/j.segy.2024.100140

Publication date
2024

Document Version
Final published version

Published in
Smart Energy

Citation (APA)
Pelka, S., Kesselring, A., Preuß, S., Chappin, E., & de Vries, L. (2024). Can behavioral interventions optimize self-consumption? Evidence from a field experiment with prosumers in Germany. *Smart Energy*, *14*, Article 100140. https://doi.org/10.1016/j.segy.2024.100140

Smart Energy 14 (2024) 100140

Contents lists available at ScienceDirect

Smart Energy

journal homepage: www.journals.elsevier.com/smart-energy

Can behavioral interventions optimize self-consumption? Evidence from a field experiment with prosumers in Germany

Sabine Pelka [a,b,*], Anne Kesselring [b], Sabine Preuß [b], Emile Chappin [a], Laurens de Vries [a]

[a] Energy and Industry Group, Faculty of Technology Policy and Management, Delft University of Technology, Jaffalaan 5, 2628, BX, Delft, the Netherlands
[b] Fraunhofer Institute for Systems and Innovation Research ISI, Breslauer Str. 48, 76139, Karlsruhe, Germany

A R T I C L E I N F O

Keywords:
Default
nudge
Demand response
Smart meter data
Smart charging

A B S T R A C T

Aligning prosumers' electricity consumption to the availability of self-generated electricity decreases CO_2 emissions and costs. Nudges are proposed as one behavioral intervention to orchestrate such changes. At the same time, fragmented findings in the literature make it challenging to identify suitable behavioral interventions for specific households and contexts - specifically for optimizing self-consumption. We test three sequentially applied interventions (feedback, benchmark, and default) delivered by digital tools in a field experiment with 111 German households with rooftop-photovoltaics. The experiment design with a control-group, baseline measurements, and high-frequency smart-meter-data allows us to examine the causal effects of each intervention for increasing self-consumption. While feedback and benchmark deliver small self-consumption increases (3–4 percent), the smart changing default leads to a 16 percent increase for active participants. In general, households with controllable electric vehicles show stronger effects than those without. For upscaling behavioral interventions for other prosumers, we recommend interventions that require little interaction and energy literacy because even the self-selected, motivated sample rarely interacted with the digital tools.

1. Introduction

Shifting consumption to the times of self-generated electricity of households with rooftop photovoltaic (PV) is a key measure to decarbonize the residential energy sector. Coordinated consumption shifts ensure a viable return-on-investment for households [1–3] and a more efficient operation of the existing energy infrastructure [4–6]. Optimization models demonstrate that households can increase their self-consumption with consumption shifts by 2–50 percentage points, depending on the optimized technology. White goods are at the lower end [7], while stationary battery systems [1,2,6,8–10] and EVs [11,12] are more promising. To unlock the emerging flexibility potential of the latter, households need to establish a new routine for using these flexible technologies [12].

Orchestrating consumption shifts is an understudied use case for behavioral interventions [6,13,14]. Behavioral interventions, as subtle changes in one's choice environment, complement price incentives. Price incentives shape the terms of household consumption and address rational reasoning (e.g., higher return-on-investment from tax exemption [15–17]). Behavioral interventions provide ongoing support for households to respond to these terms (e.g., stimulating flexibility) and to make intuitive decisions [20]. Nudges are one of the most researched behavioral interventions [18]. The ongoing rise of digital tools leads to a broader application since behavioral interventions can be easily implemented in the user interface [19,20].

Behavioral interventions can guide households in the way the choice task is structured, and the choice option is described [20,21]. The first category about structuring the choice task is known as more effective but also invasive in terms of paternalism. A frequently applied example is defaults [20,22]. In contrast, the second category about describing the choice options (e.g., feedback) is more subtle [20]. The majority of interventions for energy savings belong to this category. For instance, the realized energy savings for feedback ranged between 5 and 13 percent (e.g., Refs. [11,23–29]). In some studies (e.g., Refs. [28,29]), the effect persisted over a period of up to two years. However, most field trials took four weeks to 11 months and did not report long-term effects.

Although automated consumption shifts enable behavioral interventions from the first category (i.e., structuring the choice task), the few existing studies on consumption shifts apply behavioral interventions from the second category about the description of choice options (e.g., environmentally friendly framing). These studies show

* Corresponding author. Energy and Industry Group, Faculty of Technology Policy and Management, Delft University of Technology, Jaffalaan 5, 2628, BX, Delft, the Netherlands.

E-mail address: s.pelka@tudelft.nl (S. Pelka).

https://doi.org/10.1016/j.segy.2024.100140
Received 28 July 2023; Received in revised form 8 March 2024; Accepted 25 March 2024
Available online 29 March 2024

Abbreviations:

DiD -	difference-in-differences
EV –	electric vehicle
FE –	fixed effects
PV –	photovoltaic
TWFE -	two-way fixed-effects

one-digit improvements of provided flexibility [13,15,16]. Although this seems small, the effects are economically meaningful given the strong evidence that price incentives alone are insufficient for energy decision-making (e.g., Refs. [17,19]). The similar magnitude, the existing literature's focus on other incentive mechanisms, and the missing utilization of automation encourage us to explore further behavioral intervention of both categories for households with rooftop-PV and EVs. Thereby, we consider that these prosumers have a different asset base and, therefore, more flexibility potential in the operational phase than the general population. We contribute to the broader research question "Can behavioral interventions delivered through digital tools help prosumers increase their self-consumption?" by testing empirically the impact (i) of interventions from both categories (i.e., changing the description of choice options and the structure of the choice task) and (ii) for prosumers with and without EVs.

The range of findings in the literature makes it challenging to determine which kind of behavioral intervention fits which household and context. The efficacy of such interventions is highly context-specific, as the intervention accounts only for part of the outcome variation (self-consumption in our case) in real-life environments. Insights on the group- and context-dependent fit are therefore important but largely missing [24], while publication bias reinforces the evidence gap [25]. At the same time, methodological challenges exist: First, generally established techniques for stated preferences are less suitable for capturing intuitive choices and intervention effects of everyday life [25]. Second, shortcomings in the research design of revealed preference approaches (e.g., underpowered sample, no control-group, no baseline measurement) impede applying methods for causal effects [25,30]. Studies with larger, more heterogeneous samples tend to result in smaller effect sizes [25]. Third, behavioral intervention studies are highly context-specific, hindering interventions' comparability across single interventions [31].

Under consideration of the content-related and methodological challenges, we examine the understudied use case of behavioral interventions for consumption shifts based on smart-meter-data. In a German field experiment of the Horizon 2020 funded project NUDGE, three sequentially applied interventions support 111 participating households in shifting their electricity consumption to times of their self-generated electricity. The first two interventions adjust the description of the choice option (i.e., second category), specifically through visualization in the digital tool in the form of (a) feedback and (b) benchmarking. The third intervention is a default targeting EV charging (i.e., first category). Recent intervention studies based on smart-meter-data (e.g., Refs. [23,24,32–35]) successfully applied a difference-in-differences approach (DiD) to reveal the causal effect of interventions. We also selected this approach and compared the relative developments in self-consumption between the treatment- and control-groups over time.

We create a new comparability level by testing three interventions within the same experiment setting to minimize context-specific differences and present new evidence specific to EV-users. Learning effects during the interventions are managed by establishing previously tested interventions as new basic settings and calculating only the incremental change for each intervention. We investigate group-specific effects for prosumers with and without controllable EVs and fatigue effects during the nudging period.

In Section 2, we present the applied methodology. Section 3 contains the results with the overall, time- and group-dependent effect for each intervention. Section 4 includes the discussion, whereas we conclude our study in Section 5.

2. Methodology

2.1. Sample

We analyzed the self-consumption of 111 participating households living in, or near Mannheim, Germany. The participants are customers of the service provider Beegy[1] and responded voluntarily to its call for participation via e-mail. Smart-meter-data were collected continuously at high-frequency resolution and aggregated to daily average values at household-level for analysis. Our estimation sample starts in January 2022 and ends in June 2023. Supplementary data on household equipment and a socio-demographic survey were also recorded (see supplementary-material 1.2).

The majority are families with children (57 percent) living in a single- or semi-detached house (69 percent) [36]. The average age is 56.34 [36]. All households have rooftop-PV with an average installed PV-capacity of 8.16 kWp [36]. A sub-group of 39 participants owns a controllable EV (in the following called "EV-group"). 105 participants are equipped with battery-storage-systems and 29 with heatpumps. We divided the sample into a treatment-group ($n = 54$) and a control-group ($n = 57$) with random assignment before the first intervention. Both groups are similar in installed PV-capacity, number of controllable EVs ($n = 18$ in the treatment- and $n = 21$ in the control-group), wall boxes, heatpumps, and other technical dimensions (see TableAnnex 1). Equipment changes during the intervention period (see supplementary-material 1.2) were considered in a robustness check.

2.2. Design and Procedure

2.2.1. Interventions and experiment design

In the following, we describe the interventions and their implementation. Each intervention was presented to participants for a specific period during the experiment (see Fig. 1). Two tools, a webportal, and a smart-charging-app, exposed the participants to the interventions. The smart-charging-app is only available for participants with controllable EVs. The tools were already in use before the experiment. This real-life embedding creates authentic insights but also places restrictions on the intervention design (e.g., no social comparison is possible due to data privacy).

Flexible technologies such as battery-storage-systems ($n = 105$) and heatpumps ($n = 29$) were automatically optimized for increasing self-consumption (see Section 4).

Given concerns about fatigue effects for participants, we first implement two interventions that change the description of the choice options (but leave the choice task as before, i.e., second intervention category) and end with an intervention that simplifies the choice task (i.e., first intervention category). The two earlier interventions are visualizations that bundle more than one design element. They present information on how behavior translates to savings in both monetary terms and CO_2 emissions. They are designed to require more active user engagement than the last intervention.

The two earlier interventions describe the choice options more appealingly based on concurrent timing (first intervention, feedback) or more competitively with a dynamic framing (second intervention, historical benchmark). The feedback combines simple indicators on a dashboard with signaling colors (see Fig. 2). The historical benchmark with prompts reports on the previous and upcoming self-consumption in

[1] Further information on the service portfolio of Beegy can be found: https://www.beegy.com/one-pager-en/(last visited: 27/12/2023).

