Journal of Public Economics 238 (2024) 105191

Contents lists available at ScienceDirect

Journal of Public Economics

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

Norm-based feedback on household waste: Large-scale field experiments in two Swedish municipalities[☆]

Claes Ek[a,*], Magnus Söderberg[b,c]

[a] Department of Economics, University of Gothenburg, P.O. Box 640, SE-405 30 Gothenburg, Sweden
[b] Department of Accounting, Finance and Economics, Griffith University, Parklands Drive, Southport QLD 4222, Australia
[c] Ratio Institute, P.O. Box 3203, SE-103 64 Stockholm, Sweden

ARTICLE INFO

JEL classification:
C93
D13
D61
Q53

Keywords:
Household waste
Norm-based feedback
Social comparisons
Unit-based pricing
Pay-as-you-throw
Welfare cost of nudging

ABSTRACT

We conduct two large-scale randomized controlled trials to produce the first evidence that Home Energy Report-type norm feedback letters can be used to reduce household waste. We explore several feedback variants, including a novel short-run dynamic norm that emphasizes ongoing changes in waste behavior. Waste reductions are on the order of 7%–12% for all treatments, substantially larger than usually found in the energy or water domains. Effects are mostly driven by increased recycling of packaging and remain largely intact a year after the intervention ended. Feedback is highly cost-effective compared to alternative non-price waste policies. However, net social benefits depend on household preferences for receiving feedback, which we elicit in a valuation survey, and whether existing waste fees internalize the marginal social cost of waste.

1. Introduction

Large-scale interventions promoting household resource conservation through norm-based feedback have become a mainstay of applied behavioral economics over the past decade.[1] The most well-known example of such a norm-based intervention is the "Home Energy Report" (HER) developed by Opower and mailed to households across the United States (see e.g., Allcott, 2011; Ayres et al., 2012; Costa and Kahn, 2013). HERs present household-specific feedback on energy use compared to a set of similar neighbors, and includes both injunctive ("ought") and descriptive ("is") norm components.

Effects from norm feedback interventions appear to systematically differ in size across contexts and domains. Feedback on water use typically drives reductions of up to about 5% in participating households (Ferraro and Price, 2013; Bernedo et al., 2014; Brick et al., 2017; Jaime Torres and Carlsson, 2018; Jessoe et al., 2021; Goetz et al., 2022). Effect sizes for electricity use are generally smaller: in an analysis of over 100 different large-scale HER experiments in the

☆ We thank the editor, Robert Metcalfe, and three anonymous reviewers for useful comments that substantially improved our manuscript. We also thank Lina Andersson, Petyo Bonev, Pol Campos-Mercade, Fredrik Carlsson, Li Chen, Martin Dufwenberg, Olof Johansson-Stenman, Mitesh Kataria, Peter Martinsson, and Andreas Nilsson, as well as participants at the 2020 Markets and Sustainability Symposium at the University of Groningen, the 2021 Conference of the European Association of Environmental and Resource Economists (EAERE), the 2023 Workshop in Environmental Policy Evaluation at the University of St. Gallen, and seminar participants at the University of Adelaide, University College Dublin, University of Gothenburg, Griffith University, University of Hamburg, University of Helsinki, University of Queensland, and the Swedish University of Agricultural Sciences, for valuable suggestions and discussions. We are grateful to Camilla Bergquist and Malin Persson for initial assistance in building the partnerships underlying this research. Ruben Dieleman, Tommy Strandelin, and Linus Svärm provided valuable research assistance. Financial support from the Swedish research council Formas, grant nos. 2017-00225 and 2018-02603, is gratefully acknowledged. The experiments were pre-registered at the American Economic Association Registry for randomized controlled trials, trial identification no. AEARCTR-0003301. Declarations of interest: none.

* Corresponding author.
E-mail address: claes.ek@economics.gu.se (C. Ek).

[1] The rapidly growing literature, starting from Schultz et al. (2007), now includes a large number of studies evaluating specific designs (Allcott, 2011; Ayres et al., 2012; Costa and Kahn, 2013; Ferraro and Price, 2013; Dolan and Metcalfe, 2015; Jaime Torres and Carlsson, 2018; Holladay et al., 2019; Brülisauer et al., 2020), as well as longer-run effects (Ferraro et al., 2011; Allcott and Rogers, 2014; Bernedo et al., 2014), welfare implications (Allcott and Kessler, 2019), and psychological mechanisms (Alberts et al., 2016; Byrne et al., 2018).

https://doi.org/10.1016/j.jpubeco.2024.105191
Received 26 August 2022; Received in revised form 1 May 2024; Accepted 15 July 2024
Available online 30 July 2024

US, Allcott (2015) finds an average reduction of 1.31%, with a standard deviation of 0.45 percentage points. Given the documented range of effects, there has recently been some debate on the value and cost-effectiveness of feedback interventions in different settings.[2] Our paper adds to that discussion by providing the first evidence on HER-type feedback on household waste.[3] To our knowledge, no other study has tested the large-scale impact of such interventions in this domain.

In the literature on HER, it seems plausible that the range of observed effect sizes reflect differences in the cost of reducing usage. If so, there is good reason to expect larger effects from feedback on waste compared to electricity and water use. According the 2015 Residential Energy Consumption Survey of the Energy Information Administration,[4] the bulk of home electricity use in the US (the setting for most HER studies to date) is for air conditioning, refrigerators, and space and water heating, with only about 10% due to lighting. Making deep cuts in usage would thus involve either high-effort behavior change such as turning down the heat in winter, or costly physical capital investments (Allcott and Rogers, 2014). A similar, but less stark pattern seems likely to hold for water use (Bernedo et al., 2014). By contrast, recycling allows households to reduce residual waste without forgoing consumption utility. Recycling also normally requires little in the way of physical capital investment beyond buying a set of in-home waste bins. Unlike for electricity or water, this physical capital is largely a complement rather than a substitute for effort (Vollaard and van Soest, 2024), but even so, the marginal effort of increasingly diligent recycling seems less steep compared to reductions in other domains.[5] Effects of HER-style feedback may thus prove particularly large for waste.

Reducing waste is a policy priority in the EU and around the world since waste disposal, and particularly landfilling, contributes to climate change by causing emissions of various greenhouse gases; according to Eurostat, about 3% of total EU-wide greenhouse-gas emissions are due to waste. Waste also poses risks to human health and ecosystem services, e.g., through leaking landfills and plastic pollution. In addition, recycling reduces the need for virgin materials. The UN 2030 Agenda for Sustainable Development thus includes targets to "substantially reduce waste generation through prevention, reduction, recycling and reuse" (Goal 12.5). The EU has also adopted far-reaching targets within the Waste Framework Directive; the main long-term goal is for 65% of household waste to be recycled by 2035. The latest report on implementation (European Commission, 2018) found that 14 member states were at risk of missing these targets.

We run two separate experiments, involving a total of some 20,000 households, in a pair of municipalities in western Sweden. Randomizing treatment across nearly all single-family homes, we send treated households repeated and accurate feedback over a period of eight months. The feedback, typically presented in a HER-style format, is on the amount of residual waste (in kg/person) that households generate compared to their neighbors.[6] We focus on residual waste to allow households to react not only by recycling more, but also through waste prevention, e.g., buying fewer packaging-intensive products. We also test a novel design that stresses short-run waste reduction in a household's comparison group. This is inspired by recent evidence in Sparkman and Walton (2017) and Mortensen et al. (2019) that "dynamic" or "trending" norms are more effective at changing behavior. For example, compared with snapshot information stressing cross-sectional variation, ongoing changes may send a stronger signal that behavioral costs and benefits are shifting in the population as a whole, increasing the chance that recipients update their view of the proper course of action.

In both experiments, high-precision average treatment effect (ATE) estimates show that the residual-waste weights of treated addresses drop after receiving the first letter. Effects of the short-run dynamic treatment are statistically indistinguishable from the standard HER design. Notably, reductions are consistently about 7%–12%, depending on the exact regression specification. Thus, our two experiments provide independent data points suggesting considerably larger effects than in previous studies of norm-based feedback. The reductions are about 50%–75% the size of that in Vollaard and van Soest (2024), who study a Dutch crackdown on incorrect sorting involving strong measures like fines and salient bin inspections. Also, our data extend a full year beyond the end of the intervention period, allowing us to examine longer-run effects as well; these turn out to be very persistent, remaining largely intact at the end of the additional year. Thus, our results show that norm-based interventions may be deployed as effective non-price instruments to reduce unsorted waste, and that effects from feedback do indeed vary strongly across resource domains.

In principle, the effect we observe may be driven by any combination of waste prevention, increased recycling, and illicit disposal (dumping). We examine mechanisms through a unique combination of evidence, involving (i) official records on local dumping incidents, (ii) municipal data on collected food waste, (iii) an endline survey, and (iv) a major waste-composition analysis of about 1000 kg of residual waste collected separately from treated and control households. We find no indication that treatment increases illicit disposal; instead, the main mechanisms appear to be increased recycling of packaging and (to a lesser extent) food waste.

The municipalities in our setting have pre-existing systems for unit-based pricing (UBP) of waste (weight-based fees). Indeed, our particular design is tailored to such pricing schemes: we partner with public utilities that routinely weigh each bin during collection, and feedback letters are constructed from the resulting weight data. As a result, our study also relates to the ongoing examination of marginal-cost pricing of waste, also known as "pay-as-you-throw".[7] As in the case of energy or water, utilities may be reluctant to raise marginal costs further,

[2] In particular, Andor et al. (2020) has argued that the relative success of HER in the US may be limited to that setting. Allcott (2011) estimates that, for effect sizes of about 1%–3%, the Opower HERs are cost-effective per unit of carbon emissions compared to other energy conservation policies. By contrast, Andor et al. (2020) replicate the HER design in Germany and observe a substantially smaller treatment effect of 0.7%. Since baseline per-capita electricity consumption is higher in the US than in other OECD countries, the authors conclude that norm-based feedback is unlikely to be cost-effective outside of the United States.

[3] Psychology studies that examine non-HER norm interventions in the waste domain include Schultz (1999), Dupré and Meineri (2016), and Bruchmann et al. (2021).

[4] eia.gov/consumption/residential/index.php.

[5] Also, exceptions involving substitutability with effort do exist and tend to involve actions to prevent waste, e.g., by placing a "no ads" sticker on one's mailbox.

[6] Throughout this paper, the term residual waste is used for the unsorted stream of household waste, which is typically incinerated in Sweden and many other OECD countries. Similarly, what we term food waste is the biodegradable stream, which differs from wasted food in that not all food waste is avoidable.

[7] Early studies of UBP have attempted to identify the causal effect of such schemes on waste generation and recycling by making before-after comparisons (Fullerton and Kinnaman, 1996), exploiting cross-sectional variation (Kinnaman and Fullerton, 2000), or both (Dijkgraaf and Gradus, 2009). More recently, several studies have used a regression difference-in-differences approach with unit (e.g., municipality) fixed effects (Allers and Hoeben, 2010; Usui and Takeuchi, 2014; Bucciol et al., 2015; Dijkgraaf and Gradus, 2017; Carattini et al., 2018). Bueno and Valente (2019) uses a synthetic-control strategy that appears to better model unobserved heterogeneity than fixed-effects approaches do. Some studies have also used IV approaches to control for endogenous policy (Kinnaman and Fullerton, 2000; Allers and Hoeben, 2010; Huang et al., 2011).

e.g. because of acceptance concerns.[8] Non-price policies, by contrast, may prove more readily implementable. Conversely, the fact that there is unique potential to add norm-based interventions on top of pay-as-you-throw schemes implies that UBP may hold greater promise than recognized. We estimate that unit-based fees need to increase by 32%–67% to produce effects of similar size to our main ATEs. Thus, norm feedback appears highly effective compared with economic incentives. Feedback is also very cost-effective compared to other non-price waste policies such as curbside collection of packaging.

Finally, we conduct a complete analysis of the net social benefits of our treatments. A key component of that analysis is households' willingness to pay (WTP) to (not) receive feedback, which we elicit in a separate, incentive compatible survey module sent out in late 2023, four years after our original experiments. Thus, we also contribute to the small empirical literature on the "welfare cost of nudging" (Damgaard and Gravert, 2018; Allcott and Kessler, 2019; Butera et al., 2022). Given these WTP measures, results on net social benefits are mixed: in one of the two municipalities, household mean WTP is negative, and moreover existing unit-based fees already largely account for the marginal social costs of waste disposal, so overall net benefits are negative. By contrast, mean WTP is positive and marginal fees are lower in the other municipality, implying generally positive net social benefits of up to €12 per household. In a novel but somewhat speculative exercise, we also use a separate estimate of household effort to isolate the utility purely from receiving feedback, e.g., due to shame or pride: this varies by treatment, ranging from slightly negative to clearly positive. Thus, there appear to be no large "hidden costs of nudging" in our experiments. Overall, feedback nudges stand out as efficient second-best policy where pre-existing marginal incentives do not fully reflect marginal social costs.

