ARTICLES
https://doi.org/10.1038/s41560-020-00719-z

The interaction of descriptive and injunctive social norms in promoting energy conservation

Jacopo Bonan[1,2], Cristina Cattaneo[1], Giovanna d'Adda[1,3,4] ✉ and Massimo Tavoni[1,2]

Behavioural interventions that leverage social norms are widely used to foster energy conservation. For instance, home energy reports combine information on others' behaviour (descriptive feedback) and approval for norm compliant behaviour (injunctive feedback). In a randomized controlled trial, we investigated how descriptive and injunctive feedbacks interact to affect electricity use, and evaluate the effects of additional normative feedback presented in the form of descriptive or injunctive energy conservation norm primes. We found that consistent descriptive and injunctive feedback boosts the effectiveness of social information in inducing energy conservation. When descriptive and injunctive feedback are in conflict, conservation behaviour is a function of the relative strength of the two types of feedback. Additional normative feedback produces smaller gains when it reinforces existing information of the same type. These results suggest complementarities between different types of normative messages rather than superiority of any one kind of feedback.

Social information programmes that provide information on the actions or beliefs of others are widely used interventions to foster behavioural change in several domains[1–4], which include residential resource conservation[5–12]. These programmes typically feature two forms of feedback: descriptive and injunctive. In the case of residential energy, descriptive feedback generally takes the form of information on other households' average energy consumption, whereas injunctive feedback provides social approval for energy savings. The combination of descriptive and injunctive feedback within the standard design of social information programmes was inspired by the finding that descriptive information alone leads those who use less energy (low energy users) to increase their consumption, and that the addition of injunctive feedback in support of an energy conservation norm prevents this boomerang effect[13]. Injunctive information thus counterbalances descriptive information. However, the experimental evidence on the impact of descriptive and injunctive information when they exert opposing influences on behaviour is mixed and mainly focuses on short-term or self-reported outcomes[14,15].

Given the wide adoption of communication campaigns that rely on social information to promote behavioural change among both policymakers and private firms, it is important to understand how different programme features interact in real-world settings[16]. Indeed, impact evaluations of similar programmes find that they are effective in fostering energy savings, but that effect sizes vary widely across contexts and individuals[17]. Prominent explanations for such differential responses rely on the heterogeneity of consumers' traits, such as beliefs[18,19], misperceptions of one's compliance with the social norm[20] or personal values[21,22].

Here we focus on the varying effect of specific features of these messages, and particularly on how the salience, strength and consistency of the feedback they contain differ, and thus affect behaviour differently, across users. This could inform a more effective design and targeting of messages and provide more specific and nuanced guidance to prevent similar information campaigns from backfiring[16]. First, we exploited the features of the standard design of home energy reports and isolate the impact of changes in injunctive feedback. Specifically, we examined whether reinforcing the injunctive feedback has different effects on electricity use if it is accompanied by consistent descriptive feedback—as is the case for those who use more energy (high energy users), for whom both the injunctive and descriptive information encourage energy conservation—or contrasting descriptive feedback—as is the case for low energy users, for whom conforming with the descriptive feedback entails consumption increases, at odds with the injunctive feedback that praises energy saving. Second, we randomized descriptive or injunctive information that primes a social norm of energy conservation, and evaluated the effect of strengthening the injunctive feedback in the presence of either the descriptive or the injunctive prime.

We propose a conceptual framework for understanding how different features of social information programmes impact energy conservation that can be articulated in a set of hypotheses, illustrated in Fig. 1. First, the effectiveness of a normative message is maximized by the inclusion of consistent feedback of different types (that is, injunctive and descriptive; Fig. 1a). Second, when injunctive and descriptive feedbacks are in contrast (Fig. 1b), the strength of each single piece of information matters. The strength of the normative feedback may depend on several factors highlighted in the literature, from the recipient's beliefs on what relevant others think is socially approved of[18,23] to the degree of consensus or ambiguity around the norm conveyed by the information[16,24]. In our setting, we hypothesize that the effect of the descriptive information increases according to the difference between an individual's electricity consumption and the average consumption of the reference group. The effect of injunctive information instead varies according to the strength of social approval conveyed through visual cues and encouragement messages. Third, additional pieces of consistent feedback of the same type produce smaller savings (Fig. 1c).