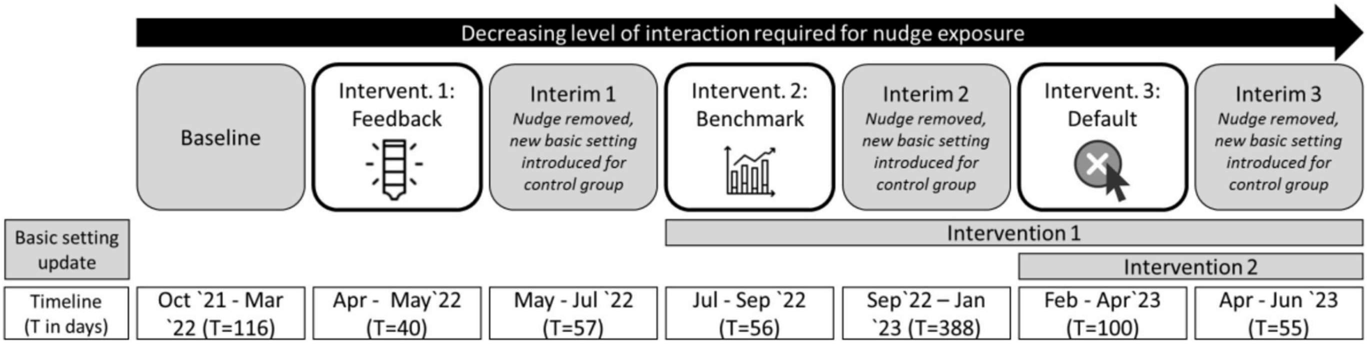

Fig. 1. Timeline of the experiment.

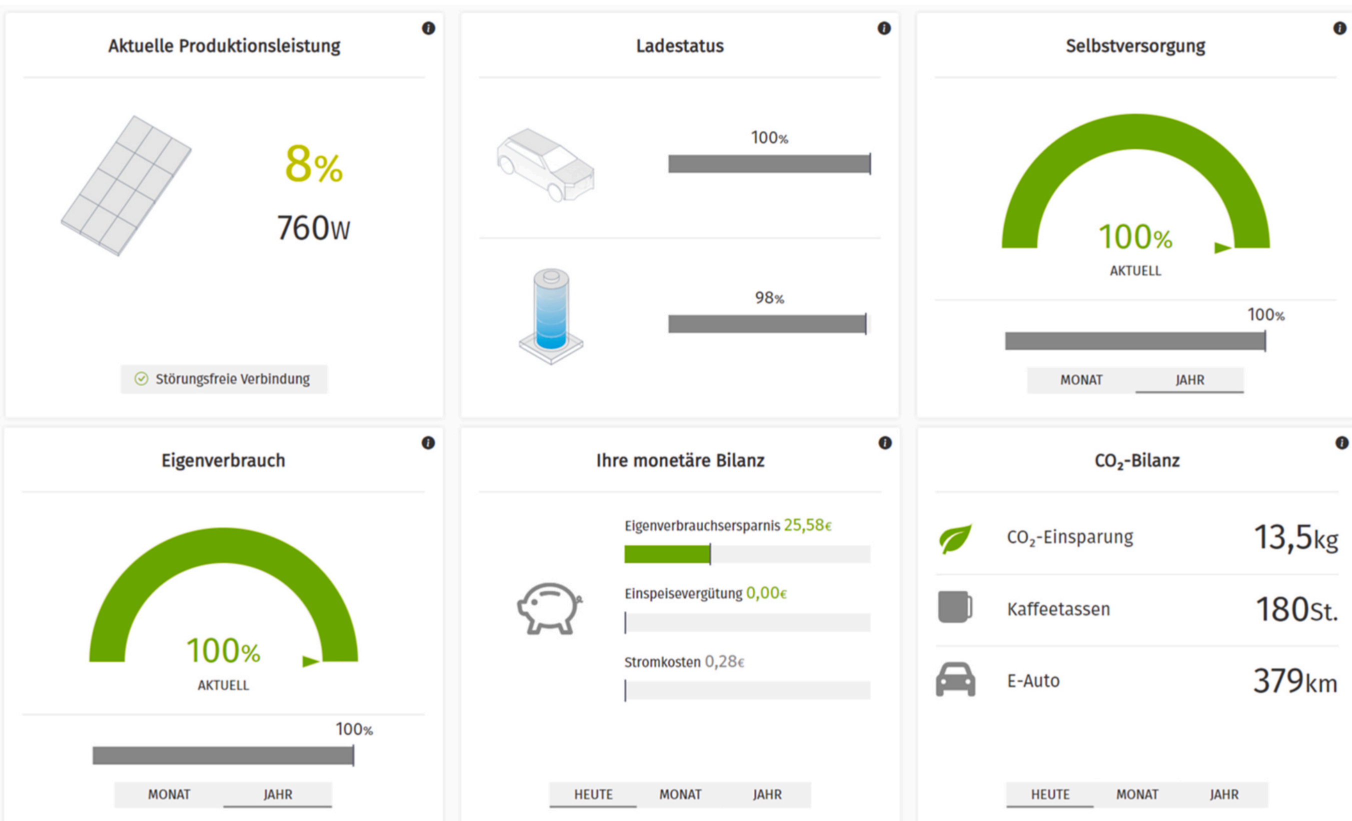

Fig. 2. Intervention 1 for the PV- and EV-group providing simple indicators in signaling color to stimulate consumption shifts or additional consumption during PV-generation by the participants, as presented to the participants and thus, in German language, see supplementary material for further information.

a bar chart and provides recommendations on how to adjust the consumption (see Fig. 3). The consumption recommendations are based on a forecast of self-generated electricity and are communicated with prompts. They encourage households to use their dishwasher, laundry, or washing machine during the hours of forecasted generation.

The third chosen intervention, a default intervention, changes the choice tasks and aims to establish new charging behaviors with low awareness and interaction requirements (see Fig. 4).[2] Therefore, a new charging mode for participants with controllable EVs was introduced. The existing charging mode of the smart-charging-app maximized self-consumption during charging, given the specified target state of charge and departure time. The new charging mode is activated on the webportal and charges the EV only with self-generated electricity.[3] Once the participants accepted the new charging mode in the webportal, it was always activated when the EV was plugged in at home.

Simultaneously with the smart-charging-default, an additional feature as part of the third intervention was introduced for all participants to keep participants without controllable EVs engaged. The feature aggregates the savings in terms of cost savings and CO_2 emissions in the form of a downloadable energy report (see FigureAnnex 3).

The first intervention is implemented on the dashboard, which is the landing page of the tool (i.e., the page that is shown once the tool is opened). The second and third interventions are implemented on pages that are accessed via the sidebar of the tools (i.e., "statistic" and "forecast" for the second and "download" and "e-mobility" for the third

[2] Figs. 2–4 illustrates the three interventions in the webportal, which are similarly implemented in the smart-charging-app (see Appendix 8.1).

[3] Provided it is not overruled by new settings in the smart-charging-app.

Fig. 3. Intervention 2 for the PV- and EV-group providing benchmark of previous and current self-consumption (top) and forecast of PV-generation with recommendations for actions (bottom) to stimulate consumption shifts or additional consumption during PV-generation by the participants, as presented to the participants and thus, in German language, see supplementary material for further information.

intervention). The participants were informed about the updates via a one-time e-mail at the beginning of the intervention and via a “new" sticker next to the page name on the sidebar.

Two special constellations in the experiment design allow us to evaluate the respective effect of each intervention. First, multi-treatment designs have to consider learning effects, which makes it difficult to separate individual interventions from the compound effect. To mitigate this issue, we introduced interim periods without an intervention after each treatment period. Second, to distinguish between persistent learning and the effect of the following intervention, we transformed the previous intervention into a basic setting for the following intervention. This means that the previous intervention was visible to the control- and treatment-groups when the following intervention was introduced. To allow the control-group to internalize the new basic setting, we have already introduced the previous intervention to them during the interim period. The difference between the treatment- and control-group provides incremental change since the control-group has only seen the previous intervention before the next treatment begins.

2.2.2. Main measures for treatment and control group

We tested the three interventions sequentially in the same setting. To respond to the interventions, participants can either shift their existing consumption or additionally consume self-generated electricity. We computed two measures to analyze participants’ responses to the nudging interventions: an absolute one (self-consumption) and a relative one to the overall consumption (autarky-rate) [35,36]. While in other studies (e.g., Refs. [4,7]) the latter is also called self-sufficiency-rate, we call it autarky-rate to avoid terminological confusion with sufficiency-research.

The absolute measure recognizes both responses but is prone to random consumption changes (e.g., vacations, construction works). The relative one absorbs these consumption changes (including additional consumption in response to the intervention).

As outlined above, the composition of the treatment- and control-group is comparable; this also applies to the mean outcome variables

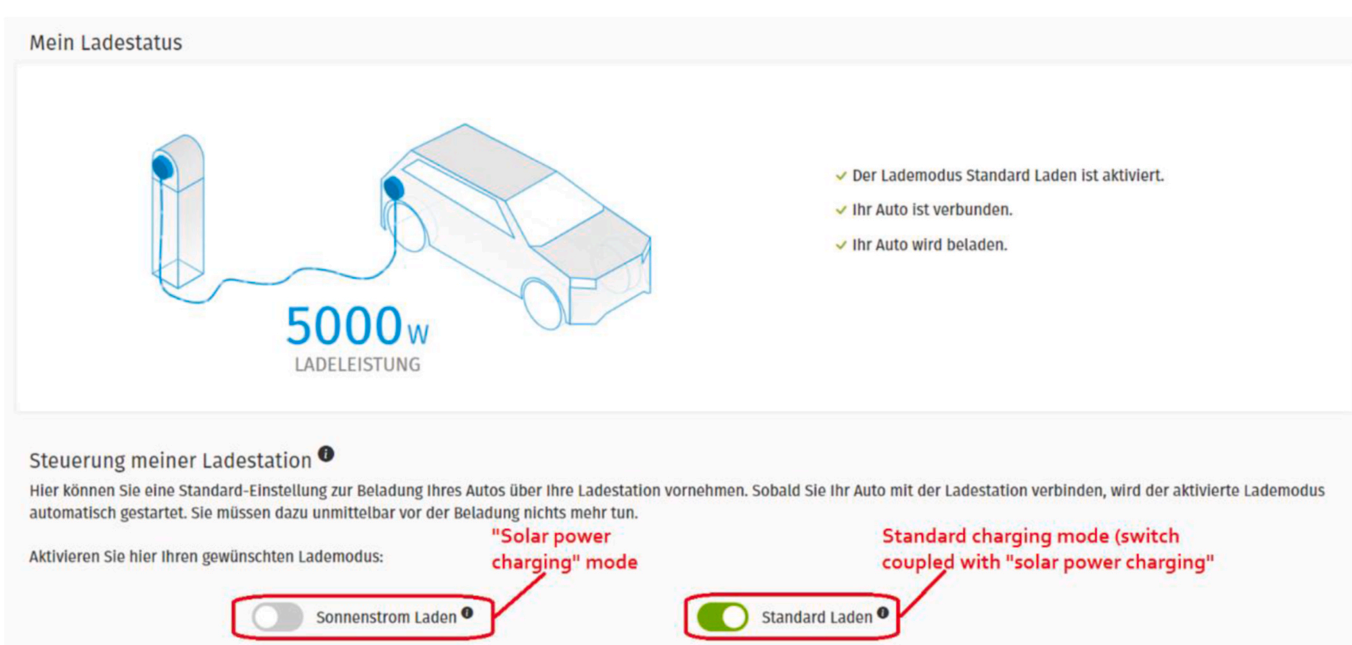

Fig. 4. Intervention 3 for the EV-group providing a new charging mode that charges the EV automatically with excess electricity from the local PV (“solar power charging" – switch on the left side, which is deactivated until its first activation). If activated, the switch on the right side for the existing charging mode “standard charging" is deactivated.