The remainder of this paper is organized as follows. In Section 2, we provide some brief institutional background on waste management in Sweden. Section 3 outlines our experimental design, while Section 4 describes our empirical strategy as well as some important features of the data. Section 5 presents main results. Section 6 moves on to various extensions, providing robustness tests and analyzing mechanisms (illicit disposal, prevention, recycling), treatment-effect heterogeneity, and longer-run effects. Section 7 evaluates the costs, benefits, and cost-effectiveness of norm feedback on household waste. Finally, Section 8 concludes the paper.

2. Waste management in Sweden

Swedish national targets for waste management largely derive from EU objectives, with the 2018 revision of the EU Waste Framework Directive requiring each members state to recycle 55% of municipal waste by 2025. Additional targets construct a trajectory where recycling targets increase by five percentage points every five years, up to 65% in 2035. The overall Swedish recycling rate stood at 38.3% in 2020. While some of the target gap is unrelated to waste from households (European Environment Agency, 2022), additional policies are likely to be needed to attain future targets. There are also more specific Swedish targets for packaging, paper, and food-waste recycling, most of which are currently not being met (Swedish Environmental Protection Agency, 2023).

Table 1
Experimental treatments in the two studies.

Municipality	Treatment	N	J
Varberg	1. *Control*: no letters sent	4971	172
	2. *Standard HER*: monthly feedback	4961	172
	3. *Short-run dynamic*: monthly feedback	5003	172
Partille	1. *Control*: no letters sent	1837	55
	2. *Monthly*: "Standard HER" feedback	1838	55
	3. *Quarterly*: "Standard HER" feedback	1844	55

Notes: Table lists treatment conditions in the two studies. For each treatment, N is the number of addresses included in our main data sets, and J is the number of clusters. Thus, there are 172 (55) blocks in Varberg (Partille).

Local waste management rests on a dual system. First, collection and treatment of residual and food waste is left to municipalities, typically being run by public utilities. Some localities have opted for curbside collection and/or unit-based pricing to encourage household recycling; the latter is in use in about 10% of Swedish municipalities. Second, packaging and paper are subject to extended producer responsibility regulations. Collection from single-family homes (the focus of our study) occurs mainly through some 5000 designated "recycling stations" where households may go to drop off packaging and paper waste. The vast majority of stations are run by a single producer-owned corporation, FTI.

Our experiments were conducted in Varberg and Partille, two municipalities in southwest Sweden. Partille is a suburb of the city of Gothenburg, while Varberg has a large central town and is otherwise mostly rural.[9] Both have used weight-based waste fees since the 1990s, and both collect residual and food waste curbside. In both municipalities, waste fees have a fixed as well as a per-unit component. The marginal-cost component remained constant throughout 2019, the year of our intervention; in Partille, it equaled approximately €0.17 per kg in euro terms, for both residual and food waste, while it was about €0.28 in Varberg. Varberg additionally requires households that do not source separate food waste to pay a per-unit surcharge, roughly doubling the per-unit price. In both areas, the variable cost component is displayed separately on all utility bills received by households.

3. Experimental design

We conduct two separate but parallel studies in the Swedish municipalities of Varberg and Partille.[10] In each locality, our study sample includes about 90% of all single-family homes; since household-specific waste weights cannot be identified in apartment buildings, no such addresses are part of either study. This leaves us with about 15,000 households in Varberg and 5000 in Partille.

In both areas, households are divided nearly equally into one control group and two treatments. However, the two treatments differ across municipalities, as shown in Table 1. All experimental interventions involve letters containing accurate and household-specific norm feedback on residual waste. These letters, stamped with the relevant municipal logo, are sent repeatedly to all treatment-group households;

[8] Additionally, there is some worry that higher fees will lead to waste dumping (Fullerton and Kinnaman, 1996; Heller and Vatn, 2017) or leakage to unregulated areas (Bucciol et al., 2015), though most empirical evidence suggests such perverse effects are small to nonexistent (Allers and Hoeben, 2010; Carattini et al., 2018; Bueno and Valente, 2019; Erhardt, 2019). Several authors have noted that recycling is driven by intrinsic motivation in addition to material concerns (e.g., Sterner and Bartelings, 1999; Berglund, 2006; Ferrara and Missios, 2012; Czajkowski et al., 2017). Viscusi et al. (2011) question the importance of social norms for waste behavior, arguing that private values are crucial; in contrast, our results confirm that waste behavior is strongly driven by norm-related concerns.

[9] Recycling stations are typically within walking distance in urban areas but often require car use in rural areas. Among the households in our sample, the median (99th percentile) geodesic distance to the nearest station is 369 meters (1.19 km) in Partille but 1.66 km (11.60 km) in Varberg. However, nearly all rural recycling stations in Varberg are located next to grocery stores, sports fields or other local points of interest, so rural households (which tend to need a car in any case) are likely able to combine dropping off recyclables with other car trips.

[10] Both experiments were pre-registered at the AEA RCT Registry under ID no. AEARCTR-0003301.

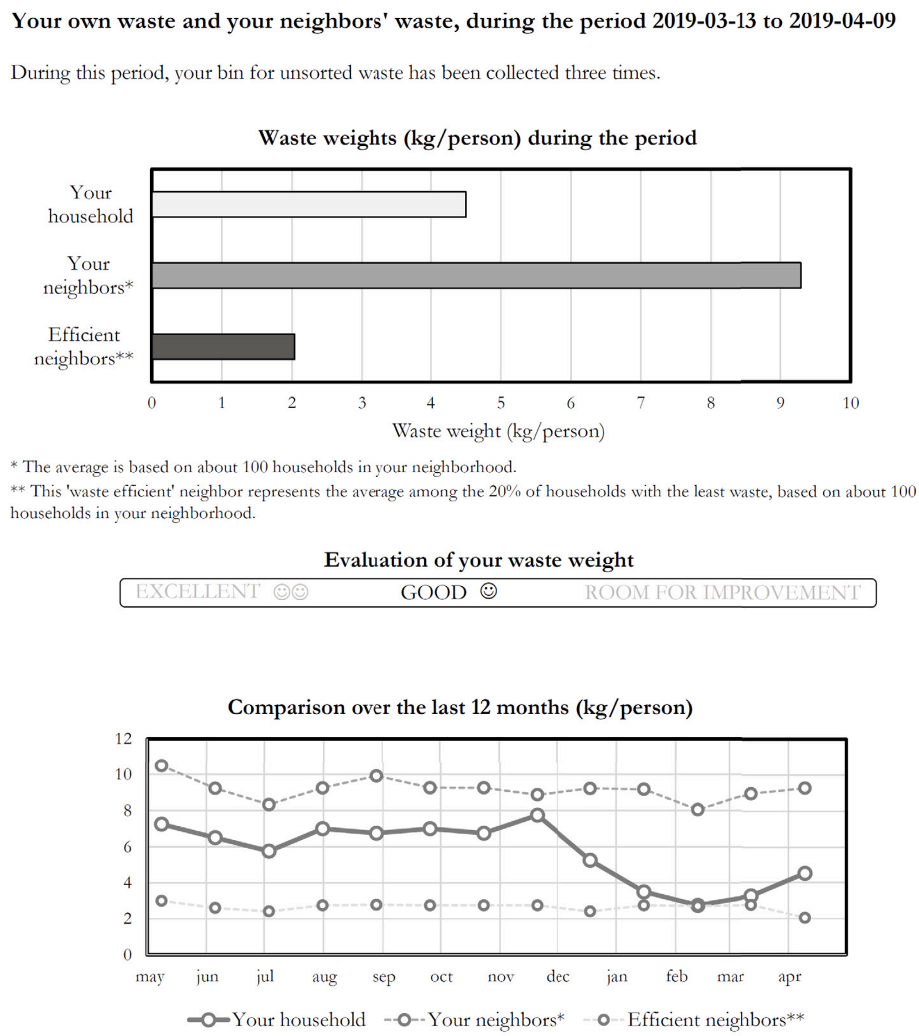


Fig. 1. Translated example of "standard HER" norm feedback.

households in the control groups receive no letters. Both interventions took place during March–October 2019, with the first letters received on 19 March in all groups.

We implement cluster randomization with blocking in both municipalities. Clusters are geographically contiguous groups of addresses that are themselves organized into larger blocks (also contiguous) of exactly three clusters each. Treatment status is randomized across the three clusters in each block such that treatments are perfectly correlated within cluster and each treatment arm is present in all blocks. As shown in Table 1, there are 172 blocks in Varberg, and 55 in Partille. We use cluster randomization to mitigate potential interference between treatment and control households, which might arise if, for instance, immediate neighbors discuss the letters. Evidence of such across-household spillovers is mixed in previous research (Allcott and Rogers, 2014; Dolan and Metcalfe, 2015; Jaime Torres and Carlsson, 2018). Although randomizing in clusters reduces power to some extent, we prefer to err on the side of caution, not least since our experiments apply norm feedback in a new domain. Furthermore, the use of blocking effectively provides stratification by neighborhood, again increasing estimator precision.[11]

The Varberg study varies the feedback type used. We attempt to go beyond standard feedback designs that present cross-sectional household comparisons and year-long time series ("standard HER") to rather highlight how waste behavior has changed since the last letter was received ("short-run dynamic"). In Partille, all households receive standard HER-type feedback, and we instead vary feedback frequency, with one treatment group receiving feedback every four weeks ("monthly"), and the other receiving feedback every twelve weeks ("quarterly"). Households in the monthly condition receive a total of nine feedback letters between March and October 2019, while households in the quarterly condition receive three feedback letters.

Fig. 1 provides an example, translated from Swedish, of the "standard HER" feedback presented to households in Varberg; the full letter page is given in online Appendix A.1. The setup is very similar to earlier studies on HERs such as Allcott and Rogers (2014) or Andor et al. (2020). In monthly conditions, each letter refers specifically to the preceding four weeks; in the quarterly condition, reference periods are the past twelve weeks. For each such period, the bar chart in the upper part of the page displays, top to bottom: (i) the receiving

[11] The clusters and blocks themselves were pre-constructed "by hand" using maps and GIS tools. We explicitly aimed to sort similar housing types into the same blocks; for more information on our randomization methodology, see online Appendix B.1. Online Appendix C presents our main power calculation, based on a variant (Ek, 2020) of the recent (Burlig et al., 2020) serial-correlation-robust procedure, showing that our sample gives at least 80% power to detect a residual-waste reduction of about 2% (4%) in Varberg (Partille). The same minimum detectable effects apply in comparisons between treatments, e.g., between monthly and quarterly feedback.

During the latest period, your waste weight was 0.3 kg/person lower than during the preceding four weeks.

Over the same period, 45 percent of your neighbors have managed to reduce their waste by more than your household.

Fig. 2. Example of a short-run dynamic norm feedback box.

household's summed residual-waste weights per person; (ii) average summed per-person weights within a reference group of roughly 100 households belonging to the same treatment arm; and (iii) average per-person weights within the subset of "waste efficient" neighbors, i.e., households in the bottom 20 percentiles of the reference-period specific weight distribution.[12]

Following standard practice, we add an injunctive component to the bar chart, with the aim of counteracting "boomerang effects", i.e., that efficient households reduce their efforts at the same time that inefficient households increase them (though Schultz et al., 2007 is the only study we are aware of that finds such effects). Below the bar chart, a summary box with three possible outcomes is displayed. First, if a household's weight is above the reference-group average, the assessment "Room for improvement" is displayed, with the other two outcomes greyed out. For weights below the reference average but above the efficient average, "Good" is displayed instead, along with one smiling emoticon. Finally, if the weight falls below even the efficient average, "Excellent" is displayed along with two smileys.