Our results are in line with these hypotheses. First, we found suggestive evidence that the standard social information message induces larger savings among high electricity users who are exposed to consistent descriptive and injunctive feedback, compared with low electricity users who are exposed to contrasting descriptive and injunctive feedback. More importantly, reinforcing the injunctive

[1]RFF-CMCC European Institute on Economics and the Environment (EIEE), Centro Euro-Mediterraneo sui Cambiamenti Climatici, Milan, Italy. [2]Department of Management, Economics and Industrial Engineering, Politecnico di Milano, Milan, Italy. [3]Department of Economics, Management and Quantitative Methods (DEMM), University of Milan, Milan, Italy. [4]Fondazione Pesenti, Milan, Italy. ✉e-mail: giovanna.dadda@unimi.it

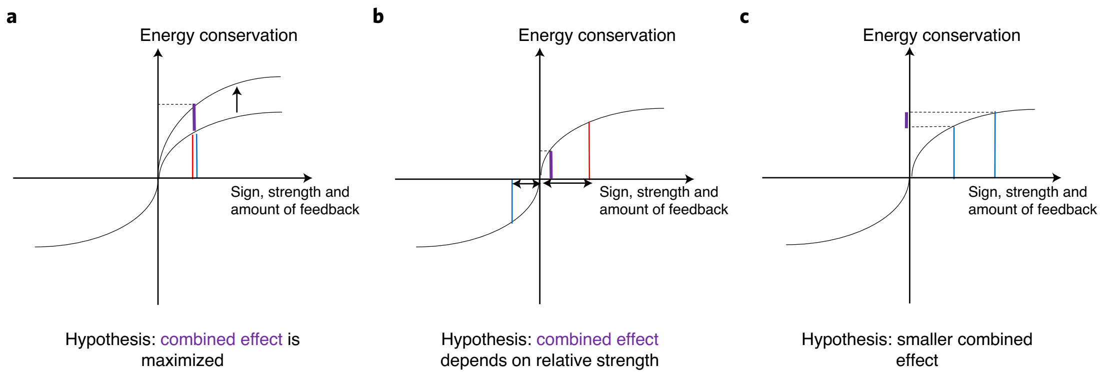

Fig. 1 | Hypothesized impact of injunctive and descriptive feedbacks in social information messages. a–c, Hypothesized effects of consistent injunctive and descriptive (blue) feedbacks (**a**), contrasting injunctive (red) and descriptive feedbacks (**b**) and additional consistent feedbacks of the same type (**c**). The black curve represents the overall impact of the normative message. Injunctive feedback is shown in red, descriptive feedback is shown in blue and their combined effect is shown in purple. The horizontal axes indicate the sign, strength and amount of normative feedback, where positive feedback values imply messages that encourage electricity savings. The vertical axes represent the behavioural outcome where positive values are associated with behaviour in compliance with the norm, which in our case is energy conservation.

feedback has the largest effect among low electricity users exposed to a consistent descriptive prime. These findings are in line with the notion that injunctive and descriptive feedbacks have a larger impact on electricity conservation when they pull behaviour in the same rather than in opposite directions. Second, reinforcing the injunctive feedback led to a reduction in consumption, but only among customers with low electricity usage. This shows that the relative strengths of the different types of feedback matters when they are contrasting. Such a reinforcement has no effect on customers with high consumption. This demonstrates the limited effect of reinforcing one type of normative feedback within a message that already contains two consistent pieces of normative information of different types. This is further confirmed by the finding that reinforcing the injunctive feedback has no effect among users exposed to a consistent injunctive prime. Together these findings suggest that additional pieces of feedback have a larger impact when they pull behaviour in the same direction and are of different types. Overall, our results support the presence of synergies between different types of feedback rather than the primacy of any one type of feedback.