Table 1
Summary statistics by group.

	Mean	SD	Min	Max	Obs
Treatment-group (n = 54)					
Consumption [Wh]	755.58	586.12	0.05	7503.91	23029
Self-consumption [Wh]	445.49	358.8	0	3863.78	23029
Autarky-rate [percentage]	0.55	0.24	0	1	23029
Control-group (n = 57)					
Consumption [Wh]	720.18	565.57	0	5987.11	24010
Self-consumption [Wh]	459.33	370.57	0	4188.53	24010
Autarky-rate [percentage]	0.60	0.23	0	1	24010

Notes: Descriptive statistics for estimation sample from January 2022 to June 2023 at daily aggregation. Self-consumption is the difference between total consumption and output to grid. Autarky-rate is the ratio of self-consumption to total consumption.

(see Table 1). Self-consumption is calculated as the mean hourly value over a 24-h period. Autarky-rate is the ratio of self-consumption to total consumption, calculated from the respective daily means. The autarky-rate takes values between 0 and 1. The final column shows the number of observations (Obs) in the panel comprising 422 days, after excluding few cases with missing values in the smart-meter reporting. The standard deviation, minimum and maximum indicate a high variation within each group across individual households. Overall, the participants' energy consumption is above the German average but falls in line with estimates addressing prosumers and EV ownership (e.g., Refs. [8, 12]).

Fig. 2 plots the data for both groups over time to complement the static representation. The three solid, vertical black lines indicate the treatment start dates of the interventions, with the dashed line indicating the end of the intervention for the treatment-group. This partitions the study period into four blocks of interest: the baseline (N = 0) and the interventions N = {1,2,3}. For the DiD approach, it is important that treatment- and control-groups are comparable and do not exhibit differential patterns at baseline (parallel trends assumption, see, e.g., Angrist and Krueger (1999) [30]). The graphical illustration supports this assumption. Both groups are similar in levels and trends, and the strong, common fluctuation over time is driven by weather conditions, as expected. The variability in August 2022 is attributable to missing values due to problems with a central data platform. We conducted a robustness check with a restricted sample to ensure that this does not bias the estimation results. The same pattern holds qualitatively for both outcomes (self-consumption and autarky-rate), despite higher volatility for self-consumption.

The parallel development between the groups also holds when comparing the particular EV-group and the group without controllable EVs (in the following called "PV-group"), providing further support for successful randomization in the design. For details, refer to the supplementary-material 1.3. The figure does not support a visible divergence during the intervention, which indicates that average effects may be small and weather effects may dominate patterns over time. We will consider these first insights in the formal analysis.

2.3. Statistical models for analysis

As outlined above, the main identification strategy is a DiD approach. The objective is to identify the causal effect of the behavioral intervention after accounting for differences across groups and differences across time that would otherwise correlate with the nudging effect. We evaluate the effect of the intervention assignment (i.e., intention-to-treat), which may include participants that did not actively interact with the content. However, this approach indeed provides a realistic projection of the expected effect of intervention in real-world settings for policy-makers and practitioners.

The model is estimated for the two outcome variables. Autarky-rate is the preferred outcome in light of the wide variation across individuals depicted in Table 1. For self-consumption, we log-transform the dependent variable to address the long right tail with high-value outliers in the distribution of the raw data.

There are two challenges to obtaining credible estimates in our setting. First, the European energy crisis: We address this with time-fixed effects at the daily level absorbing shocks in the environment that are common to both groups. This includes behavioral adjustments driven by price spikes and political announcements. For example, Pelka et al. (2022) [31] document that search volume on Google Trends accounts for part of the variation in self-consumption. Time-fixed effects also account for weather variation, which applies to both groups and shows in the raw data (see Fig. 5). Second, we want to compare the three nudging treatments with each other. We, therefore, estimate separate coefficients for each intervention instead of a single treatment effect.

With these considerations, we chose a two-way fixed-effects model (TWFE) with multiple treatment periods (see e.g., Ref. [37]). Formally, the regression equation (1) is:

$$y_{it} = a_i + b_N\, T_{it}\, N_t + c\, G_i + d\, N_t + p_t + e_{it} \tag{1}$$

Where i indicates individuals and t indicates time periods (days). The indicator T equals 1 for the treatment-group, and zero for the control-group. N is a categorical variable that takes value 0 at baseline and then has six non-zero values. The three active intervention periods $N = 1, N = 2, N = 3$, and the interim periods (see Fig. 5). The coefficient of interest is b_N for all $N = \{1, 2, 3\}$, which captures the DiD treatment effect from the interaction of T and N. The estimate represents the differential development of the treated households during the nudging period measured relative to the control-group.

The TWFE model absorbs individual-specific intercepts (a_i) and period-specific intercepts (p_t, see discussion above). The individual fixed-effects a_i absorb level differences across households in a within-transformation. This accounts for time-constant factors such as household size, stock of appliances, or pre-existing behavioral differences. Robust standard errors are calculated with the common Huber-White adjustment. From a purely statistical perspective, the model obtains coefficients also for the interim periods: N = {0,1, …6} (see supplementary-material 1.3). The interim coefficients $b_{N>3}$ capture the relative difference across groups, not the counterfactual development without any nudging.

Overall, we chose the methodology in light of the data structure and the objective to deliver causal effects. The DiD is state of the art [38], and allows us to leverage the experiment design with control group and panel data. The addition of two-way fixed effects provides further control over the granularity in the time-series and cross-sectional variation [39–41]. Relative to simpler regression designs, we lose degrees of freedom, but gain the ability to address the complex variation pattern.

We then add heterogeneity analysis and robustness checks. First, we consider the dynamic nature of the treatment effect by running an event study for all three interventions. This addresses concerns that the treatment effect diminishes over time due to fatigue. By contrast, the behavior might change with a time lag to treatment, so the prediction is ambiguous. Additionally, the pre-treatment coefficients support parallel trends assumption.

The second extension is a sub-group analysis for the EV and PV-group. This is implemented through an additional interaction term in the main regression equation for the sub-groups. The working hypothesis predicts a stronger effect on the EV-group because this sub-group receives the treatment with an additional interface – the smart-charging-app. This analysis is particularly interesting for Intervention 3, since that intervention is a two-part treatment with additional functionalities specific to the EV-group. We then use additional information from the tool to test whether tool users were able to shift self-consumption from the evening hours (in which they tend to charge [42]) to the midday-PV-peak. Formally, this is tested with a regression (2) specific to Intervention 3 using data at *hourly* frequency:

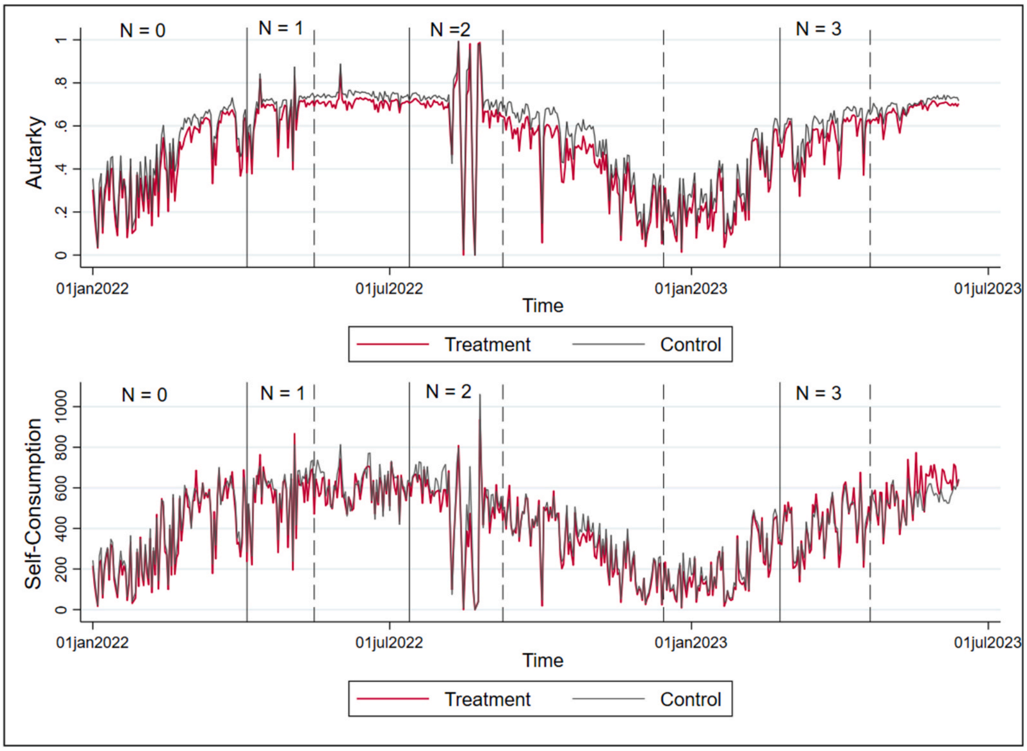

Fig. 5. Outcomes by group over time.

Table 2
Main results for DiD design.