The lower graph shows the evolution of own-household weights as well as reference and efficient averages over the past twelve months. Like the upper chart, this time series is updated with each additional feedback letter. Finally, at the bottom of the page is a link to a municipal web page with more information, including some "frequently asked questions". A translated version of such an FAQ section may be found in online Appendix A.2. Recipients are also informed that the FAQ web page includes a service where they may opt out of receiving letters in the future. Households that do so receive no further feedback letters during the entire intervention period.[13]

By the end of the project, 1453 households had opted out in Varberg, while 186 households had done so in Partille. These figures amount to 14.6% and 5.1% of treated households, respectively; by comparison, in studies of HERs, opt-out rates tend to be less than 1% on average (Allcott, 2015). Some households are likely to view simply being exposed to the letters as a utility cost (Allcott and Kessler, 2019); the high opt-out rates we observe suggest the magnitude of such costs may be domain-specific, i.e., larger than for energy-use feedback. In any case, households that opted out remain included in our final data set, and thus their decision not to participate does not bias our results. However, for privacy reasons, our data do not identify households that opted out, so we are unable to analyze how, e.g., participating and non-participating households differ.[14]

For the short-run dynamic condition, received only by a subset of households in Varberg, we dropped the time-series graph featured in the standard HER feedback letter and instead added a prominently placed text box. This box, shown in Fig. 2 (again, see online Appendix A.1 for the full letter) reports how waste weights have changed over the immediately preceding four-week period. For households that have reduced their weight, the share of neighbors with an even larger reduction is given. Households that increase their waste weight from one period to another receive similar feedback, but with the sentence on neighbor behavior reporting the proportion that have reduced their waste by any amount. Thus, households are always provided with a relevant benchmark for comparison.[15]

Each feedback letter also includes text on the back, with general information on recycling options in the recipient's municipality as well as some specific tips on how to reduce waste (e.g., by planning food purchases or putting a no-ads sticker on the mailbox). This page did not change over the course of the experiment, although there was some variation across the two municipalities. An example back page (for Varberg) is given in online Appendix A.1.

The final feedback letter, sent in late October 2019, informed households that no more letters would be sent and included a link to an endline survey. This information was sent without feedback to addresses in the Partille quarterly condition (which had already received feedback one month before) as well as to control households, but not to households that had opted out. Survey items included questions regarding knowledge of and attitudes to the project, waste behavior over the preceding months, as well as project-related contacts with other households. Translated versions of the survey are included in online Appendix A.3.

4. Data and empirical strategy

Both participating municipalities weigh all waste bins during collection. The resulting weight records contain one line per bin-specific collection event, typically including a non-zero weight measured to a precision of 0.5 kilograms. These events form our main data source.[16]

[12] Other studies have made the comparison with the 20th percentile instead; we believe the average is easier to explain. A second difference is that we did not include information on potential monetary savings.

[13] Overall, there are few major differences between the "standard HER" feedback received by households in Partille and Varberg. The most substantial difference is that, due to municipality concerns regarding public acceptance, households in Partille do not receive a textual evaluation of the bar chart. Valenced feedback is reduced to the use of emoticons at the right end of the bar chart, aligned with the upper (own-household) bar. The number of smileys shown for any outcome is the same as in Varberg, but we do not display the set of possible assessments not given to a household.

[14] We did, however, separate opt-out rates by treatment arm. In Varberg, 13.4% (15.8%) of "standard HER" (short-run dynamic) feedback households opted out. In Partille, 8.3% (1.8%) of monthly (quarterly) feedback households similarly did not participate. Thus, variation appears largely driven by (i) a municipality effect, and (ii) a lower opt-out rate for quarterly feedback.

[15] The "short-run dynamic" letters do retain the (static) bar chart and its associated text-and-emoticon evaluation at the bottom. Similarly, some information on short-run dynamics can be gleaned from the time series included in standard HER letters. It follows that our added treatment does not fully isolate the effect of short-run dynamic feedback as such; nevertheless, compared to the standard HER design, it does put much stronger emphasis on the period-to-period changes that are occurring in the reference population, as in Sparkman and Walton (2017) and Mortensen et al. (2019).

[16] They were also used to compile accurate feedback letters throughout the course of our intervention; for information on how this was done, see online Appendix B.2.

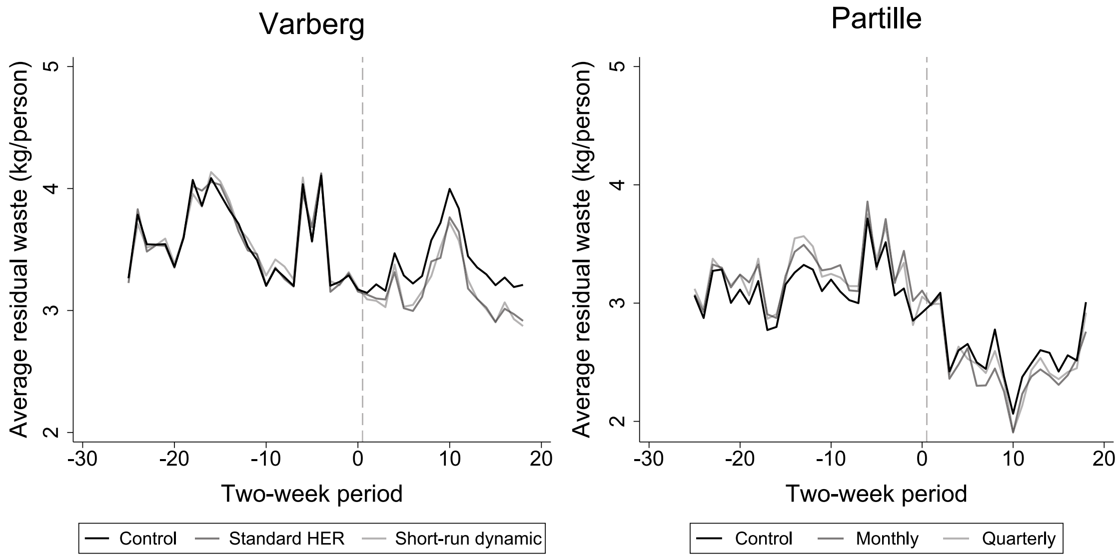


Fig. 3. Residual-waste averages by treatment arm and two-week period.

To construct our main experimental data set, we perform a number of operations on the raw data, as further detailed in online Appendix B.3. Most (or, in Varberg, all) households have biweekly collection cycles, with collection from different households roughly evenly staggered across each cycle. Therefore, we organize our data as an address-by-two-week-period panel. The panel, starting on 19 March, 2018, includes 26 pre-experimental periods ($t \leq 0$), and 18 post-experimental periods ($t \geq 1$). Thus, monthly feedback was received in periods 1, 3, 5, etc; and quarterly feedback was received in periods 1, 7, and 13.[17] The data include all addresses that either received at least one feedback letter, or else are flagged as part of a control cluster (14,935 in Varberg, and 5519 in Partille).

Our main outcome variable y_{ijkt} is the amount of residual waste collected from household i, belonging to cluster j and block k, in period t. Waste is expressed in kg/person by dividing the raw weights (summed within period) by the number of household members as given by register data from the Swedish Tax Authority.[18] In accordance with our pre-analysis plan, certain observations and addresses are considered outliers and are dropped from the data.[19] Our main regression estimates

$$y_{ijkt} = \lambda_{kt} + \beta_1 T_j^1 + \beta_2 T_j^2 + \theta \bar{y}_i^{PRE} + \boldsymbol{\gamma}\mathbf{X}_i + \epsilon_{ijkt} \quad (1)$$

where, since we employ cluster randomization, we consistently cluster robust standard errors at the cluster level (Abadie et al., 2017). λ_{kt} are block by two-week period fixed effects, and $\mathbf{X}_i$ is a set of predetermined address-level controls.

Eq. (1) is an ANCOVA regression, replacing address fixed effects with $\bar{y}_i^{PRE}$, the baseline (periods −25 to 0) average of residual-waste weights for household i. ANCOVA can be viewed as an efficient convex combination of difference-in-differences and an *ex-post* comparison of means across treatment arms. It yields weakly higher precision than either component estimator, with efficiency gains compared to difference-in-differences increasing as serial correlation approaches zero (McKenzie, 2012). ANCOVA regressions are run only on post-treatment observations, allowing the treatment t subscript to be dropped. Our main regressions additionally exclude period 1, when households first received feedback, although results are robust to not doing so.

5. Results

Fig. 3 provides a first look at the experimental results. It tracks average per-person residual-waste weights for each treatment arm and all periods, separately for Varberg (left panel) and Partille (right panel). Vertical lines, placed between period 0 and 1, mark the start of treatment.

To the extent that randomization has successfully eliminated average differences between treated and non-treated units, each set of three lines should coincide throughout the pre-treatment period. Reassuringly, this is clearly the case in Varberg despite some rather pronounced seasonal effects.[20] It is not so apparent in Partille, where treatment is randomized over fewer clusters and outcome balance is correspondingly less likely. However, note that, on either side of the dashed vertical line representing the start of treatment, the relative position of each treatment-arm average is roughly constant over time. Given that pre-treatment trends thus appear reasonably parallel, it may be appropriate to apply difference-in-differences as a secondary identification strategy. We return to this point below.

Next, to the extent that our interventions are effective, we would expect control and treatment averages to diverge beyond each vertical line. Indeed, it seems that this is happening in both municipalities. In Varberg, from about period 1 onward, control averages are consistently above treated averages, suggesting a residual-waste reduction of about the same magnitude in both feedback groups. In fact, since pre-treatment trends essentially coincide, these average treatment effects

[17] The periods run from Monday to Sunday at the end of the following week, and do not coincide with the four-week and twelve-week intervals used for feedback purposes, which always run from a Wednesday to a Tuesday. For example, the initial set of monthly letters was compiled on 13 March, 2019 and covered the period 13 February-12 March, which partially overlaps periods −2 to 0.

[18] For addresses where the tax authority data does not report any household members, values are imputed using the relevant 2019 municipality average for single-family homes from publically available Statistics Sweden data (3.0 persons/household in Partille, 2.7 in Varberg).

[19] Specifically, we exclude (i) all addresses with an average residual or food-waste weight above 15 kg/person; (ii) addresses with >90% missing or zero observations for both residual and food waste, across all periods; and (iii) any single data point with residual or food-waste weight above 50 kg/person. In both municipalities, about 2% of remaining observations (3% of remaining addresses) are dropped as a result, with nearly all exclusions due to condition (i) or (ii). Our results are robust to retaining these observations.

[20] Varberg is a popular domestic summer resort, explaining the peak around periods −18 to −15, at the height of the Swedish summer holidays. Other peaks roughly coincide with national holidays (Easter, Christmas).

Table 2
The effect of treatment on per-person residual waste: ANCOVA.

	Varberg		Partille	
	(1)	(2)	(3)	(4)
Standard HER (monthly)	−0.218***	−0.264***		
	(0.026)	(0.028)		
Short-run dynamic (monthly)	−0.240***	−0.290***		
	(0.028)	(0.030)		
Monthly (standard HER)			−0.236***	−0.227***
			(0.035)	(0.035)
Quarterly (standard HER)			−0.178***	−0.187***
			(0.038)	(0.037)
Baseline waste average	0.761***	0.755***	0.702***	0.707***
	(0.010)	(0.011)	(0.016)	(0.016)
p value, $\beta_1 = \beta_2$	0.466	0.424	0.126	0.277
Block by period FE	Yes	Yes	Yes	Yes
Additional controls	No	Yes	No	Yes
Observations	250,145	216,026	86,430	83,609
Addresses	14,918	12,868	5414	5235
R^2	0.375	0.373	0.393	0.390

Table presents ANCOVA regression estimates for average treatment effects on per-person residual waste. Additional controls are: (i) household size, (ii) age of the oldest member of the household, (iii) gender of the oldest member of the household, (iv) whether the household includes at least one child below five years of age; (v) distance, in kilometers, to the nearest FTI recycling station; and (vi) whether the household's waste collection cycle is two weeks. Robust standard errors clustered at the cluster level reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3
The effect of treatment on per-person residual waste: difference-in-differences.

	Varberg	Partille
Standard HER (monthly)	−0.217***	
	(0.028)	
Short-run dynamic (monthly)	−0.246***	
	(0.029)	
Monthly (standard HER)		−0.270***
		(0.037)
Quarterly (standard HER)		−0.212***
		(0.038)
p value, $\beta_1 = \beta_2$	0.371	0.138
Block by period FE	Yes	Yes
Address FE	Yes	Yes
Additional controls	No	No
Observations	629,469	217,965
Addresses	14,935	5519
R^2	0.490	0.518
Within R^2	0.000	0.001

Table presents regression difference-in-differences estimates for average treatment effects on per-person residual waste. Within R^2 relates to remaining variation after absorbing both address and block-by-period fixed effects. Robust standard errors clustered at the cluster level reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

will be roughly equal to the gap between the lines, suggesting a reduction on the order of 0.25 kg/person in both conditions. In relative terms, this is about 7% of the post-treatment control average of roughly 3.40 kg/person. The effect is substantially larger in magnitude than typically found in the literature on Home Energy Reports (e.g., Allcott, 2011, 2015). In Partille, applying differences-in-differences reasoning, we note that control averages are slightly below monthly and quarterly-group averages up until the start of treatment, and are consistently above thereafter. That pattern again suggests a negative-sign treatment effect, although the magnitude of the effect is less immediately clear than in Varberg.