Field experiment

Our setting consists of a randomized controlled trial implemented by an Italian energy company that provides almost half-a-million households with information on their electricity use relative to that of their neighbours[6,17]. The social information is included in a Home Energy Report distributed to customers via email (eHER). The programme was rolled out in 2016 and involved 464,523 customers ($n=418{,}178$ treatment, $n=46{,}345$ control). The core feature of the eHER is the neighbour comparison, which combines descriptive and injunctive normative information. The descriptive norm graphically compares the customer's electricity use over the previous month with the average use in two reference groups: 100 similar customers who live nearby (that is, neighbours) and the 15% most-efficient neighbours. The injunctive norm takes the form of thumbs-up symbols next to the descriptive norm graph: three thumbs up ('excellent') for users who consume within the top 15% most-efficient neighbours, two thumbs up ('good') for those more efficient than the average neighbour and one thumbs up ('you can do better') for the others. Figures 2a,b shows the eHER for users receiving three and two thumbs up, respectively.

We collaborated with the energy company to augment this set-up with a message displayed at the bottom of the eHER delivered in April–May 2018. The utility randomly allocated half of the treated sample at that time ($n=256{,}487$) to receive either the descriptive ($n=127{,}899$) or the injunctive ($n=128{,}588$) message priming an energy conservation norm (Fig. 2c, Supplementary Methods and Supplementary Fig. 1). The descriptive norm prime emphasizes that a large majority of customers try to save energy, that is, adopt behaviours consistent with a social norm of saving electricity. The injunctive norm prime claims that a majority of customers hold electricity saving as a personal value, which thus supports the belief that electricity saving is approved by relevant others. The two primes use fellow customers of the same utility as the reference group. The information on energy saving behaviours and values featured in the primes was taken from an online survey that we conducted with about 3,000 utility customers (Methods).

We have access to data on monthly electricity consumption from July 2015 to December 2019. The daily average electricity usage, normalized with respect to the control group consumption in the intervention period, was our main outcome variable. Pre-intervention daily electricity usage in a month was calculated over the period July 2015 to June 2016. Our data also include information on the contents of customers' reports and on whether customers open or click on them. We provide details on the programme implementation and data in the Methods, and descriptive statistics and balance tests in Supplementary Tables 1 and 2 and Supplementary Note 1. Samples are balanced across all available dimensions.

Impact of the social information programme

The impact evaluation of the standard programme indicates a statistically significant reduction of normalized electricity usage in its first year (coefficient $=-0.353$, standard error (s.e.) $=0.113$, $P=0.002$; equation (1) in Methods and Supplementary Table 3, column 1). The impact of the treatment increases with baseline consumption, although this result is not robust to the measure of electricity consumption used, that is, discrete or continuous (Supplementary Table 3, columns 2 and 5); its statistical significance varies with how the sample is defined (Supplementary Table 4) and with the time frame considered (Supplementary Tables 5–7) and it does not always hold after multiple hypotheses corrections. Exploiting data on engagement with the reports and on changes in feedback over time, we found that the impact of social information is magnified among users who actually read it and who experience upgrades in feedback (see Supplementary Tables 8 and 9 and Supplementary Note 2 for further details).