Panel A: Results for Autarky-Rate	(1)	(2)	(3)	(4)
	Basic	Time FE	**Twoway FE**	Weather
N = 1	0.0146**	0.0185***	**0.0209*** **	0.0201***
	(2.09)	(3.45)	**(5.57)**	(4.40)
N = 2	0.0145*	0.0201***	**0.0212*** **	0.0180***
	(1.77)	(3.72)	**(5.05)**	(2.65)
N = 3	−0.00868	−0.00493	**−0.00935** **	−0.000671
	(-1.22)	(-0.87)	**(-2.26)**	(-0.05)
R2	0.255	0.582	**0.778**	0.639
Obs	46409	46409	**46409**	34004
FE	none	time	**time + household**	household

Panel B: Results for Self-Consumption	(1)	(2)	(3)	(4)
	Basic	Time FE	**Twoway FE**	Weather
N = 1	0.0160	0.0291	**0.0291* **	0.0259
	(0.54)	(1.23)	**(1.73)**	(1.25)
N = 2	0.0450	0.0606**	**0.0280**	0.0165
	(1.25)	(2.15)	**(1.32)**	(0.59)
N = 3	0.0569**	0.0704***	**0.111*** **	0.105*
	(2.09)	(3.43)	**(6.64)**	(1.81)
R2	0.111	0.473	**0.702**	0.525
Obs	45928	45928	**45928**	33564
FE	none	time	**time + household**	household

Notes: DiD estimation for dependent variables autarky (upper panel) and self-consumption (lower panel). Robust standard errors (Huber-White) in parentheses. Significance levels: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. N = {1,2,3} refers to interventions 1 and 2, and 3, respectively. Columns differ in the fixed-effects structure, indicated in the bottom row.

$$y_{it} = a_i + b_H A_{it} H_t + c H_t + p_t + e_{it} \tag{2}$$

Where A is an indicator for households that actively engage with the app, and H is a categorical variable for AM (6–10am), midday (11am-3pm), and PM (4pm–8pm). The base level is AM, and we exclude night-time hours. The coefficient of interest is b_H, which indicates whether active app users realize larger shifts during a specific Time Block H. We again use a TWFE model and robust standard errors.

3. Results

3.1. Treatment effects

The analysis based on the DiD approach delivers treatment effects for each intervention. For interpretation, two particular aspects of our design are important (see Section 2). First, the treatment effects are measured relative to the control-group, which has never seen the respective intervention before. Second, the treatment effects capture incremental changes: each coefficient gives the effect of the newly introduced intervention. Table 2 presents the main regression results, with autarky-rate as the dependent variable in the upper panel and the natural logarithm of self-consumption in the lower panel. Throughout all results, we refer to self-consumption meaning this logarithmic transformation for ease of exposition. The first column shows the basic model with no controls or fixed effects. Column (2) adds time-fixed effects to address day-specific shocks common to both groups. Column (3) is a two-way fixed-effects model with both time and household-specific fixed effects. This very conservative estimation is most demanding regarding variation but also most credible in eliminating the potential confounders discussed previously. Note that the group and period indicators are omitted due to the collinearity with the fixed effects. Column (4) replaces the time fixed-effects with the continuous variable solar radiation based on the insights from Fig. 2. The household fixed-effects are kept. The number of observations is lower because radiation data are not available for December 2022 (interim period after Intervention 2) and after May 2023 (last part of Intervention 3).

Before turning to treatment effects, we assess model selection. We use the coefficient of variation (R^2) in the bottom panel as proxy for model fit. Moving from column (1) to columns (2) and (3), the R^2 increases substantially with the addition of fixed effects. The simplest model explains 26 percent of the variation in the outcome autarky, which increases by more than 30 percentage points when time-fixed effects are added. After adding household-fixed effects, the TWFE model in column (3) accounts for 78 percent of the variation. Notably, substituting time-variant weather controls for the time-fixed effects results in a substantial drop in the R^2, suggesting that time patterns are not driven entirely by weather as an exogenous force. We find a very similar pattern for self-consumption in the lower panel. For both outcomes, the sign of the coefficients is robust across all four columns, but the effect sizes and the standard errors increase as we build towards the TWFE model. This indicates that care must be taken in accounting for household and time heterogeneity, as the simpler models tend to understate the estimated treatment effect.

Based on these preliminaries, we consider column (3) the main estimate of the analysis. In the following, we focus on this column. For the feedback intervention (N = 1), there is a small, positive treatment effect. The coefficient for autarky indicates that the intervention increased autarky by 2.1 percentage points, a moderate improvement of 3.8 percent when evaluated against the mean outcome of 0.55 (Table 1 for reference). The coefficient on self-consumption indicates a 2.9 percent increase in self-consumption. Evaluated at the sample mean, this translates to an improvement of 13 Wh per hour on average. While the effect sizes are similar for both outcomes, the estimate is highly significant for autarky, but not for self-consumption (only at the 10 percent-level of confidence).

Regarding the benchmark intervention (N = 2), the effects are again positive and of similar magnitude as Intervention 1. For self-consumption, the effect sizes vary substantially across columns, and the result is not statistically significant in the conservative estimates (Columns 3 and 4). This likely reflects the higher volatility of self-consumption relative to autarky-rate, which leads to unstable coefficients in specifications that do not control for heterogeneity across households. Comparing the estimates in the TWFE model, a Wald test fails to reject the null hypothesis of equal coefficients (p-value = 0.955 for autarky-rate, p-value = 0.961 for self-consumption, see supplementary-material 1.3). This indicates that the feedback and the benchmark intervention do not differ in their effectiveness.

The consistent picture of small, positive effects from the first two interventions does not carry to the default intervention (N = 3). For autarky, the effect size is negligible from an economic perspective despite the statistical significance. Yet, for self-consumption, there is a sizable increase of 11 percent in self-consumption. Evaluated at the hourly sample mean, this translates to an increase of 49 Wh. Wald tests against Intervention 1 and Intervention 2 reject equality of coefficients for both outcomes (all p-values $<$0.001), indicating that Intervention 3 does indeed work differently.

When self-consumption rises, but autarky remains unaffected, the likely explanation is that households simultaneously increased total energy consumption. Autarky-rate as the ratio would then be constant. To substantiate this interpretation, we also ran the same model with total energy consumption as the outcome variable (not shown here). We found a significant increase of about 8 percent, which suggests that households increased both the denominator and the numerator of the autarky-rate. Intervention 3 is found to be more effective in increasing self-consumption, but ineffective for autarky. This finding suggests that more is needed to understand the mechanism of Intervention 3 compared to the other two interventions. We explore this further in the following sections.

Finally, we conduct a number of robustness checks and run the regression separately for each intervention to support the stability of the estimates. The list of robustness checks is included in Appendix 8.2. Code and documentation are available from the authors upon request.

3.2. Short-run effects

One explanation for the small average effects in the main results above could be that consumers quickly lose interest rather than adapting their routine due to the interventions [43]. We test this with the event study design displayed in Fig. 6. Time is centered to zero as the day a specific intervention becomes effective. For a better overview, the specification reports only 20 lead and lag terms (daily coefficients before and after the intervention started) and aggregates the other daily coefficients as endpoints (see Ref. [44]). Individual coefficients are plotted as black circles; the endpoints are represented as hollow circles. Across all interventions and for both outcome variables, the point estimates are clustered closely around the horizontal line at zero. The confidence intervals also span zero in the vast majority of cases. Corresponding to the main results, the time pattern appears less volatile for autarky than self-consumption at least for Interventions 1 and 2. Overall, the event study does not support a clear time trend within the study period. In fact, individual coefficients are insignificant, which indicates that single-day effects are small and the positive average found in the main result

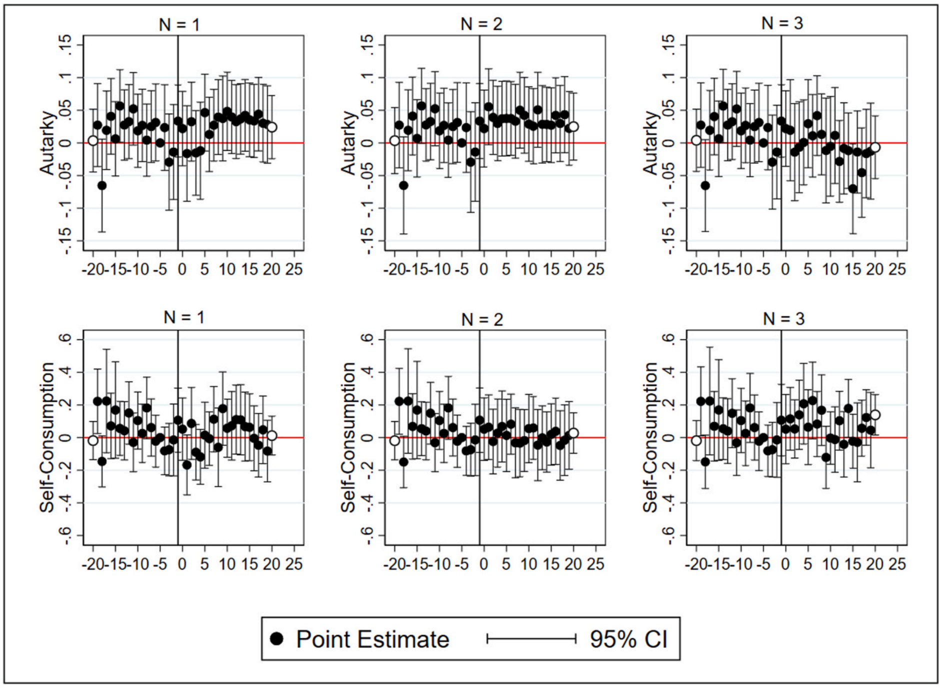

Fig. 6. Event study results.