Table 2 presents ANCOVA regression results. The regression in column 1 corresponds to Eq. (1) absent covariate vector $\mathbf{X}_i$, confirming a waste reduction of 0.2–0.25 kg/person from both treatments in Varberg. At the bottom of the table, we also report p values for the test that both treatment effects are equal in magnitude. Clearly, this null hypothesis cannot be rejected, reflecting the fact that our design is likely underpowered to detect such small effect differences (see footnote 11).

Then, in column 2, we add an additional set of household characteristics at baseline, i.e., immediately before the first letter was received.[21] With these covariates, treatment-effect estimates are very similar to column 1, indeed slightly larger at 0.25–0.30 kg/person. However, the sample is effectively selected on non-missing covariate data; when we run the regression specification of column 1 on the subsample where covariates are available, we obtain estimates nearly identical to those in column 2. For Partille (columns 3 and 4), ANCOVA estimates for the monthly treatment are similar to either Varberg intervention, both with and without added covariates. Point estimates for quarterly feedback are somewhat smaller than for monthly feedback, though the difference is not significant. With the municipality control post-treatment average at about 2.55 kg/person, these estimates correspond to a decrease of 7%–9%.[22]

Given the lack of pretreatment outcome balance in Partille, in Table 3 we run regressions like those in column 1 and 3 of Table 2 using difference-in-differences in place of ANCOVA.[23] Note that predetermined controls are invariant within household and thus cannot be used with difference-in-differences. Online Appendix E.1 supports these regressions by testing the parallel-trend assumption through a series of placebo treatment tests, counterfactually assuming that interventions had begun at various points throughout the actual pretreatment period. These regressions confirm our earlier conjecture that pre-treatment trends are roughly parallel in both municipalities. Thus, for Partille, the difference-in-difference analysis in Table 3 is our preferred specification. In Varberg, results in Table 3 are nearly identical to the earlier ANCOVA estimates. Point estimates for Partille are slightly larger, at about 0.25 kg/person.

To put these seemingly large reductions into perspective, it is useful to examine what fee increase might have produced similar effects if applied throughout the post-experimental period. In a synthetic-control study of Italian unit-based pricing, Bueno and Valente (2019) conclude that a €0.09 per liter volume-based fee reduces unsorted waste by 37.5%, a percentage effect size 4–5 times larger than ours. Assuming a conversion factor of 0.2 kg/liter of residual waste,[24] the Bueno

[21] These are: (i) household size, (ii) age of the oldest member of the household, (iii) gender of the oldest member of the household, (iv) whether the household includes at least one child below five years of age; (v) distance, in kilometers, to the nearest FTI recycling station; and (vi) whether the household's waste collection cycle is two weeks. In Partille, about 90% of households in the data have two-week collection cycles, while in Varberg, the figure is exactly 100%, so this covariate is not added there. Our results are robust to simply dropping all households with collection cycles not equal to two weeks.

[22] Beyond average effects, a few HER studies (Allcott, 2011; Costa and Kahn, 2013) have also evaluated the causal effect of injunctive norm content, e.g., of receiving the rating "Good" versus "Room for improvement" (one or no smiley). Since these ratings are based on the cutoff criterion that a household lies at or below the relevant reference-group average, (sharp) regression discontinuity analysis may be applied around the cutoff. In online Appendix D, we do so, finding all estimates to be nonsignificant, except one: in Varberg, there is a positive effect of about 0.2 kg/person from being labeled "Excellent" rather than "Good" ($p < 0.01$). Previous studies have consistently found null results, e.g. Allcott (2011).

[23] All results in both Tables 2 and 3 are robust to applying a Bonferroni correction for multiple hypothesis testing within regression, adjusting the critical values of the two treatment coefficients as well as that of the $\beta_1 = \beta_2$ test (i.e., $m = 3$).

[24] This conversion factor, also used by Dijkgraaf and Gradus (2004), is supported in the raw waste data. We interpret the upper end of the distribution of weights (in kg) collected from a given type of waste bin as an approximation of its capacity. Since the data also list each bin type's volume, we can calculate kg/l factors by dividing, e.g., the 99th percentile of collected weights by the volume. This consistently yields conversion factors around 0.2.

and Valente (2019) fee translates into €0.45/kg. Thus, as a rough estimate of the equivalent fee increase, we simply divide €0.45/kg by four or five, yielding an increase of about 32%–40% of the current Varberg unit-based fee (€0.28), and 54%–67% of the Partille fee (€0.17). The effects we observe thus translate into quite large price increases which, strikingly, also exceed the equivalent price increase of 11%–20% reported for electricity by Allcott (2011).

6. Extensions

6.1. Robustness tests

The linear-in-parameters regressions presented in the previous section fail to account for two potentially important features of our data. First, about 15% of all weight observations in both municipalities are equal to zero. Although such corner solutions may be due to stringent recycling efforts, they are perhaps more likely the result of factors not directly related to waste behavior, such as time spent away from the household. Second, the distribution of weights is highly right-skewed; indeed, viewed in a histogram, the empirical distribution essentially decreases monotonically for all positive weights and features a long right tail. This suggests that it may be more appropriate to model the conditional mean as exponential in the covariates.

In online Appendix H, therefore, we perform a robustness test accounting for both of these features by estimating the lognormal hurdle model of Cragg (1971). Unlike a standard Type I Tobit model, this approach has the benefit of assuming separate variables and/or coefficients driving corner solutions compared to weight choices conditional on weights being strictly positive. It also assumes an exponential rather than linear model for the interior outcomes. The resulting treatment coefficients (online Appendix Table H.1) are comparable but somewhat larger than in Tables 2 and 3, exceeding 0.3 kg/person in some cases and implying reductions of up to 12% compared to the control group. Additionally, treatment-effect differences between monthly and quarterly feedback in Partille are now significant.

Online Appendix H also presents several other variants of the analysis in Tables 2 and 3: (i) in Table H.2, we do not control for block status in any way, instead including only period fixed effects; (ii) in Table H.3, we collapse the data at the cluster level, using clusters as our unit of analysis and cross-sectional cluster averages of residual waste as outcome variable; (iii) in Table H.4, we pool both municipal data sets and run difference-in-differences regressions that include three treatment variables, where one represents monthly HER-standard feedback in both municipalities; and (iv) in Table H.5, we re-run our preferred regressions using residual waste per *household* as outcome variable. All four exercises yield results that confirm those already reported.

A different issue concerns potential spillovers between treatment and control, which will typically bias effect estimates toward zero. For example, members of control households may hear about the feedback letters from receiving neighbors, possibly motivating them to reduce their own waste. However, in our endline survey (full responses are given in online Appendix Table A.1), only about 5% of control respondents claim to both be aware of the project and to have discussed it with others, suggesting that spillovers are not a major issue in our data. However, among those who have discussed the letters, most did so with someone other than immediate neighbors.[25]

6.2. Mechanisms

What strategies do households use to reduce residual waste in response to feedback? Generally speaking, there are three options available to households. First, as already noted, they may be turning to illicit disposal, i.e., dumping. Second, they might increase their sorting efforts, thus diverting waste from the residual bin to various recyclable streams. Third, they may reduce the amount of waste generated, for example by buying more packaging-free products. Quantitative analysis is complicated by the fact that, as in the wider economic literature on waste management, little reliable data is available for any of these three waste-reduction categories. Nevertheless, we will discuss each mechanism in turn.

6.2.1. Illicit disposal

The Partille survey included an item on dumping, asking respondents whether they thought the letters had "made any of their neighbors dispose of their waste in an illegal way": only about 2.5% of respondents thought so. We are able to complement the survey data by accessing municipal records on dumping incidents related to household waste. These necessarily represent a partial measure, since some types of dumping (e.g., in lakes) are unobservable in the short run. Nevertheless, we would expect any substantial effect on dumping to show up in the records.

In Partille, dumping data are available for 2018 (6 incidents) and the intervention year of 2019 (7 incidents), suggesting no major treatment effect on illicit disposal. Since illicit disposal might also occur across the border of small municipalities like Partille, we also check for dumping incidents in neighboring Härryda municipality, where monthly data on dumping of household waste are available from 2015 through to late 2020. Incident frequency is increasing prior to 2019, so we add a linear time trend in addition to a dummy that equals one from March 2019 onward. The dummy is non-significant ($p = 0.896$), a result which does not change when instead we "switch off" the dummy after our intervention concluded in October 2019.

In Varberg, municipal records exist from 2015 onward and are disaggregated by waste stream, allowing us to consider dumping of household waste separately from waste types not targeted by our intervention, such as chemicals, scrap vehicles, and building materials. We do find a spike in dumping incidents related to household waste in 2019 (8 instances, compared to 1–4 during 2015–2018). However, a similar increase appears in 2019 for non-household waste (9 incidents, compared with 2–6 in earlier years), suggesting the variation is unrelated to our experiment. In any case, as in Partille, all incident numbers are clearly extremely small in relation to the number of treated households, so effects on dumping (if any) are likely very minor.

6.2.2. Recycling and prevention

Given that illicit disposal can arguably be ruled out as a mechanism, we now turn to recycling of waste. Recyclables include food, paper, and packaging waste. Starting with food waste, this waste type is collected by our partner utilities and is available in our main data sets. Thus, Fig. 4 depicts raw time series for collected food waste (in kg/person). Unlike in Fig. 3, average pre-treatment weights appear roughly to coincide for all treatment arms in either panel, suggesting ANCOVA regressions may be run in both municipalities. The averages also appear to diverge in the post-treatment period, although the effect is much less pronounced than found in Fig. 3 for residual waste; note that increased recycling translates into more food waste being collected. Online Appendix Table H.7 shows the corresponding ANCOVA

[25] Some caution is advised in interpreting these results, given that response rates are generally low: only about 5% (10%–18%) of treated (control) households. However, in online Appendix Table H.6, we also check for spillovers directly in the data. For Partille, we add a dummy for the 597 control households that are directly adjacent to some treated household(s), thus capturing the differential effect compared to control households that do not border on treated clusters. For Varberg, we instead interact both treatment variables with a dummy for whether a given block is in a rural area (68 out of 172 blocks). The idea is that since houses are spaced further apart in rural areas, spillovers are perhaps less likely to occur and thus treatment estimates should exhibit less bias there. We find no indication of spillovers in either model.

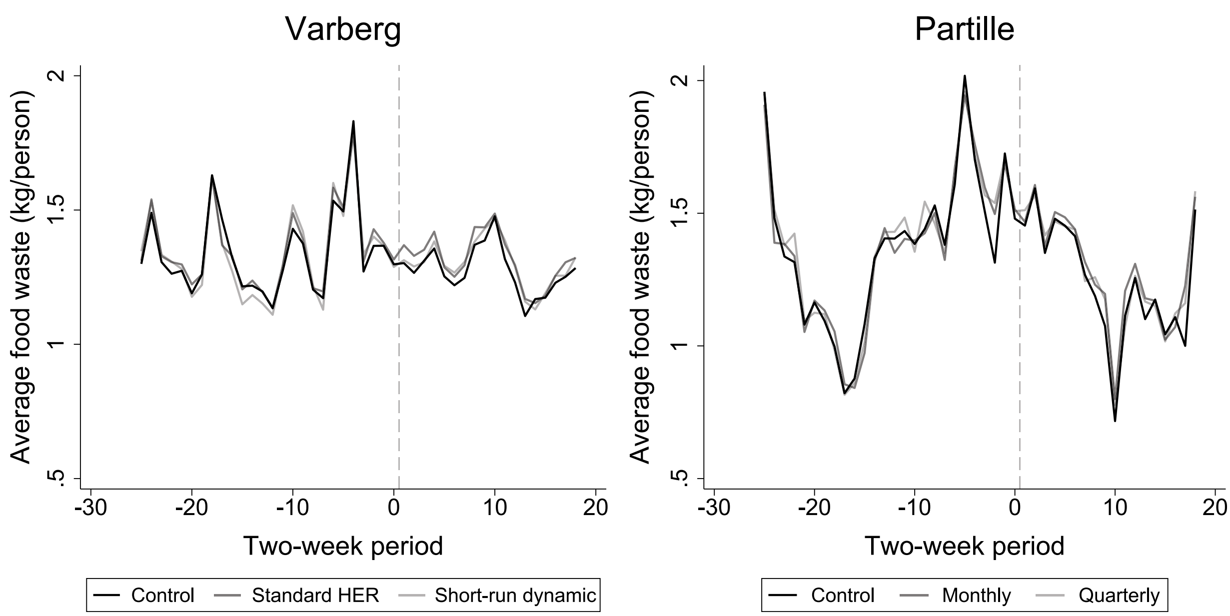


Fig. 4. Food-waste averages by treatment arm and two-week period.