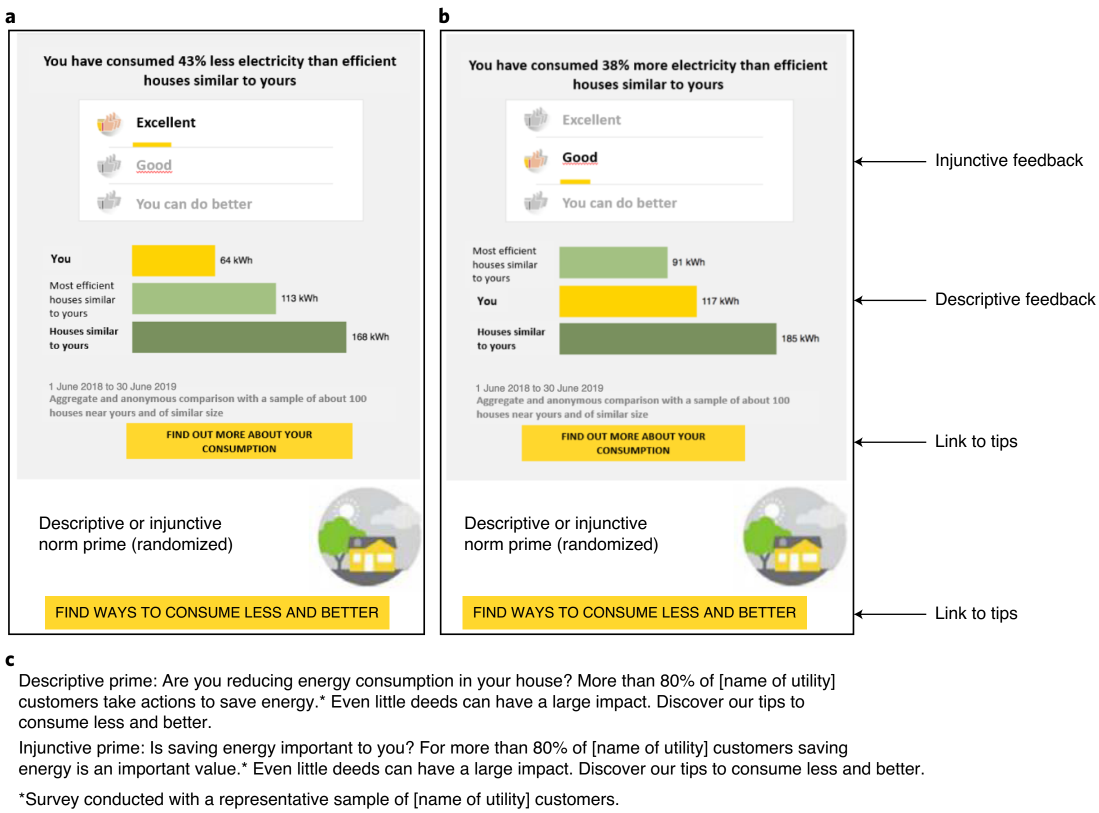

Fig. 2 | Home Energy Report. a,b, Layout and content of a Home Energy Report for a user receiving three thumbs up (**a**) and a user receiving two thumbs up (**b**). Both versions of the report contain the injunctive feedback, that is, the thumbs up (top), and the descriptive feedback, that is, the energy consumption bars (bottom). **c**, Text of the randomized norm primes. Credit: Copyright 2016-2020 Oracle. All rights reserved.

Although the magnitude of the average savings from the programme (−0.353%) is outside the range of those generated by similar ones in the United States (minimum = 0.88%, maximum = 2.55%) (ref. [18]), they are in line with the existing evidence from Europe[19]. Various factors, such as lower average consumption in Europe than that in the United States, the specific features of the programme we studied or differences in beliefs across contexts, may be responsible for these differences. The heterogeneous effects, although not robust and only marginally statistically significant, are qualitatively in line with the existing evidence on the larger impact of social information on high electricity users[17,20,25] and on the absence of boomerang effects among low users[13].

These results provide initial, albeit weak, support for our conceptual framework. For high users, normative and injunctive feedbacks pull behaviour in the same direction, which results in a reduction in electricity almost twice as large as that in the average treatment effect. For low electricity users, conforming to the reference groups' behaviour motivates a consumption increase ('boomerang'), but the injunctive feedback included in the eHER counterbalances the negative effect of the descriptive feedback. The injunctive feedback therefore induces stronger behavioural reactions among high electricity users, who are also exposed to the supporting descriptive feedback, than that among low electricity users, for whom the two types of feedback are at odds. Although such an interpretation is only suggestive based on the evidence presented so far, it shows how established findings are consistent with our conceptual framework.