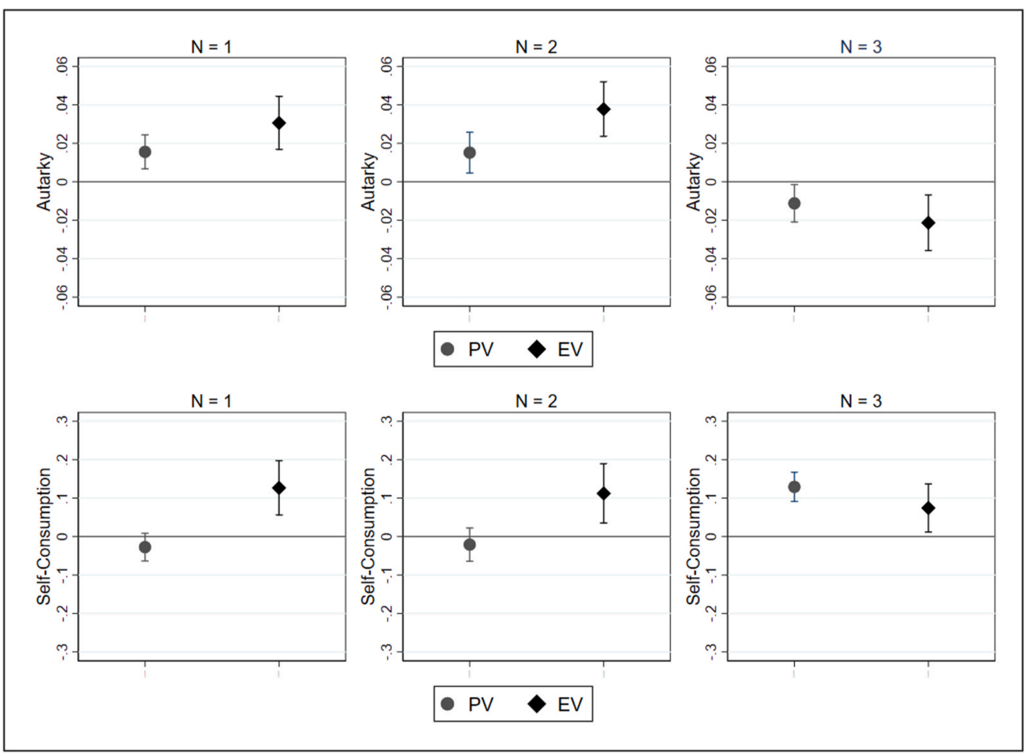

Fig. 7. Sub-group analysis.

emerges only in the aggregate. Given power constraints with the small cross-section relative to the number of parameters, this is expected. On the flip side, the study also lends credibility to the parallel trends assumption, as the pre-treatment effects are tightly clustered around zero. In economic terms, the event study further supports that the three tested interventions have small effects within the ecosystem of prosumers' energy consumption.

3.3. Sub-group analysis

A unique feature of the experiment is the sub-division into the participants without ("PV-group") and with controllable EVs ("EV-group"). We estimate the effects separately for these sub-groups to explore heterogeneity. This is shown in Fig. 7. The specification is the same TWFE model as in the main results but displayed in graphical form for exposition: the circle and diamond symbols represent the coefficients, i.e.,

Table 3
Intra-day shifts during intervention 3.

	(1)	(2)	(3)	(4)	(5)	(6)
	Autarky-rate		Self-Consumption		Total Consumption	
Active x Midday	**−0.00217**	**0.0132**	**0.165****	**0.157****	**0.165****	**0.135***
	(-0.15)	(0.94)	(2.28)	(2.17)	(2.26)	(1.84)
Active x PM	**0.0114**	**−0.00871**	**−0.0587**	**−0.130**	**−0.0117**	**−0.0313**
	(0.42)	(-0.32)	(-0.39)	(-0.87)	(-0.15)	(-0.39)
Midday	0.184***	0.173***	0.721***	0.802***	0.176***	0.260***
	(80.26)	(44.97)	(66.91)	(42.21)	(23.91)	(20.68)
PM	−0.0209***	0.00600	0.118***	0.260***	0.0972***	0.170***
	(-7.42)	(1.29)	(8.59)	(11.25)	(13.83)	(14.69)
R^2	0.352	0.322	0.200	0.193	0.250	0.196
Obs.	72802	26325	69347	25270	72802	26325
Control-group	All	EV only	All	EV only	All	EV only

Notes: Regression testing for intra-day shifts during Intervention 3. Data at hourly frequency. Baselevel is AM (6:00–10:00). Active is an indicator for interaction with the app. Robust standard errors in parentheses. Significance Levels: *p < 0.10, **p < 0.05, ***p < 0.01.

point estimates for the marginal effect, and the vertical extensions the 95 percent confidence interval. Autarky-rate is displayed in the upper panel, self-consumption (log-transformed) in the lower panel. The columns correspond to the three interventions.

For Interventions 1 and 2, the EV-group appears more responsive than the PV-group. While the confidence intervals overlap for autarky-rate, the sub-group differences are statistically significant for self-consumption. The EV-group has self-consumption treatment effects in the range of 10–12 percent, which is substantially above the average effect of 2–3 percent in the main analysis. Across both outcomes, the analysis suggests that the positive average effect is driven more by the EV-group.

However, this does not hold for Intervention 3. The confidence intervals of the two sub-groups overlap substantially for both outcomes, and the associated p-values do not support sub-group differences (not reported here). This result is surprising because especially Intervention 3 was targeted to the EV-group. The PV-group only received the energy report, whereas the EV-group had a new charging mode. The sub-group analysis does not support the interpretation that the increase in self-consumption is driven by the EV-group, which was the working hypothesis derived from the main results.

However, the presented effects come from assignment to the treatment-group (intention-to-treat effect). Using the additional information available from the smart-charging-app, we explore intraday-shifts separating those households that activated the new charging mode of Intervention 3 (n = 9), and those that did not. The hypothesis is that the active group changed their charging behavior to longer plug-in times, so the smart-charging-mode would shift consumption to the midday-PV-peak. We add energy consumption (again log-transformed) based on the insights from the main result.

The regression results are shown in Table 3 below. The coefficients of interest are in bold in the top two rows: the interaction terms reveal whether the active group shifts more into midday (11am-3pm), i.e., relatively more than the control-group. For each outcome, the first column uses all inactive households as the control-group; the second column uses only those in the EV-group, i.e., only participants who had access to the mode. This means a loss of observations but serves as a robustness check against concerns that those in the PV-group are not suitable control-group for the EV sample. As before, the results show no significant effects for autarky-rate but a strong positive shift to the midday hours for self-consumption and total consumption (Active x Midday). The effect sizes of 15–17 percent for self-consumption are substantially larger than the main result. Total consumption increases by a similar magnitude. We do not find significant differences in the evening hours (Active x PM) across all outcomes, which indicates that the midday increase is not offset by opposite changes during evening hours. The base effects for Midday and PM in the lower rows conform to expectations from normal load profiles. The results are interpreted as revealing the potential of the new mode, thus indicating that the weak effects in Fig. 7 stem from a low activation level. By contrast, participants that activate the new charging mode are able to use their PV-generation more effectively. In brief, the default intervention has a high potential for consumption shifts that is not captured in the overall sample because a relatively small sub-group drives it.

Connecting the insights from sections 2.2 and 2.3, the question is whether the default intervention can also induce more regular and, thus, sustainable behavior changes. With the small sample and limited uptake, we can only provide indicative, descriptive evidence here. Testing for variance equality (Levene-test) shows a minor *increase* in the variance of self-consumption, but fails to reject the null hypothesis. By contrast, the correlation between solar radiation and self-consumption increases sharply from 0.22 before the default to 0.73 afterwards. This indicates that the intervention increases the alignment with the relevant variation (solar radiation), but does not decrease the *unconditional* variation in the outcome.

4. Discussion

The treatment-group's positive, small intervention effects are of similar magnitude as other studies estimating causal effects regarding energy-saving behavior (e.g., Refs. [23,24,32–35]). We did expect our results to be at the lower end of the effect spectrum in the literature on behavioral interventions due to the publication bias and lack of causal effect methods in other studies. At the same time, the larger effects for the active EV-group even range between the few available studies with EVs (e.g., Refs. [45–47]) and model-based studies optimizing self-consumption under optimal conditions (e.g., Refs. [11,12]). In summary, we show that the tested interventions for feedback and benchmarking are suitable for increasing self-consumption by changing the described choice options. Additionally, the charging default increases self-consumption effectively by re-structuring the choice task. In the following, we provide a methodological reflection, position the results, and suggest subjects for further research.

Estimating causal effects with smart-meter-data requires careful consideration of the identification strategy to extract the relevant variation from the overall noise, which our results demonstrate. We provide treatment effects using a conservative TWFE specification of the broader DiD estimation, which applies microeconomic methods in this interdisciplinary setting. In the process, we showcase the difficulty of assessing treatment effects from real-life settings: the bulk of the variation in the smart-meter-data stems from general differences across households and time. Interventions on (self-)consumption behavior cannot change the external conditions, leaving a limited margin for optimization because much of the variation is "pre-determined". However, the third intervention also indicates strong opportunities for new behavioral routines (EV-charging) that are *aligned* with exogenous variation (solar radiation).

Our employed model is a strong improvement relative to pooled ordinary least-squares, but it is not a panacea for all confounders. The fixed-effects strategy rests on the assumption that within-household behavior is constant over time and can therefore be partialled out (see Ref. [42]). This strategy does not address time-variant confounders such as newly added assets. This is most critical for intervention 3, which begins during heating season. We ensure that heatpump ownership is balanced across all four sub-groups. However, we cannot completely rule out this potential confounder (e.g., different operation across households). This is similar to the battery-storage-systems, which are operated all year long but more intensively during the summer period.

Similarly, time fixed-effects absorb factors like solar radiation common to all households on a given day, but the treatment effect is the average effect across all households. Essentially, we take the assumption that factors like weather and energy prices are common to the treatment and the control group on a given day. In practice, the approach assumes for energy prices that this is a common shock to all households, and that the groups respond similarly on average – not only regarding the direct price effect, but also how susceptible households are to energy-related information in the intervention. The tested interventions show limited behavioral effects relative to the ecosystem, but the *capacity* for exploiting solar radiation indirectly impacts how prosumers optimize self-consumption. In the summer, when some households are close to complete autarky-rate, there may not be room left for the intervention to increase it further. In the winter, there are days with very little radiation and potential to exploit. Hence, interpreting effect sizes across seasons deserves a note of caution. Moreover, the seasonal yield differences imply that, for the intervention design, interventions should stimulate additional consumption for the excess generation in summer and focus on shifts of existing consumption in winter.

The rise of smart-metering makes data for such estimations easily accessible. At the same time, the data quality is prone to technical and human failures, such as connection issues, which increase noise that is difficult to separate from systematic variation. If such issues cannot be fully mitigated, it is key to understand their implication on the results. For instance, for some participants, the winter break created longer disconnection times. If these disconnection times are due to absence from home, we would conclude that data is missing from a below-average consumption period. Studying such correlation between data issues and human behavior and deriving best practices for handling them are subjects for further research.

Human behavior could also impact the results due to social desirability. Participants were aware of being part of an experiment and, thus, may intentionally pay special attention. However, showing social desirability in everyday life over a time span of 1.5 years appears difficult [48]. In addition, the event study did not show differences over time. Thus, we believe that social desirability did not affect the results (largely). Studies examining long-term effects of behavioral interventions may consider and assess this in more detail.

Self-selection in our sample creates limitations regarding the external validity of our results. With the highly motivated prosumers and their sizable asset portfolio to optimize over (see supplementary-material 1.2), the sample does not represent the German population. At the same time, it shows typical characteristics of early adopters of rooftop-PV and EVs [49,50], who are the current target group for this kind of intervention. Learnings from this group give insights into how to support other households at the later stages of the diffusion curve [3, 51]. The increasing diffusion is expected to lead to greater household-dependent variations and an elevated need to tailor interventions to household conditions. Thus, our work lays a basis for further research.