Table 4
Post-treatment composition of residual waste.

	Food	Paper & packaging	Other, avoidable	Other	Sum
Control ($N = 238$)					
Weight (kg)	165.5	200.2	43.1	99.1	507.9
Weight share (%)	32.6	39.4	8.5	19.5	100.0
y_{it} (kg/person)	0.831	1.005	0.216	0.498	2.550
Pooled treatment ($N = 423$)					
Weight (kg)	147.6	192.2	71.2	110.1	521.1
Weight share (%)	28.3	36.9	13.7	21.1	100.0
y_{it} (kg/capita)	0.654	0.852	0.315	0.488	2.309
Reduction: y_{it} (kg/person)	0.177	0.154	−0.099	0.010	0.241

Table presents results from a bin audit on a sample of 661 households in Partille. Rows "y_{it}" estimate weights in kg/person for control and pooled treatment as a whole, multiplying the weight shares in the audit data by either the control-group post-treatment mean (2.550 kg/capita), or by the same value less the average of the Partille ATE point estimates in column 2 of Table 3 (0.241). Row "Reduction" gives the resulting stream-specific differences across control and pooled treatment.

regression estimates, which are consistent with the figure: except for an insignificant and near-zero coefficient for the quarterly treatment, ATEs on food waste cluster around 0.03 kg/person, about one eighth of the reduction in residual waste.[26] We conclude that most of that reduction must be due to other mechanisms.

Paper and packaging are not available in our data and thus cannot be analyzed in the same way.[27] Instead, we perform a bin audit to check the composition of residual waste in participating households. The idea is that learning how the post-experimental content of residual waste differs across control and treated households allows us, at least in principle, to infer recycling and prevention behavior for all waste streams. For example, a lower share of packaging thrown in the residual bin suggests that recycling and/or prevention of packaging waste has increased. We carried out such an analysis in late November 2019, less than one month after the final feedback letter was sent.

[26] For additional robustness, we run difference-in-differences regressions in online Appendix Table H.8 (with supporting placebo regressions in online Appendix E.2); the results are very similar to those in online Appendix Table H.7.

[27] Recall from Section 2 that these waste types are not collected curbside; instead, households dispose of them at designated recycling stations. While the producer-owned corporation responsible for station maintenance does record the amount of collected paper and packaging by municipality and year, these data do not allow us to estimate experimental impacts.

The procedure was the following. First, a contractor collected all residual waste generated during a single waste cycle from a (sub)sample of 661 participant households in Partille. The sample was nonrandom but included all single-family households in a particular area within the municipality, and thus was split roughly equally across treatment arms. Two sets of separate but nearly concurrent collection runs were made: one for control households, and one for both treatments. Once all waste had been collected, a random sample of about 500 kilograms (10%–20% of collected waste, depending on the group) was made from the waste totals of each collection run. The composition analysis is on that subsample.

The sampled weight of each waste type is presented in Table 4. We also calculate the weight proportions of each waste stream. Next, supposing those proportions applied to all participating addresses in Partille, we multiply them either by the control-group post-experimental mean of 2.55 kg/capita (control), or by the same value less the average of the treatment-effect estimates given in column 2 of Table 3, which is 0.241 (pooled treatment). As a result, post-treatment residual waste is decomposed by waste stream in the rows labeled "y_{it}". Finally, we compare the results across control and pooled treatment. Note that we are unable to distinguish recycling from prevention: reductions in Table 4 may be due to either or both mechanisms.

Based on this composition analysis, treated households throw 0.177 kg/person less food waste in the unsorted bin than do control households. This is a much larger reduction than the corresponding increase in sorted food waste found in Fig. 4 and online Appendix Table H.7. Several explanations are possible. First, as noted, some of the reduction computed in Table 4 may be due to prevention and will therefore not show up as increases in sorted food waste. Second, the analysis may not be representative of Partille as a whole, due to sampling and/or measurement error. Third, the bin audit treated unopened packaged food as part of the food-waste stream, but the packaging itself would obviously not add to sorted food waste if such items were correctly source separated. In any case, the reduction in paper and packaging waste is roughly equal to that in food waste, suggesting households respond to treatment by reducing both streams more or less equally. These reductions are somewhat offset by *increases* in the "other, avoidable" stream, including for instance organic (garden) waste, cloth, utensils, etc.

Our endline survey, finally, may also shed light on the relative contribution of recycling and prevention across various streams. In panel B of online Appendix Table A.1, we ask control and treated households whether their waste behavior had changed during the intervention period; and if so, what the most substantial change was.[28] Here, nearly half of households state increased recycling of paper and packaging as their main response to treatment; though stated prevention behavior, while rarer, is also not negligible in our sample.[29]

6.3. Heterogeneous treatment effects

A general finding in the existing literature on norm feedback, whether with respect to water or energy usage, is that treatment effects are largely driven by high users (see e.g., Allcott, 2011; Ferraro and Price, 2013; Jaime Torres and Carlsson, 2018; Andor et al., 2020). Does such a pattern hold within the waste domain as well? To check this, we interact both treatment variables with indicators for each household's position in the distribution of average pre-experimental ($t \leq 0$) waste weights.[30]

For each municipality and treatment, Fig. 5 plots the resulting decile treatment effects and associated confidence intervals. Estimates follow the expected pattern, with more pronounced effects among households with high baseline generation. Indeed, reductions in the highest deciles are about 0.5 kg/person (15%–20%), twice as large as our estimated ATEs. At the lower end of the baseline distribution, effects instead approach zero, so there is little evidence of a "boomerang" effect such that low-decile households generate more waste when treated (Schultz et al., 2007).

6.4. Longer-run effects

In early 2021, we obtained an additional batch of waste data from our partner municipalities, allowing us to examine longer-run effects in the year after the intervention concluded. The new data set extends to November 2020 and starts in late November 2019 at the exact endpoint of the main data explored thus far (effectively adding post-experimental periods 19–44).

Most HER studies have found that treatment effects are remarkably persistent for both energy (Allcott and Rogers, 2014; Alberts et al., 2016) and water use (Ferraro et al., 2011; Bernedo et al., 2014), often with more than half of the original effect remaining after a year or more. The pattern of immediate change and slow reversion is consistent with some combination of changes in habits and capital stock (Allcott and Rogers, 2014). For example, Brandon et al. (2017) exploit the fact that in the Opower HER experiments, letters were discontinued upon the sale of a home and not sent to the incoming household. Thus, any residual treatment effect occurring after sale seems likely to reflect changes in physical capital, and Brandon et al. (2017) estimate that this channel accounts for 35%–55% of the total effect.[31]

Physical capital plays a different role in the waste domain, however. Waste-related physical capital investments such as in-home recycling bins or "no-ads" stickers are unlikely to mechanically persist after home sales, limiting the scope for identifying the two channels through home sales. More importantly, effort and physical capital are complements rather than substitutes, as stressed by Vollaard and van Soest (2024). As a result, both short-run and longer-run effects will mostly reflect behavior change, with physical capital improvements reduced to a multiplicative effect.[32]

With these points in mind, we examine longer-run effects from all four treatments in Fig. 6. We run a single regression in each municipality. These interact both relevant treatment variables with all periods except $t = 0$, allowing us to track effect sizes over time. Thus, in the figure, both sets of estimates within a given municipality derive from the same regression, and are presented in separate subfigures purely for clarity. The Varberg estimates (panel a), being derived from an ANCOVA specification, use only post-treatment periods, while the Partille difference-in-difference regression (panel b) performs a full event study that includes all periods. Dashed gray lines mark the start and end of our original post-treatment period. Both regressions include block-by-period fixed effects but no other covariates (except for baseline averages in Varberg).

Strikingly, we observe quite limited reversion-to-zero effects over the course of the additional year. In Partille, such backsliding is practically nonexistent; in Varberg about one quarter (half) of the "standard HER" (short-run dynamic) norm effect is lost. The patterns are consistent with norm feedback inducing a change in highly sticky, habitual behavior, and are confirmed in additional regressions where treatment status is interacted with a set of longer durations of about six periods each (online Appendix Table H.9).

[28] Simple two-proportions tests confirm that changed waste practices are more common among treated respondents ($p = 0.037$ in Varberg, $p < 0.001$ in Partille). By contrast, chi-square homogeneity tests fail to reject the null hypothesis that the distribution of specific changes made is equal across treatment and control. Of course, a potential issue in these comparisons is the fact that response rates are low, and substantially lower in treated households compared to control (online Appendix Table A.1).

[29] For Partille, we are able to support this final point with additional evidence: in July and August 2020, nine months after the last letter was sent, we had research assistants check for "no-ads" stickers on the mailboxes of all 2398 single-family homes (split roughly equally across treatment arms) in three specific areas of the municipality. Sticker rates were found to be about 6.4 percentage points higher in the monthly group (two-proportions test: $p = 0.011$), and 4.1 percentage points higher in the quarterly group ($p = 0.101$), compared to control households. We interpret this as evidence of prevention of paper waste, though the amount of waste thus avoided is unknown. Overall longer-run effects are discussed in Section 6.6.

[30] For Varberg, the specification used is the modified ANCOVA regression

$$y_{ijkt} = q_i^{dec} \times (\lambda_{kt} + \beta_1 T_j^1 + \beta_2 T_j^2) + \theta \bar{y}_i^{PRE} + \epsilon_{ijkt}$$

where q_i^{dec} represents a full set of dummy variables for the baseline waste-weight decile to which address i belongs. Thus, we vary both treatment effects and block-specific time trends by decile. For Partille, where our preferred model is a difference-in-differences regression, we retain all pre-treatment periods in the data and run a specification like the above, except with time-varying treatment status and with address fixed effects in place of baseline averages.

[31] Note that these figures apply to the wider sample only if movers are representative in terms of their chosen combination of habit formation and capital investment; in other words, if there is no selection into being a mover.

[32] In principle, we might use changes in customer IDs to infer where moves have occurred (as in Ferraro and Miranda, 2013; Bernedo et al., 2014). Our Varberg data do not include such IDs, however, making it infeasible to correctly flag movers without close examination of personal data (e.g., name changes). In any case, our best estimate is that, in both municipalities combined, no more than about 1000 households moved at some point in the "longer-run" periods 19–44, implying that our power to decompose effects into different mechanisms is likely quite low.