Impact of strengthening the injunctive feedback

Our conceptualization can guide the analysis and interpretation of the effect of intensifying the injunctive feedback, with the descriptive feedback kept unchanged, for low and high users. We isolated the causal impact of the strength of the injunctive feedback via a regression discontinuity (RD) estimation (Methods). We exploited the fact that the injunctive feedback (number of thumbs up) changes discretely as a customer's consumption crosses the two thresholds given by the average consumption of the efficient neighbours (three versus two thumbs-up cutoff), and the neighbours' average consumption (two versus one thumbs-up cutoff), whereas the descriptive feedback (bars of electricity use) changes continuously across the thresholds. We focused on customers around the thresholds, whose assignment to a given injunctive feedback category was plausibly random. Indeed, although customers that belong to the three normative feedback categories differ on average along various baseline characteristics (Supplementary Note 3 and Supplementary Table 10), individuals close to the thresholds are similar (Supplementary Tables 11 and 12, columns 1–3).

We conducted two separate RD estimations, one for each cutoff, on the sample of treated customers who received the eHER sent in April–May 2018 ($n = 256{,}487$) to allow a direct comparison with the analysis presented below. In each estimation, we compared users in the two feedback categories adjacent to the cutoff and estimated the marginal effect of receiving one additional thumbs up. Figure 3 presents the results in terms of level changes in electricity usage. Although there are no statistically significant changes in the effect of the eHER when crossing the threshold between the one and two thumbs up (Fig. 3a), the discrete shift in the injunctive norm reduces electricity use when moving from the two to three thumbs up (Fig. 3b). The corresponding empirical estimates are presented in Table 1 (columns 1 and 2).

We can attribute these effects to the social information contained in the report rather than to other content, namely electricity saving tips—tips can only be accessed through a clickable link on the