Remarkably, even our self-selected sample takes up the interventions only to a limited extent. In particular, the efficacy of the charging default is weakened since only half of the EV-group activated the feature. While no significant effect for the overall EV-group is found, a comparison of the non- and activated participants shows a 15–17 percent increase in self-consumption for activated prosumers. Acknowledging the risk of low uptake, we recommend designing interventions with only a minimum amount of required interaction. In our case, we believe a charging default without the need for an initial activation would likely be more effective.

Such low uptake of interventions demonstrates limitations, which policymakers may face when upscaling behavioral interventions as policy measures. In this sense, the main results for the entire sample are a more accurate projection for policy measures since they measure the effect for the ones that were assigned to the treatment (intention-to-treat) and not the sub-group that was certainly exposed to the treatment. The uptake is likely to decrease further when the interventions are rolled out to the - likely less motivated - German population. At the same time, since the German population is younger than the sample, a rollout would target more digital natives, which may increase the chances for an uptake.

The results of the feedback and benchmark interventions confirm the efficacy of their common design aspects, i.e., condensed information presented in an appealing manner for describing choice options. Although we recognize an incremental improvement when a benchmark is added as a second intervention, both effects are not significantly different. Consequently, no conclusions can be derived from the distinctive design aspects that come with describing choice options (i.e., whether signaling colors are a more effective stimulus than benchmarking).

The effects are driven by the sub-group with controllable EVs. The increased effect size emphasizes the opportunity for technology-specific intervention designs that align with the strong exogenous drivers of the outcome of interest. The stronger results for the EV-group in interventions 1 and 2 (compared to the non-EV group) suggest that there is potential for implementing interventions while EVs and other electrified residential technologies are still emerging and new routines around them are created. From a different angle, the large intra-day effects in intervention 3 fit with this interpretation, albeit conditional on active utilization. Since these emerging technologies are already equipped with digital interfaces, behavioral interventions could also be integrated at a low cost. However, in our study, the additional interface for the EV-group does not allow us to clearly distinguish between the impact of the technology and the interface. Future research could disentangle both factors. Furthermore, it could test the effect on other flexible technologies (e.g., heatpumps) and on households who are not yet prosumers, further assessing the heterogeneity and context-specificity of behavioral interventions. Thereby, other aspects of behavioral interventions from the literature could be further examined, e.g., (i) the relation of applied interventions to normativity [52], (ii) their link to economic incentives [53], and (iii) their focus on the individual's or the society's welfare [54].

5. Conclusion

This paper has studied the effectiveness of three behavioral interventions within the same field experiment using a more rigorous estimation framework than much of the previous literature. We find small, positive effects for interventions through feedback and benchmarking, both in absolute self-consumption and in relative terms (autarky-rate). Sub-group analysis shows that the EV-group mainly drives the average effects. The default intervention stands out as different from the others: it increases self-consumption substantially but is ineffective for autarky-rate as total consumption increases simultaneously. We are able to show that the low uptake of the intervention explains the weak average effect. By contrast, the prosumers adopting the smart-charging-mode can increase self-consumption by 16–17 percent. As a subject for further research, we suggest exploring how behavioral interventions interact with the households' charging routine.

Overall, we contribute novel evidence on stimulating prosumers to optimize self-consumption, which is a previously understudied use case of behavioral interventions with growing potential in the energy transition and consumption shifts. The study extracts intervention effects from a real-life field experiment, which reveals that behavioral interventions target a relatively small component within the ecosystem of household energy consumption.

The uptake is likely to deteriorate for other, less dedicated prosumers. Also, the prosumers' level of energy literacy is likely to be lower, which makes interventions that change the described choice options (e.g., feedback) less attractive than interventions that re-structure the choice task (e.g., default). Future applied work could explore specifically how intervention design can be better embedded in the ecosystem.

Subtle interventions require supporting regulatory, technical, and digital conditions. The other way around, restrictive self-consumption regulation, unappealing digital interfaces, and mal-functioning flexible technologies can easily overrule the small, positive treatment effects. At the same time, if behavioral interventions are thoughtfully aligned to these conditions, they can unlock hard-to-reach flexibility potential. Orchestrating a grid-friendly operation of large consumption technologies, such as EVs and heatpumps, is a promising future case for behavioral interventions in light of emerging flexibility markets, digitalization, and other grid regulations.

CRediT authorship contribution statement

Sabine Pelka: Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization. **Anne Kesselring:** Writing – original draft, Software, Methodology, Formal analysis. **Sabine Preuß:** Writing – review & editing, Methodology, Investigation, Conceptualization. **Emile Chappin:** Writing – review & editing. **Laurens de Vries:** Writing – review & editing.

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.

Data availability

The data that has been used is confidential.

Acknowledgements

This work was supported by the European Union's Horizon 2020 research and innovation program project NUDGE under grant agreement no. 927012. The authors gratefully acknowledge the contributions of Werner Neumeier and Dorina Rauth from the service provider Beegy and Johanna Isenhuth and Christian Schäfer from the utility MVV Energie AG for carrying out the field experiment. The collection of the smart-meter-data was orchestrated by the service provider domX (in particular, Stratos Keranidis, Dimitris Voulgarakis, Polychronis Symeonidis), and the collection of the survey data by the research institute imec (in particular, Stephanie van Hove, Emma Martens). Thank you for your enduring support and endless patience. We also would like to express our gratitude to Peter Conradie, Merkouris Karaliopoulos, Katharina Wohlfarth, and Marian Klobasa for their guidance in approaching this paper. Furthermore, we thank Peter Spraul for supporting the analyses with complementary survey evaluations.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.segy.2024.100140.

Appendix

8.1 Design of the Interventions for the Smart-Charging-App (intervention 1 and 2) and the Monthly Report Download (intervention 3)

A description of the interventions can be found in the supplementary-material 1.1.

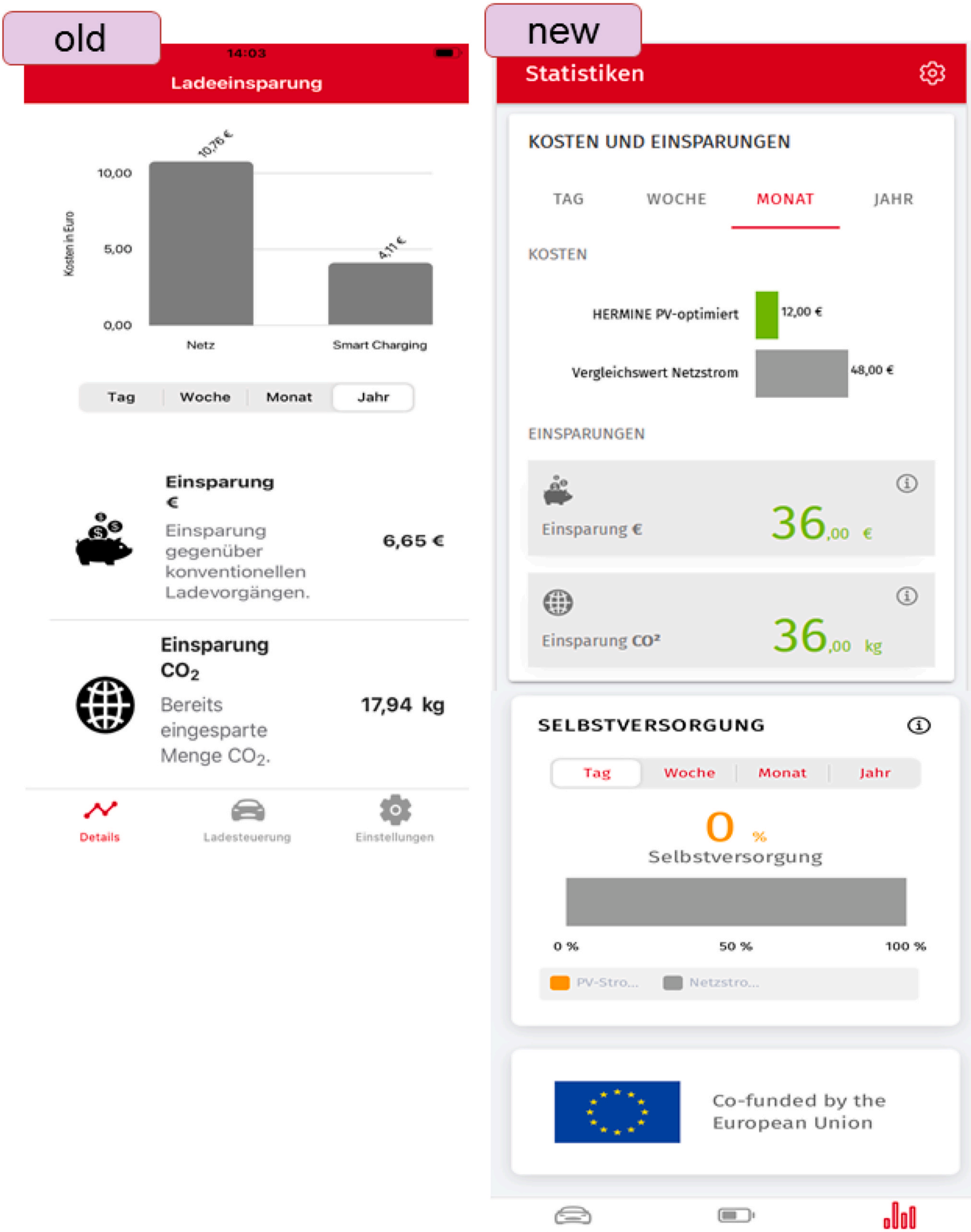

Fig. Annex 1. Intervention 1 for the EV-group providing simple indicators in signaling color.

Fig. Annex 2. Intervention 2 for the EV-group providing benchmarks of previous and current self-consumption of charging and upcoming optimized charging process.