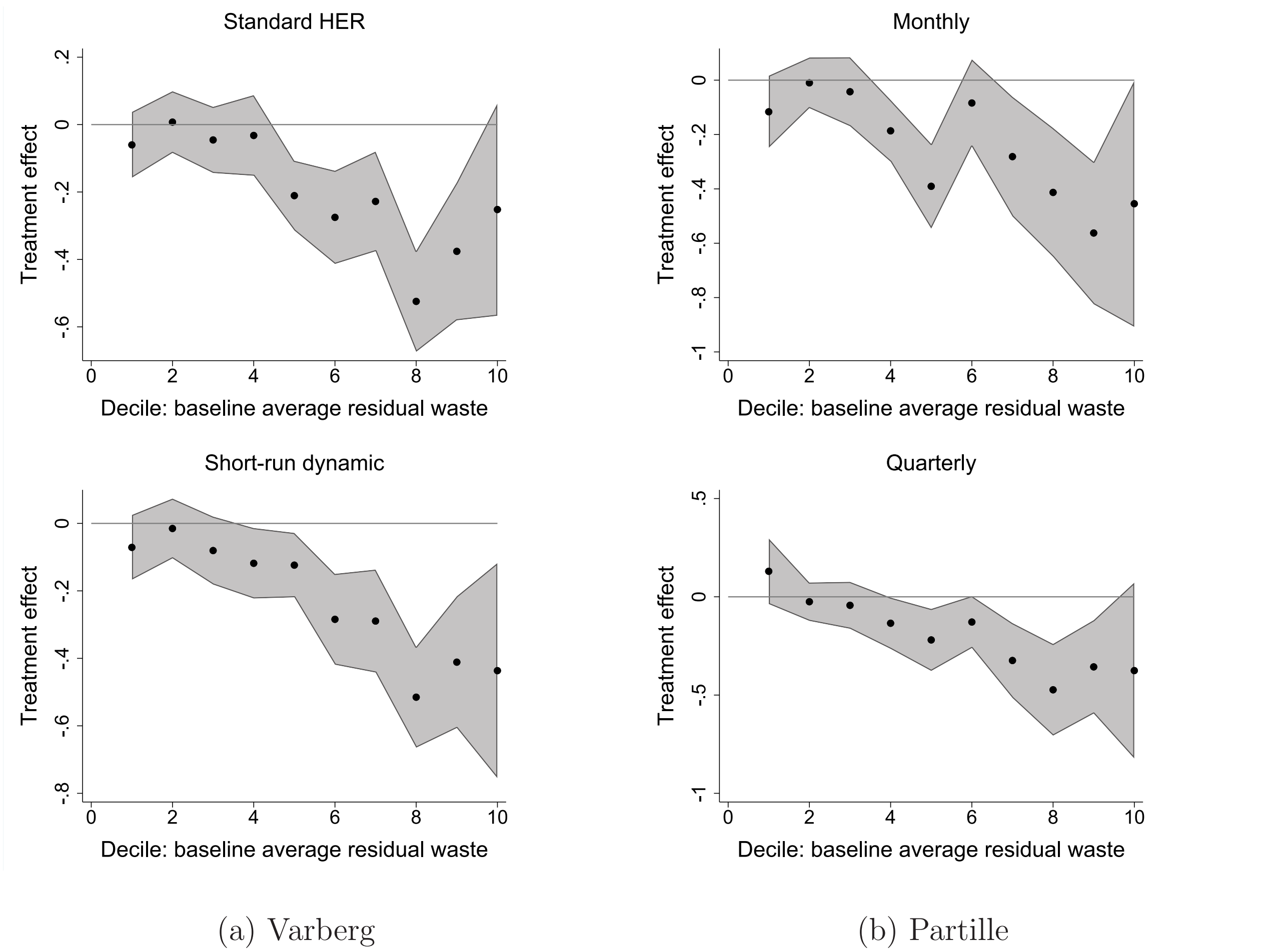


Fig. 5. Heterogeneous treatment effects by baseline waste generation: point estimates by decile, with 95% confidence intervals.

7. Costs, benefits, and effectiveness

In this section, we summarize a cost–benefit analysis of our intervention, presented in full in online Appendix F. We use Swedish data wherever possible. A major component of this analysis is data on household WTP for (not) receiving additional feedback letters. In November 2023, we sent a valuation survey to all Varberg and Partille addresses that were treated in our 2019 experiments. We obtained 3718 usable responses at a response rate of about 20%. There were four survey variants, each presenting respondents with an example of one of the four 2019 feedback treatments and then eliciting their WTP to receive that feedback type over a period of equal length to that of the original intervention (i.e., nine letters in "monthly" treatments, and three letters in the "quarterly" treatment), starting in early 2024. For valuation, we used a variant of the Becker–DeGroot–Marschak (BDM) similar to Butera et al. (2022).[33] The valuation survey is presented in online Appendix A.5.

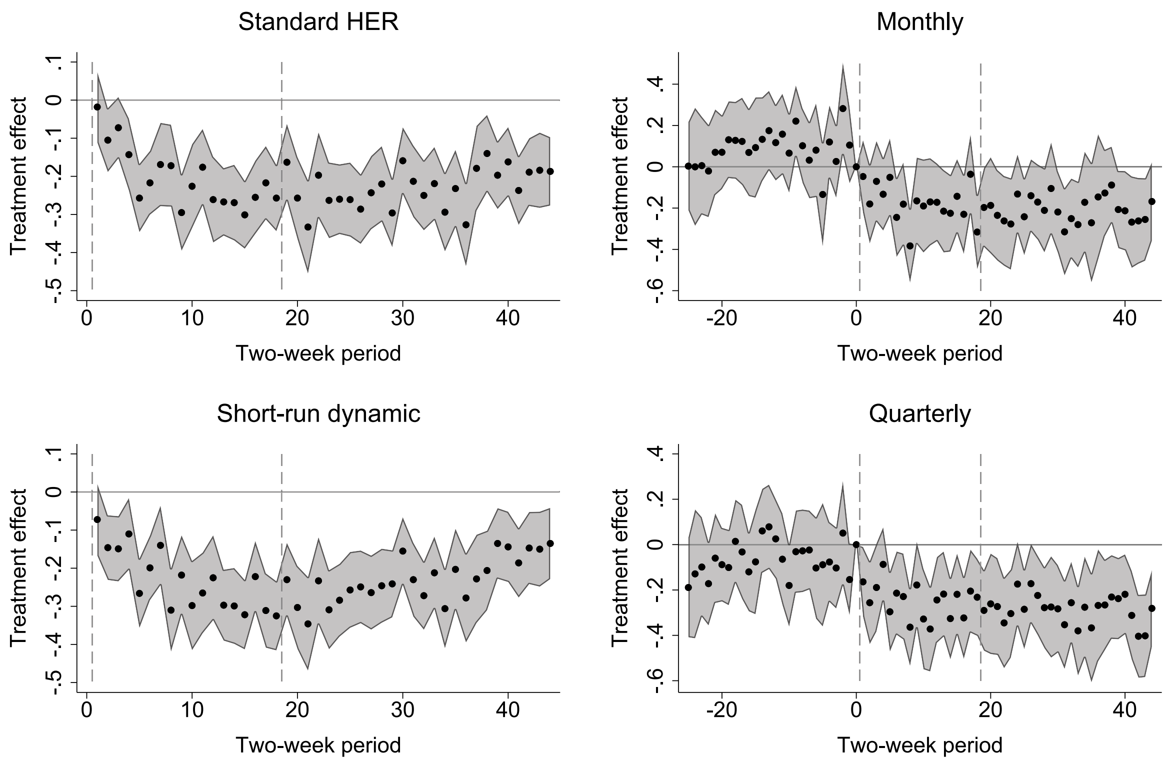
Allcott and Kessler (2019) conduct a similar valuation of specifically the second year of an Opower HER intervention. By contrast, we interpret our results as the net benefits of an intervention's initial 1.7 years, corresponding to the longer-run data period in Section 6.4. Our key assumption is that WTP is consistent with a scenario where feedback letters are again sent to all eligible households in Varberg and Partille, e.g., from early 2024. Since this is more than four years since our original treatments concluded (October 2019), we assume that any lingering effects have then fully dissipated and that the hypothetical intervention would have effects identical to those that can be estimated from our experimental data.[34] On the other hand, it is likely that many survey respondents still remember what it was like to receive the 2019 letters and thus are able to provide an accurate valuation. Indeed, we maximize prior familiarity with each feedback type by targeting each survey variant specifically to the addresses included in the corresponding 2019 treatment. Thus, our cost–benefit analysis is for that hypothetical intervention.

Fig. 7 illustrates the structure of the analysis. The horizontal axis measures the amount of residual waste generated by a representative household over the entire course of the intervention. The initial demand curve D_0, assumed linear for simplicity, reflects the household's marginal benefit of generating unsorted waste, or equivalently, marginal costs of reducing it. The presence of unit-based pricing implies pre-existing incentive ϕ.

Our intervention works by shifting the demand curve downward to D_1; the shift is parallel by assumption. Since marginal incentives remain at ϕ, the result is the ATE shown in the figure. This ATE has several welfare implications. First, private and external costs from the

[33] The first batch of letters, stamped with the logo of the University of Gothenburg rather than a municipal utility, were sent in April 2024. All payments to respondents were made through a Swedish payment service called Swish, used by some 90% of the Swedish population.

[34] The length of the hypothetical intervention is otherwise comparable to Allcott and Kessler (2019), who consider a set of four bimonthly letters. Online Appendix F.4.2 presents a linear extrapolation exercise that, for Varberg, shows that effects fully revert to zero about four and a half years after the start of the 2019 treatment, i.e., by Autumn 2023.



(a) Varberg (b) Partille

Fig. 6. Longer-run effects from treatment: period-by-period point estimates and 95% confidence intervals.

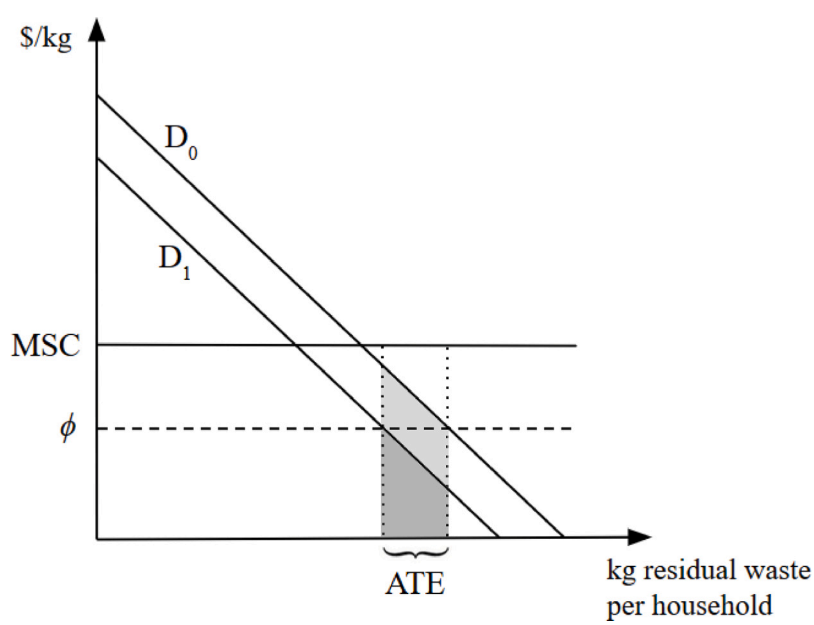


Fig. 7. The behavioral and welfare effects of norm-based feedback.

collection and incineration of residual waste drop. Also, private costs and external benefits associated with the collection, treatment and recycling of food and packaging waste increase. The sum of all these marginal costs and benefits, expressed per kilogram of reduced residual waste, is shown in Fig. 7 as the marginal social cost (MSC) of residual waste. Again for simplicity, we draw the line as horizontal. We also place it above marginal incentive ϕ, implying that a net market failure remains and that recycling rates are initially suboptimally low. Whether that is the case for our actual setting remains to be seen.

Furthermore, as residual waste drops, households experience various costs and benefits. We assume that our WTP measures reflect the sum of three such components. First, in response to treatment, households bear higher abatement costs through some combination of effort and capital investments. In the figure, these costs are given by a trapezoid area below the demand curve, with width equal to the

Table 5
Summary of costs and benefits, by treatment.

	Varberg		Partille	
	Standard HER	Short-run dynamic	Monthly	Quarterly
ϕ (€/kg)	0.28	0.28	0.17	0.17
MSC, low (€/kg)	−0.34	−0.34	−0.34	−0.34
MSC, high (€/kg)	0.31	0.31	0.31	0.31
Direct intervention costs (€/household)	4.57	4.57	4.57	1.52
A. Net benefits, full population, high MSC:				
Mean WTP (€/household)	−2.69	−2.38	1.61	2.57
ATE (kg/household, all periods)	26.33	27.57	38.14	27.74
Effect on net market failure (€/household)	0.83	0.87	5.39	3.92
Net benefits (€/household)	−6.43	−6.09	2.43	4.97
B. Utility of receiving feedback, full population:				
Fee savings (€/household)	5.53	5.79	4.86	3.54
Area below D_0 (€/household)	8.40	8.85	9.08	6.09
Area below D_1 (€/household)	6.34	6.59	3.89	3.34
Utility of receiving feedback, D_0 (€/household)	0.18	0.68	5.82	5.12
Utility of receiving feedback, D_1 (€/household)	−1.87	1.58	0.63	2.38
C. Net benefits, full population, extrapolated effect:				
Extrapolated ATE (kg/household, all periods)	47.24	33.87	105.09	81.36
Number of post-intervention periods	120	72	120	120
Effect on net market failure (€/household)	1.49	1.06	14.86	11.51
Net benefits (€/household)	−5.78	−5.89	11.90	12.55

ATE. However, as argued by Allcott and Kessler (2019), it is not clear whether the area lies below D_0 or below D_1. If a feedback intervention works exclusively by addressing an internality, i.e., by providing information or correcting recipient biases regarding consumption utility, then household disutility from reduced consumption would reflect only the shaded area below D_1. In the waste domain, such an internality might be caused by sticky habits causing households to remain at suboptimally low recycling effort. Conversely, if feedback does not correct an internality but operates exclusively through e.g. moral incentives, then the disutility instead reflects the (light and dark) shaded region below D_0. Second, by reducing residual waste, household also avoid fee payments. The municipalities in our setting apply UBP to both residual and food waste, but to the extent that households reduce residual waste without a corresponding increase in sorted food waste, the avoided fee payments are given as the rectangle with width ATE below the ϕ line. Third and finally, there may be a household net (dis)utility, not shown in the figure, purely from receiving feedback, e.g. from experiencing shame or pride.

We then calculate net social benefits as the sum of mean household WTP, the reduction in the net market failure (the rectangle between the MSC and ϕ lines; note that fee payments are transfers but included in mean WTP), and direct project administration costs of our intervention, also not shown in the figure.

Table 5 presents our estimates of the components of the analysis, organized by municipality and treatment (survey variant). Since our main focus is on the full population that would be targeted in a repeated feedback intervention, we apply inverse probability weighting to the raw WTP data to account for survey-respondent selection on covariates.[35] At the top of the table, we report general parameters. Here, the most complex task is to compute the MSC; this is done in online Appendix G by combining results from life cycle analyses (Ambell et al., 2010; Slorach et al., 2019) with environmental shadow prices (Ahlroth and Finnveden, 2011).[36] Reflecting a range of estimates from valuation studies, environmental shadow prices may take either a low or a high value. As a result, the MSC of residual waste may similarly be high or low. The two estimates are very different, mostly because high and low environmental benefits from recycling of packaging differ by a factor of 22. Our analysis focuses on the high MSC value which, at €0.31/kg, is slightly larger than the pre-existing incentive ϕ in Varberg and nearly double the rate used in Partille.[37]

Net benefits are calculated in panel A of Table 5. First, we report mean WTP, re-weighted to the full population: WTP is generally higher in Partille than in Varberg.[38] Next, to compute ATEs, we sum the longer-run estimates in online Appendix Table H.9 across all 44 data periods after the first letter was sent. For the Varberg Standard HER treatment, for instance, this total effect equals 26.33 kg/household. As a result, the effect on the net market failure is $(0.31-0.28)\times 26.33 =$ €0.83 for a representative household in this treatment. Finally, we subtract direct intervention costs as given by the project budget, yielding net social benefits as, e.g., $-2.69+0.83-4.57 =$ €−6.43/household.

Much of the variation in the resulting net benefits is clearly explained by two facts: mean WTP is lower in Varberg than in Partille, and marginal social costs are already largely internalized in Varberg, with $\phi \approx MSC$. Thus, unlike in Partille, even low-cost interventions to reduce residual waste further are difficult to justify in Varberg based on economic costs and benefits. Notably, the smaller ATE from quarterly feedback is more than compensated for by lower operation costs, suggesting that less frequent feedback is more efficient.

In panel B of Table 5, we attempt to isolate the household "welfare cost of nudging", the utility purely from receiving feedback. Recall that WTP is assumed to reflect the sum of (i) such utility impacts, (ii) avoided fee payments, and (iii) household costs when reducing waste, either in relation to D_0 or D_1. Since the fee rates are public information,

[35] The weights are derived from probit regressions (run separately for each municipality) that estimate the probability of being a survey respondent in 2023 using new covariate data from that year. In online Appendix F, we also consider net benefits within the respondent sample, with similar results as here.

[36] To make this calculation, we assume that each kilogram of residual waste reduced is matched by a corresponding 750-gram increase in packaging waste, and a 250-gram increase in food waste. Marginal social benefits of recycling (under high shadow prices) are higher for packaging waste than food waste, so the MSC increases with the share attributed to packaging.

[37] Indeed, the MSC turns out to be *negative* for low shadow prices, in which case any intervention to reduce residual waste will also reduce net benefits; see Kinnaman (2006) for one cost–benefit analysis with similar results. However, existing targets and policies for waste strongly suggest that policy makers believe the marginal social benefits of recycling are positive, which is why we focus on the high value.

[38] The differences are in line with the differential opt-out rates observed in footnote 14. All WTP figures adjust for likely censoring, given that a significant fraction of respondents state WTP equal to either the minimum possible value of −200 SEK or the maximum value of 200 SEK. In online Appendix F.2, we also check that WTP is correlated with other survey items (e.g., attitudes toward the letters) in expected ways, validating the measure.

component (ii) can be calculated directly.[39] Thus, given a credible estimate of component (iii), it is possible to isolate the utility impact of receiving feedback, component (i). As Fig. 7 shows (iii) as bounded by the areas below demand curves D_0 and D_1, respectively, we again use existing results on unit-based fees (Bueno and Valente, 2019) to infer the slope of these demand curves and calculate the shaded areas. Clearly, the implied bounds on the utility from receiving feedback differ across treatments but especially across municipalities: the utility purely from receiving feedback is near zero or slightly negative in Varberg, but positive in Partille.

As observed in Section 6.4, there are nonzero effects remaining in all treatments at the end of the data period. A reasonable but somewhat speculative exercise, therefore, is to calculate net benefits for some hypothetical "total" ATE obtained by extrapolating treatment effects into the future. In online Appendix F.4.2, we fit a linear curve to the full post-experimental period and keep adding effects until these have fully receded to zero. For instance, in the Varberg "standard HER" treatment, this happens in period 120, more than four and a half years after the start of treatment.[40] The results are shown in panel C of Table 5. Extending the effects in time does little to change results in Varberg, but has a large impact on net benefits in Partille, which are now about €12 per household. The large increase in Partille arises because the MSC is high compared to ϕ and the extrapolated effects do not attenuate over time. This underscores the point that net social benefits are largely determined by whether the marginal costs of residual waste are already internalized through the weight-based fee.

In online Appendix F.4.3 and F.4.4, we conduct two additional extensions of the analysis. First, recalling from Section 6.3 that ATEs are largely driven by households with high baseline waste generation, we evaluate a simple "profiling" design where only the 50% of households with baseline-average residual waste (weakly) above the median are treated. Because only half of all households are treated, profiling has opposing effects: it lowers both ATEs and direct intervention costs, and shifts mean WTP toward zero. We find that, indeed, net benefits do not necessarily increase with such profiling, and overall results are reasonably similar compared to our main analysis in panel A of Table 5.

Second, our particular experimental design clearly relies on the pre-existence of unit-based pricing. Thus, a policy-relevant question to ask is whether norm feedback could be applied in areas where such pricing is not in use. If so, would net benefits be positive? While providing a complete answer to this question is beyond the scope of this paper, we are able to conduct a simple preliminary evaluation. For a policy maker in an area where $\phi = 0$, even if there is no permanent shift to weighing, it may be cost-effective to lease trucks capable of weighing for the duration of an intervention. A temporary weight database can then be used to construct accurate feedback. Since the net benefits presented in Table 5 are generally largest for quarterly feedback, we assume that this is the frequency chosen, and that eventual ATEs are equal to those estimated for that treatment.

The added cost of installing and operating weighing equipment naturally drives up project implementation costs. Using cost data from a Swedish firm that supplies such solutions (Botek Systems AB, personal communications) and an estimate from Partille municipality of the number of trucks needed per single-family homes treated, we find that net benefits are positive as long as the fixed installation costs can be spread over multiple interventions (e.g., by leasing the same trucks to several municipalities), *and* mean WTP in the targeted area(s) exceeds €−2.29/household. If so, the additional cost of weighing is more than offset by the larger benefit of reducing waste where no pre-existing marginal incentive exists.

Finally, beyond net benefits, we might also consider whether norm feedback is cost-effective compared with other non-price policies that municipalities could use to reduce household waste. Foremost among these is arguably the replacement of drop-off facilities for packaging with curbside collection, thus reducing household effort and (to some extent) monetary costs associated with recycling. A policy report by the Swedish Waste Management Association (2016), a stakeholder organization mainly representing municipalities and public utilities, estimates the added administrative cost of operating such systems. Two types of curbside collection are considered: optical sorting and four-compartment waste bins. For municipalities with similar characteristics to Varberg and Partille, additional costs of the two system variants are calculated at 130 and 60 SEK per household and year, respectively. Taken over a period of 1.7 years, these figures translate into €21.82 and €10.07 per household.

By comparison, our total operation cost of providing monthly norm feedback is €4.57 per household, and quarterly feedback costs only €1.52 per household.[41] As a result, curbside collection would need to produce effects on residual waste several times larger than those reported in this paper to compete with feedback interventions. Existing policy evaluations (e.g., Bucciol et al., 2015; Best and Kneip, 2019) suggest effects of about 20%, so the cost-effectiveness of norm feedback seems at the very least on par with alternative policies for reducing household waste.

8. Concluding remarks

This paper has presented results from two large-scale behavioral interventions that apply Home Energy Report-type feedback in the waste domain, providing information to households on how much residual (unsorted) waste they generate compared to neighbors. Depending on the treatment and specification, we estimate immediate average treatment effects (waste reductions) of 7%–12%. These appear to be driven mainly by increased recycling of food and especially packaging waste. The effects are very persistent, with one-third or more of the immediate reduction remaining one year after treatment ended. Indeed, in one experiment, effects show no sign of attenuation.

We have also used incentive-compatible survey measures of household WTP for feedback to estimate the intervention's net social benefits. Results largely depend on the sign of mean WTP and on whether existing waste fees internalize the marginal social cost of residual waste. Feedback has positive net benefits mainly in Partille, where mean WTP is positive and weight-based fees are significantly below a (high) estimate of marginal social cost. Nevertheless, feedback seems highly cost-effective compared to curbside packaging collection, arguably the main alternative non-price policy to reduce household residual waste.

We also conduct an exploratory extension of the cost–benefit analysis to the common limiting scenario where there is no prior UBP or weighing of household waste. In this case, net benefits are positive whenever WTP is larger than €−2.29/household (which holds true for Partille but not Varberg). If so, the added cost of installing weighing equipment is more than offset by the benefit of reducing waste where no marginal incentive exists. These results may be conservative for

[39] Both municipalities apply UBP to both residual and food waste, so fee savings are zero whenever a household reduces residual waste through increased sorting of food waste. Given our assumption that food-waste recycling is behind 25% of feedback-driven reductions in residual waste, we compute fee savings as $\phi \times ATE \times 0.75$.

[40] In Partille, the fitted lines suggest that effects are without bound, increasing rather than diminishing over time. It seems prudent not to use these coefficients for extrapolation. Nevertheless, to provide a rough comparison with results for Varberg, we first sum over the longer-run estimates in online Appendix Table H.9 and, for periods $t > 18$, simply keep adding the last effect estimated in that table up until period 120.

[41] Including our estimate from online Appendix F.1.3 of the "switching cost" associated with fitting collection trucks with weighing equipment would increase these figures by €4.82 per household. However, the costs of curbside collection reported above also do not include likely substantial fixed switching costs in terms of purchasing waste bins and vehicles, etc.

two reasons. First, the analysis assumes treatment effects of equal magnitude as in our experiments, but effects from feedback may be larger in areas without UBP, where more "low-hanging fruits" are likely to remain than in already regulated areas.[42] Second, municipal utilities may wish to conduct weighing for other reasons, including simply to collect accurate waste data, implying co-benefits in terms of improved local governance.

One complication in relation to policy is that waste from individual households can typically be identified only in single-family housing. Thus, accurate household-level feedback cannot be applied in apartment blocks. Unlike for electricity, where smart metering is often available, there is no corresponding ready-made solution in the waste domain. Technical potential does exist: for example, utilities could issue household-specific waste bags and base feedback on the number of bags thrown. An alternative which remains little explored is to provide feedback at a more aggregated level, e.g., by apartment building. Clearly, this may introduce substantial free-riding incentives between households. Nevertheless, existing feedback interventions already presuppose some degree of cooperation between household members (as also noted by Brülisauer et al., 2020), and we believe aggregate feedback should be explored in future research.

Another refinement of the HER design that may be particularly pertinent for waste is to improve reference-group comparability. Our feedback follows most of the HER literature in adjusting for relatively few household characteristics, which runs the risk of reducing engagement if respondents perceive the comparisons as irrelevant or unfair. While this problem is not unique to our study, the relative ease of deep cuts in residual waste means that, for example, it is difficult to distinguish true low-use behavior from temporary absences (traveling, etc.) in our data. Thus, beyond controlling for more household observables in reference-group construction (possibly with the aid of predictive modeling), future studies in the waste domain could combine waste data with high-frequency data on electricity and/or water use where absent residents can be more accurately identified and adjusted for. Such rich data would also make it possible to provide feedback along multiple dimensions within a single sample, allowing effect sizes to be more directly compared across domains than we are able to do in this paper.