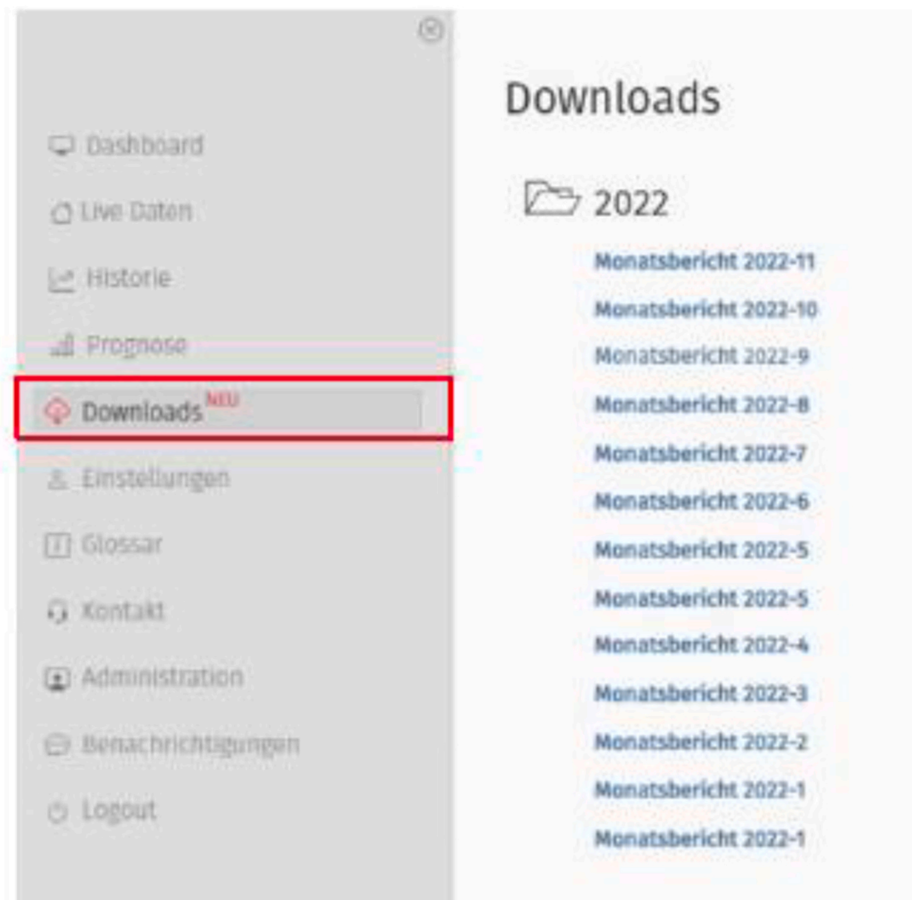

Fig. Annex 3. Intervention 3 for both groups providing aggregated information on past self-consumption in form of energy reports.

8.2. List of Robustness Checks

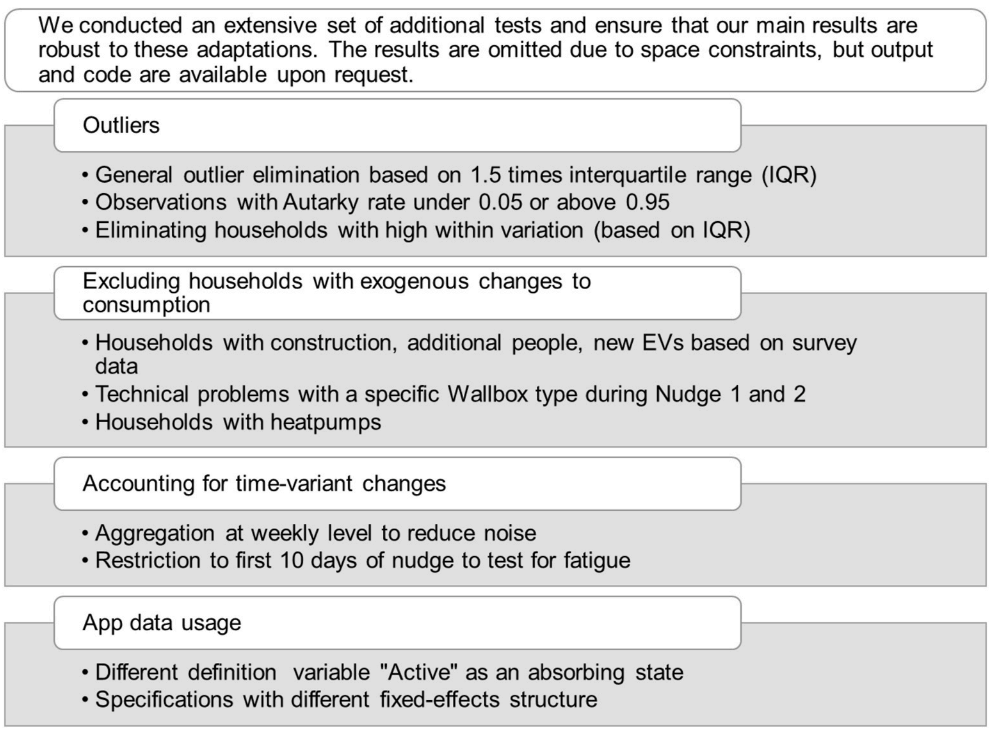

Fig. Annex 4. List of robustness checks.

8.3. Descriptive Development between EV and PV-group over Time

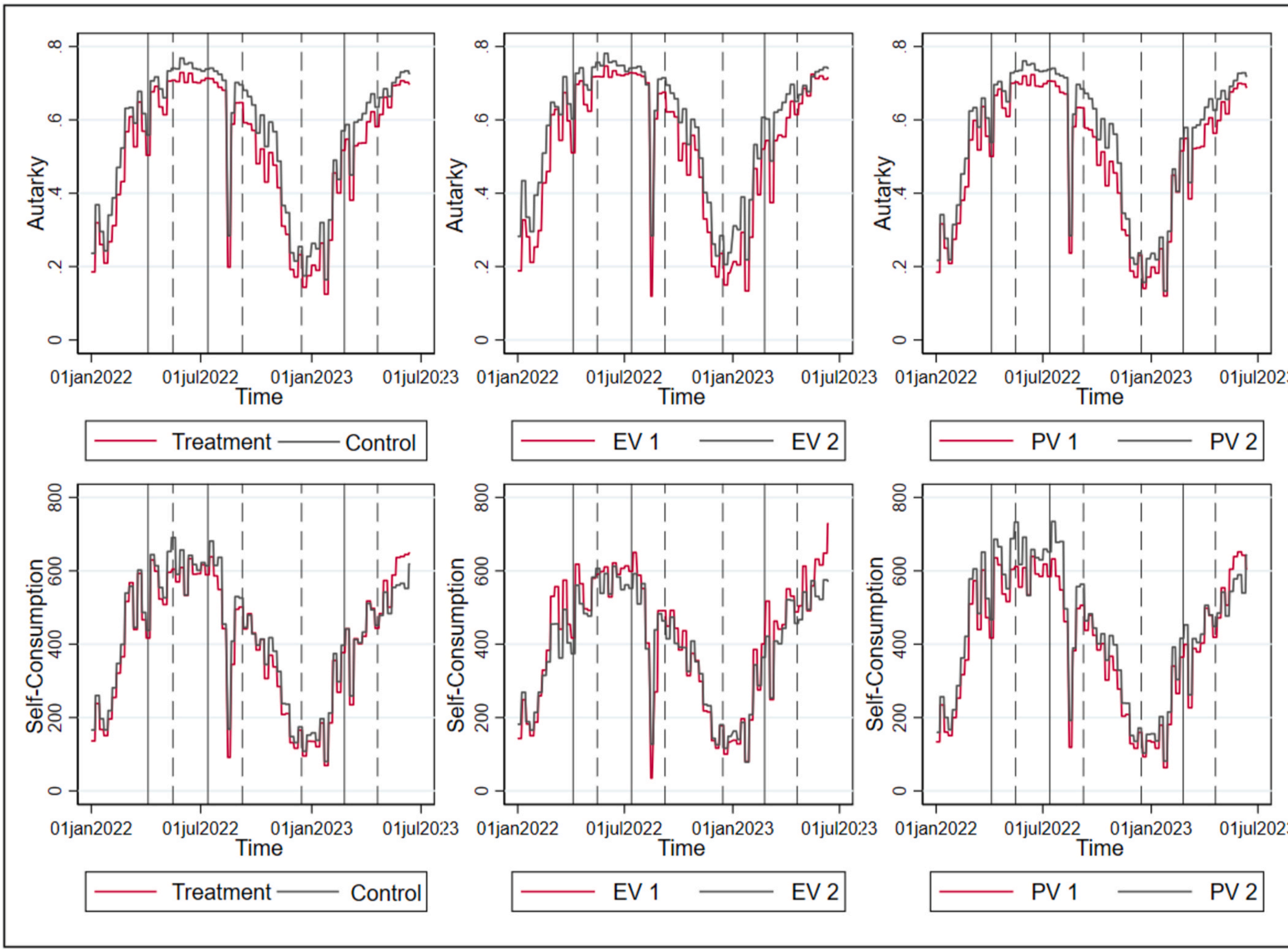

Fig. Annex 5. Outcomes by sub-group over time at weekly aggregation
Notes: Left panel refers to full sample, middle panel to EV-group, right panel to PV-group. Index 1 refers to sub-sample of treatment-group, index 2 to control-group. Aggregated to weekly means for exposition.