In the meantime, our results already provide clear proof of concept that norm-based feedback can be a useful tool for reducing waste in line with policy goals. More broadly, they imply that the (cost-)effectiveness of feedback is not limited to the relatively well-researched water and energy domains, and may be even greater for other types of behavior.

Contribution

Neither author has received funding from any interested party. Financial support from the Swedish research council Formas [grant nos. 2017-00225 and 2018-02603] is gratefully acknowledged. The funders have had no further involvement in study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the article for publication.

[42] Potential factors running in the opposite direction are site-selection bias in our two experiments (Allcott, 2015); and the finding of Myers and Souza (2020) that norm feedback may fail to reduce resource demand among households that do not pay for use. The setting in Myers and Souza (2020) was distinctive in some respects: a highly environmentally conscious student sample was used, possibly limiting the scope for improvement; and feedback specifically targeted heating rather than overall energy use. Nevertheless, future research should investigate whether their findings apply to other domains.

Ethical approval

In accordance with the Swedish Ethical Review Act (SFS 2003:460), which applies in the country where the authors' institutions are located, no IRB approval was required since no sensitive personal data was processed and the experiment did not involve a risk of harming subjects physically or psychologically.

Declaration of competing interest

none

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jpubeco.2024.105191.

References

Abadie, A., Athey, S., Imbens, G.W., Wooldridge, J., 2017. When Should You Adjust Standard Errors for Clustering? NBER Working Paper No. 24003.

Ahlroth, S., Finnveden, G., 2011. Ecovalue08 — A new valuation set for environmental systems analysis tools. J. Clean. Prod. 19 (17–18), 1994–2003.

Alberts, G., Gurguc, Z., Koutroumpis, P., Martin, R., 2016. Competition and norms: A self-defeating combination? Energy Policy 96, 504–523.

Allcott, H., 2011. Social norms and energy conservation. J. Public Econ. 95 (9–10), 1082–1095.

Allcott, H., 2015. Site selection bias in program evaluation. Q. J. Econ. 130 (3), 1117–1165.

Allcott, H., Kessler, J.B., 2019. The welfare effect of nudges: A case study of energy use social comparisons. Am. Econ. J.: Appl. Econ. 11 (1), 236–276.

Allcott, H., Rogers, T., 2014. The short-run and long-run effects of behavioral interventions: experimental evidence from energy conservation. Amer. Econ. Rev. 104 (10), 3003–3037.

Allers, M.A., Hoeben, C., 2010. Effects of unit-based garbage pricing: A differences-in-differences approach. Environ. Resour. Econ. 45 (3), 405–428.

Ambell, C., Björklund, A., Ljunggren Söderman, M., 2010. Potential för Ökad Materialåtervinning av Hushållsavfall och Industriavfall. Report, TRITA-INFRA-FMS 2010:4.

Andor, M.A., Gerster, A., Peters, J., Schmidt, C.M., 2020. J. Environ. Econ. Manag. 103, 102351.

Ayres, I., Raseman, S., Shih, A., 2012. Evidence from two large field experiments that peer comparison feedback can reduce residential energy usage. J. Law Econ. Organ. 29 (5), 992–1022.

Berglund, C., 2006. The assessment of households' recycling costs: The role of personal motives. Ecol. Econ. 56 (4), 560–569.

Bernedo, M., Ferraro, P.J., Price, M., 2014. The persistent impacts of norm-based messaging and their implications for water conservation. J. Consum. Policy 37 (3), 437–452.

Best, H., Kneip, T., 2019. Assessing the causal effect of curbside collection on recycling behavior in a non-randomized experiment with self-reported outcomes. Environ. Resour. Econ. 72, 1203–1223.

Brandon, A., Ferraro, P.J., List, J.A., Metcalfe, R.D., Price, M.K., Rundhammer, F., 2017. Do the Effects of Social Nudges Persist? Theory and Evidence from 38 Natural Field Experiments. NBER Working Paper No. 23277.

Brick, K., De Martino, S., Visser, M., 2017. Behavioral Nudges for Water Conservation: Experimental Evidence from Cape Town. Working Paper.

Bruchmann, K., Chue, S.M., Dillon, K., Lucas, J.K., Neumann, K., Parque, C., 2021. Social comparison information influences intentions to reduce single-use plastic water bottle consumption. Front. Psychol. 12, 612662.

Brülisauer, M., Goette, L., Jiang, Z., Schmitz, J., Schubert, R., 2020. Appliance-specific feedback and social comparisons: Evidence from a field experiment on energy conservation. Energy Policy 145, 111742.

Bucciol, A., Montinari, N., Piovesan, M., 2015. Do not trash the incentive! monetary incentives and waste sorting. Scand. J. Econ. 117 (4), 1204–1229.

Bueno, M., Valente, M., 2019. The effects of pricing waste generation: A synthetic control approach. J. Environ. Econ. Manag. 96, 274–285.

Burlig, F., Preonas, L., Woerman, M., 2020. Panel data and experimental design. J. Dev. Econ. 144, 102458.

Butera, L., Metcalfe, R., Morrison, W., Taubinsky, D., 2022. Measuring the welfare effects of shame and pride. Amer. Econ. Rev. 112 (1), 122–168.
Byrne, D.P., La Nauze, A., Martin, L.A., 2018. Tell me something I don't already know: Informedness and the impact of information programs. Rev. Econ. Stat. 100 (3), 510–527.
Carattini, S., Baranzini, A., Lalive, R., 2018. Is taxing waste a waste of time? Evidence from a supreme court decision. Ecol. Econ. 148, 131–151.
Costa, D.L., Kahn, M.E., 2013. Energy conservation "nudges" and environmentalist ideology: evidence from a randomized residential electricity field experiment. J. Eur. Econom. Assoc. 11 (3), 680–702.
Cragg, J.G., 1971. Some statistical models for limited dependent variables with application to the demand for durable goods. Econometrica 39 (5), 829–844.
Czajkowski, M., Hanley, N., Nyborg, K., 2017. Social norms, morals and self-interest as determinants of pro-environmental behaviors: The case of household recycling. Environ. Resour. Econ. 66 (4), 647–670.
Damgaard, M.T., Gravert, C., 2018. The hidden costs of nudging: Experimental evidence from reminders in fundraising. J. Public Econ. 157, 15–26.
Dijkgraaf, E., Gradus, R.H.J.M., 2004. Cost savings in unit-based pricing of household waste: The case of The Netherlands. Resour. Energy Econ. 26 (4), 353–371.
Dijkgraaf, E., Gradus, R., 2009. Environmental activism and dynamics of unit-based pricing systems. Resour. Energy Econ. 31 (1), 13–23.
Dijkgraaf, E., Gradus, R., 2017. An EU recycling target: What does the dutch evidence tell us? Environ. Resour. Econ. 68 (3), 501–526.
Dolan, P., Metcalfe, R., 2015. Neighbors, Knowledge, and Nuggets: Two Natural Field Experiments on the Role of Incentives on Energy Conservation. Working Paper No., 2589269, Becker Friedman Institute for Research in Economics.
Dupré, M., Meineri, S., 2016. Increasing recycling through displaying feedback and social comparative feedback. J. Environ. Psychol. 48, 101–107.
Ek, C., 2020. Serial-Correlation-Robust Power Calculation for the Analysis-of-Covariance Estimator. Working Paper.
Erhardt, T., 2019. Garbage in and garbage out? On waste havens in Switzerland. Environ. Resour. Econ. 73 (1), 251–282.
European Commission, 2018. Report on the implementation of EU waste legislation, including the early warning report for member states at risk of missing the 2020 preparation for re-use/recycling target on municipal waste. COM/2018/656 final.
European Environment Agency, 2022. Early warning assessment related to the 2025 targets for municipal waste and packaging waste.
Ferrara, I., Missios, P., 2012. A cross-country study of household waste prevention and recycling: Assessing the effectiveness of policy instruments. Land Econom. 88 (4), 710–744.
Ferraro, P.J., Miranda, J.J., 2013. Heterogeneous treatment effects and mechanisms in information-based environmental policies: Evidence from a large-scale field experiment. Resour. Energy Econ. 35 (3), 356–379.
Ferraro, P.J., Miranda, J.J., Price, M.K., 2011. The persistence of treatment effects with norm-based policy instruments: Evidence from a randomized environmental policy experiment. Amer. Econ. Rev. 101 (3), 318–322.
Ferraro, P.J., Price, M.K., 2013. Using nonpecuniary strategies to influence behavior: evidence from a large-scale field experiment. Rev. Econ. Stat. 95 (1), 64–73.
Fullerton, D., Kinnaman, T.C., 1996. Household responses to pricing garbage by the bag. Amer. Econ. Rev. 86 (4), 971–984.
Goetz, A., Mayr, H., Schubert, R., 2022. Beware of Side Effects? Spillover Evidence from a Hot Water Intervention. Working Paper.
Heller, M.H., Vatn, A., 2017. The divisive and disruptive effect of a weight-based waste fee. Ecol. Econ. 131, 275–285.
Holladay, S., LaRiviere, J., Novgorodsky, D., Price, M., 2019. Prices versus nudges: What matters for search versus purchase of energy investments? J. Public Econ. 172, 151–173.
Huang, J.-C., Halstead, J.M., Saunders, S.B., 2011. Managing municipal solid waste with unit-based pricing: Policy effects and responsiveness to pricing. Land Econom. 87 (4), 645–660.
Jaime Torres, M.M., Carlsson, F., 2018. Direct and spillover effects of a social information campaign on residential water-savings. J. Environ. Econ. Manag. 92, 222–243.
Jessoe, K., Lade, G.E., Loge, F., Spang, E., 2021. Spillovers from behavioral interventions: Experimental evidence from water and energy use. J. Assoc. Environ. Resour. Econ. 8 (2), 315–346.
Kinnaman, T.C., 2006. Policy watch: Examining the justification for residential recycling. J. Econ. Perspect. 20 (4), 219–232.
Kinnaman, T.C., Fullerton, D., 2000. Garbage and recycling with endogenous local policy. J. Urban Econ. 48 (3), 419–442.
McKenzie, D., 2012. Beyond baseline and follow-up: The case for more T in experiments. J. Dev. Econ. 99 (2), 210–221.
Mortensen, C.R., Neel, R., Cialdini, R.B., Jaeger, C.M., Jacobson, R.P., Ringel, M.M, 2019. Trending norms: A lever for encouraging behaviors performed by the minority. Soc. Psychol. Pers. Sci. 10 (2), 201–210.
Myers, E., Souza, M., 2020. Social comparison nudges without monetary incentives: Evidence from home energy reports. J. Environ. Econ. Manag. 101, 102315.
Schultz, P.W., 1999. Changing behavior with normative feedback interventions: A field experiment on curbside recycling. Basic Appl. Soc. Psychol. 21 (1), 25–36.
Schultz, P.W., Nolan, J.M., Cialdini, R.B., Goldstein, N.J., Griskevicius, V., 2007. The constructive, destructive, and reconstructive power of social norms. Psychol. Sci. 18 (5), 429–434.
Slorach, P.C., Jeswani, H.K., Cuéllar-Franca, R., Azapagic, A., 2019. Environmental sustainability of anaerobic digestion of household food waste. J. Environ. Manag. 236, 798–814.
Sparkman, G., Walton, G.M., 2017. Dynamic norms promote sustainable behavior, even if it is counternormative. Psychol. Sci. 28 (11), 1663–1674.
Sterner, T., Bartelings, H., 1999. Household waste management in a Swedish municipality: Determinants of waste disposal, recycling and composting. Environ. Resour. Econ. 13 (4), 473–491.
Swedish Environmental Protection Agency, 2023. Sveriges Återvinning av Förpackningar. Report, (In Swedish).
Swedish Waste Management Association, 2016. Beräkning av Avfallshanteringskostnader i Svenska Kommuner. Report 2016:29.
Usui, T., Takeuchi, K., 2014. Evaluating unit-based pricing of residential solid waste: A panel data analysis. Environ. Resour. Econ. 58 (2), 245–271.
Viscusi, W.K., Huber, J., Bell, J., 2011. Promoting recycling: Private values, social norms, and economic incentives. Amer. Econ. Rev. 101 (3), 65–70.
Vollaard, B., van Soest, D., 2024. Punishment to promote prosocial behavior: A field experiment. J. Environ. Econ. Manag. 124, 102899.