References

[1] Nyholm E, Goop J, Odenberger M, Johnsson F. Solar photovoltaic-battery systems in Swedish households – self-consumption and self-sufficiency. Appl Energy 2016; 183:148–59. https://doi.org/10.1016/j.apenergy.2016.08.172.
[2] Schopfer S, Tiefenbeck V, Staake T. Economic assessment of photovoltaic battery systems based on household load profiles. Appl Energy 2018;223:229–48. https://doi.org/10.1016/j.apenergy.2018.03.185.
[3] van der Stelt S, AlSkaif T, van Sark W. Techno-economic analysis of household and community energy storage for residential prosumers with smart appliances. Appl Energy 2018;209:266–76. https://doi.org/10.1016/j.apenergy.2017.10.096.
[4] Venizelou Venizelos, Philippou Nikolas, Hadjipanayi Maria, George Makrides, Efthymiou Venizelos, Georghiou George E. Development of a novel time-of-use tariff algorithm for residential prosumer price-based demand side management. Energy 2018;142:633–46. https://doi.org/10.1016/j.energy.2017.10.068.
[5] Dehler J, Keles D, Telsnig T, Fleischer B, Baumann M, Fraboulet D, et al. Self-consumption of electricity from renewable sources. In: Europe's energy transition - insights for policy making. Elsevier; 2017. p. 225–36.
[6] Luthander Rasmus, Widén Joakim, Nilsson Daniel, Palm Jenny. Photovoltaic self-consumption in buildings: a review. Appl Energy 2015;142:80–94. https://doi.org/10.1016/j.apenergy.2014.12.028.
[7] Luthander R, Widén J, Munkhammar J, Lingfors D. Self-consumption enhancement and peak shaving of residential photovoltaics using storage and curtailment. Energy 2016;112:221–31. https://doi.org/10.1016/j.energy.2016.06.039.
[8] Linssen J, Stenzel P, Fleer J. Techno-economic analysis of photovoltaic battery systems and the influence of different consumer load profiles. Appl Energy 2017; 185:2019–25. https://doi.org/10.1016/j.apenergy.2015.11.088.
[9] Kaschub T, Jochem P, Fichtner W. Solar energy storage in German households: profitability, load changes and flexibility. Energy Pol 2016;98:520–32. https://doi.org/10.1016/j.enpol.2016.09.017.
[10] Kuckshinrichs W, Ball CS, Aniello G. Levelized profits for residential PV-battery systems and the role of regulatory and fiscal aspects in Germany. Energy, Sustainability and Society 2023;13(1). https://doi.org/10.1186/s13705-023-00390-8.
[11] Higashitani T, Ikegami T, Uemichi A, Akisawa A. Evaluation of residential power supply by photovoltaics and electric vehicles. Renew Energy 2021;178:745–56. https://doi.org/10.1016/j.renene.2021.06.097.
[12] Kern T, Dossow P, Morlock E. Revenue opportunities by integrating combined vehicle-to-home and vehicle-to-grid applications in smart homes. Appl Energy 2022;307:118187. https://doi.org/10.1016/j.apenergy.2021.118187.
[13] Wolske KS, Gillingham KT, Schultz PW. Peer influence on household energy behaviours. Nat Energy 2020;5(3):202–12. https://doi.org/10.1038/s41560-019-0541-9.
[14] Andor MA, Fels KM. Behavioral economics and energy conservation – a systematic review of non-price interventions and their causal effects. Ecol Econ 2018;148: 178–210. https://doi.org/10.1016/j.ecolecon.2018.01.018.
[15] Alipour M, Salim H, Stewart Rodney A, Sahin Oz. Residential solar photovoltaic adoption behaviour: end-to-end review of theories, methods and approaches. Renew Energy 2021;170:471–86. https://doi.org/10.1016/j.renene.2021.01.128.
[16] Sunstein C. Behaviorally informed mandates? Internalities, externalities, and fuel economy rules. 29 NYU Environmental Law Journal 2021:2021.
[17] Gerarden TD, Newell RG, Stavins RN. Assessing the energy-efficiency gap. J Econ Lit 2017;55(4):1486–525. https://doi.org/10.1257/jel.20161360.
[18] Thaler RH, Sunstein CR. Nudge: improving decisions about health, wealth, and happiness. Yale University Press; 2008.

[19] Weinmann M, Schneider C, vom Brocke J. Digital nudging. Business & Information Systems Engineering 2016;58(6):433–6. https://doi.org/10.1007/s12599-016-0453-1.

[20] Hummel D, Maedche A. How effective is nudging? A quantitative review on the effect sizes and limits of empirical nudging studies. Journal of Behavioral and Experimental Economics 2019;80:47–58. https://doi.org/10.1016/j.socec.2019.03.005.

[21] Johnson EJ, Shu SB, Dellaert BGC, Fox C, Goldstein DG, Häubl G, et al. Beyond nudges: tools of a choice architecture. Market Lett 2012;23(2):487–504. https://doi.org/10.1007/s11002-012-9186-1.

[22] Momsen K, Stoerk T. From intention to action: can nudges help consumers to choose renewable energy? Energy Pol 2014;74:376–82. https://doi.org/10.1016/j.enpol.2014.07.008.

[23] Bager S, Mundaca L. Making 'Smart Meters' smarter? Insights from a behavioural economics pilot field experiment in Copenhagen, Denmark. Energy Res Social Sci 2017;28:68–76. https://doi.org/10.1016/j.erss.2017.04.008.

[24] Myers E, Souza M. Social comparison nudges without monetary incentives: evidence from home energy reports. J Environ Econ Manag 2020;101:102315. https://doi.org/10.1016/j.jeem.2020.102315.

[25] Houde S, Todd A, Sudarshan A, Carrie Armel K. Real-time feedback and electricity consumption: a field experiment assessing the potential for savings and persistence. EJ, https://doi.org/10.5547/01956574.34.1.4; 2013.

[26] Asensio OI, Delmas MA. Nonprice incentives and energy conservation. Proc Natl Acad Sci U S A 2015;112(6):E510–5. https://doi.org/10.1073/pnas.1401880112.

[27] Ruokamo E, Meriläinen T, Karhinen S, Räihä J, Suur-Uski P, Timonen L, et al. The effect of information nudges on energy saving: observations from a randomized field experiment in Finland. Energy Pol 2022;161:112731. https://doi.org/10.1016/j.enpol.2021.112731.

[28] Dominicis S de, Sokoloski R, Jaeger CM, Schultz PW. Making the smart meter social promotes long-term energy conservation. Palgrave Commun 2019;5(1). https://doi.org/10.1057/s41599-019-0254-5.

[29] Schleich J, Faure C, Klobasa M. Persistence of the effects of providing feedback alongside smart metering devices on household electricity demand. Energy Pol 2017;107:225–33. https://doi.org/10.1016/j.enpol.2017.05.002.

[30] Angrist JD, Krueger AB. Empirical Strategies in Labor Economics;vol. 3: 1277–1366. https://doi.org/10.1016/S1573-4463(99)03004-7.

[31] Pelka S, Conradie P, Vries L de, Anatolitis V, Martens E, Chappin E, et al. Self-consumption rises due to energy crises? An evaluation of prosumers' consumption behavior in 2022. 2023 19th International Conference on the European Energy Market (EEM) 2023:1–6. https://doi.org/10.1109/EEM58374.2023.10161968.

[32] Imbens GW, Wooldridge JM. Recent developments in the econometrics of program evaluation. J Econ Lit 2009;47(1):5–86. https://doi.org/10.1257/jel.47.1.5.

[33] Weigert A, Hopf K, Günther SA, Staake T. Heat pump inspections result in large energy savings when a pre-selection of households is performed: a promising use case of smart meter data. Energy Pol 2022;169:113156. https://doi.org/10.1016/j.enpol.2022.113156.

[34] Brown Z, Johnstone N, Haščič I, Vong L, Barascud F. Testing the effect of defaults on the thermostat settings of OECD employees. Energy Econ 2013;39:128–34. https://doi.org/10.1016/j.eneco.2013.04.011.

[35] Schleich J, Schuler J, Pfaff M, Frank R. Do green electricity tariffs increase household electricity consumption? Appl Econ 2022:1–12. https://doi.org/10.1080/00036846.2022.2102574.

[36] Gabriel M, Keranidis S, Voulgarakis D, Neumeier W, Vandewalle E, Nad L, et al. Report on pilot results: interim report: deliverable D4.1/D4.3 of the H2020 project NUDGE, for further information visit. https://www.nudgeproject.eu/wp-content/uploads/2022/11/NUDGE-D4.1-D4.3-Report-on-pilot-results-interim-report-_-FINAL-SUBMITTED.pdf2022.

[37] Greene W. Estimating econometric models with fixed effects. Working papers 01-10. New York University, Leonard N. Stern School of Business, Department of Economics; 2001.

[38] Andor M, Bensch G, Fels K, Kneppel N. Per stups zum energiesparen? Perspekt Wirtsch 2019;20(4):352–82. https://doi.org/10.1515/pwp-2018-0039.

[39] Collischon M, Eberl A. Let's talk about fixed effects: let's talk about all the good things and the bad things. Köln Z Soziol 2020;72(2):289–99. https://doi.org/10.1007/s11577-020-00699-8.

[40] Imai K, Kim IS. On the use of two-way fixed effects regression models for causal inference with panel data. Polit Anal 2021;29(3):405–15. https://doi.org/10.1017/pan.2020.33.

[41] Gangl M. Causal inference in sociological research. Annu Rev Sociol 2010;36(1):21–47. https://doi.org/10.1146/annurev.soc.012809.102702.

[42] Morrissey P, Weldon P, O'Mahony M. Future standard and fast charging infrastructure planning: an analysis of electric vehicle charging behaviour. Energy Pol 2016;89:257–70. https://doi.org/10.1016/j.enpol.2015.12.001.

[43] Sunstein CR. Nudges that fail. Behav. Public Policy 2017;1(1):4–25. https://doi.org/10.1017/bpp.2016.3.

[44] Callaway B, Sant'Anna PH. Difference-in-Differences with multiple time periods. J Econom 2021;225(2):200–30. https://doi.org/10.1016/j.jeconom.2020.12.001.

[45] van der Kam M, Peters A, van Sark W, Alkemade F. Agent-based modelling of charging behaviour of electric vehicle drivers. JASSS 2019;22(4). https://doi.org/10.18564/jasss.4133.

[46] Huber J, Weinhardt C. Waiting for the sun - can temporal flexibility in BEV charging avoid carbon emissions? Energy Inform 2018;1(S1). https://doi.org/10.1186/s42162-018-0026-2.

[47] Huber J, Jung D, Schaule E, Weinhardt C. Goal framing in smart charging - increasing bev users' charging flexibility with digital nudges. Association for Information Systems; 2019.

[48] Allcott H, Taubinsky D. Evaluating behaviorally motivated policy: experimental evidence from the lightbulb market. Am Econ Rev 2015;105(8):2501–38. https://doi.org/10.1257/aer.20131564.

[49] Wesche JP, Dütschke E. Organisations as electricity agents: identifying success factors to become a prosumer. J Clean Prod 2021;315:127888. https://doi.org/10.1016/j.jclepro.2021.127888.

[50] Plötz P, Schneider U, Globisch J, Dütschke E. Who will buy electric vehicles? Identifying early adopters in Germany. Transport Res Pol Pract 2014;67:96–109. https://doi.org/10.1016/j.tra.2014.06.006.

[51] Sarfarazi S, Deissenroth-Uhrig M, Bertsch V. Aggregation of households in community energy systems: an analysis from actors' and market perspectives. Energies 2020;13(19):5154. https://doi.org/10.3390/en13195154.

[52] Carlsson F, Gravert CA, Kurz V, Johansson-Stenman O. Nudging as an environmental policy instrument. Available at: SSRN CeCAR Working Paper Series No. 4 2019:3711946. https://doi.org/10.2139/ssrn.3711946.

[53] Congiu L, Moscati I. A review of nudges: definitions, justifications, effectiveness. J Econ Surv 2022;36(1):188–213. https://doi.org/10.1111/joes.12453.

[54] Schubert C. Green nudges: do they work? Are they ethical? Ecol Econ 2017;132:329–42. https://doi.org/10.1016/j.ecolecon.2016.11.009.