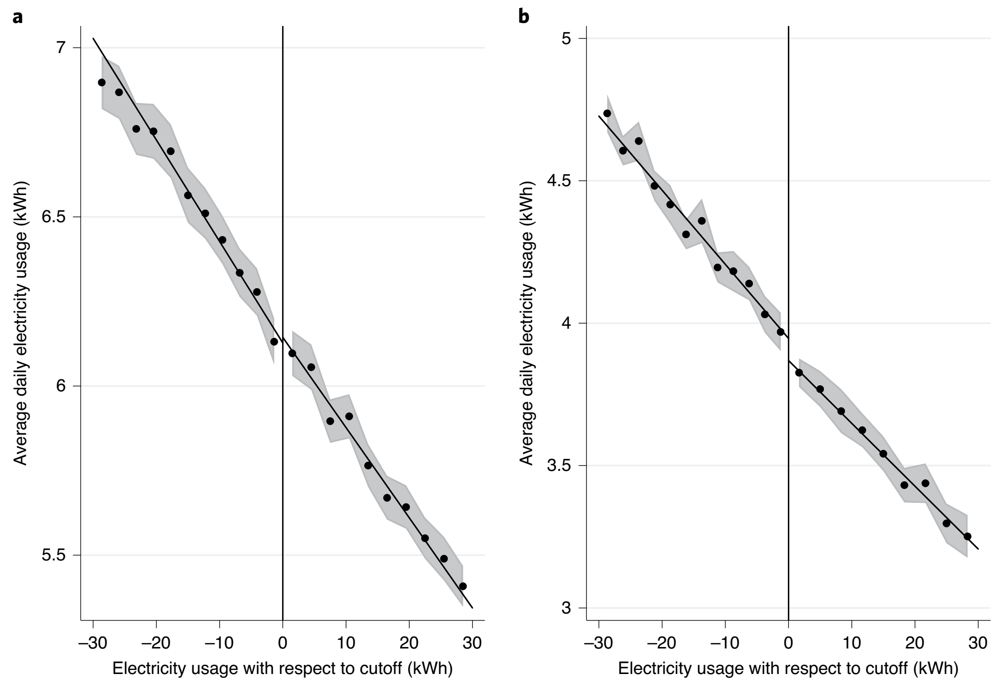

Fig. 3 | Impact of the injunctive feedback on electricity usage. a,**b**, Each dot represents the average daily electricity usage in the 3 months after the receipt of the April–May 2018 eHER around the two versus one thumbs-up cutoff ($n = 216{,}328$) (**a**) and three versus two thumbs-up cutoff ($n = 130{,}466$) (**b**) within evenly spaced bins. The solid line represents the local linear fit, estimated separately on either side of the cutoff and the shaded area shows 95% confidence intervals. The number of bins was selected through the integrated mean squared error (m.s.e). The running variable (horizontal axis) reports the individual difference between each customer's monthly consumption in the period reported in the eHER and the relevant cutoff. The cutoff is then represented by the vertical line, which is set to zero. For positive values of the score, customers get an extra thumbs up with respect to those with negative values of the score. In **a**, customers on the right of the cutoff consume less than the average neighbour and more than the top 15% most efficient and get two thumbs up. In **b**, customers on the right of the cutoff consume within the 15% most efficient neighbours and get three thumbs up. Bandwidths (BWs) and kernel are set following the data-driven process described for formal RD estimations of impacts reported in Table 1.

report and we see no difference in click shares across the cutoffs (Supplementary Table 12). The impact of the shift in injunctive feedback is persistent even after 6 and 12 months (Supplementary Table 13). The results are robust to different specifications of the RD estimate (Table 1).

These findings are consistent with our second hypothesis. Strengthening the approval for electricity savings delivered through the injunctive feedback affects customers around the three versus two thumbs-up cutoff more because, for them, it reinforces the relative strength of the injunctive information in the presence of conflicting injunctive and descriptive feedback. On the contrary, and in agreement with our third hypothesis, for consumers around the two versus one thumbs-up cutoff, the reinforcement of the injunctive feedback has a smaller marginal effect, as it adds strength to already consistent feedback types.

Impact of additional injunctive and descriptive feedback

To further test the impact of an additional piece of normative information, we combined the features of the standard eHER with the randomized addition of descriptive and injunctive information through the primes. To exploit the interaction between the discontinuities in the eHER's injunctive feedback and the randomly delivered primes, we repeated the RD analysis across the two cutoffs (two versus one and three versus two thumbs up) separately for the subsamples of customers who received the two types of prime.

The results are reported in Fig. 4 and Table 1 (columns 3–6). Across the cutoff between one and two thumbs up, we observe no statistically significant changes in consumption, regardless of whether the descriptive (Fig. 4a) or the injunctive (Fig. 4b) prime is present. Conversely, a discrete shift in the injunctive feedback across the three versus two cutoff causes electricity reduction, but only when combined with the descriptive prime that nudges energy efficiency (Fig. 4c). The results are robust to adjustments for multiple hypothesis testing (Table 1) and are persistent over longer time horizons (Supplementary Table 13).

This evidence, consistent with our first hypothesis, suggests synergies between different types of feedback: adding supportive descriptive information increases the impact of the shift in injunctive feedback. The marginal contribution of additional feedback of the same type instead decreases: customers exposed to the injunctive prime do not react to the reinforcement of the injunctive feedback across the three versus two thumbs-up cutoffs (Fig. 4d). Similarly, strengthening the injunctive feedback across the two versus one thumbs-up threshold makes no difference, regardless of whether the descriptive (Fig. 4a) or injunctive information (Fig. 4b) is added. In this case, the descriptive and injunctive feedback within the standard neighbour comparison already pull behaviour in the same direction. Further priming either type of normative feedback does not generate incremental electricity conservation.

To determine whether the overall effect of crossing the three versus two thumbs-up threshold is exclusively due to the presence within the eHER of the descriptive prime, we performed the RD estimation on a standard eHER (February–March 2018). We found statistically significant effects (coefficient $= -0.855$, s.e. $= 0.368$, $P = 0.02$; Table 2, column 1). Therefore, the effect of reinforcing the injunctive feedback for low users does not depend on the presence of the descriptive prime within the report. In addition, we observe that the overall effect of crossing the three versus two thumbs-up

Table 1 | Regression discontinuity estimates of the impact of the injunctive norm and normative prime message on electricity usage

	All		Descriptive prime		Injunctive prime	
	1	2	3	4	5	6
Three versus two thumbs up						
Conventional	−1.217***	−1.165**	−2.426***	−2.388***	−0.0550	0.209
	(0.461)	(0.456)	(0.619)	(0.611)	(0.686)	(0.676)
			[0.001]	[0.001]	[1]	[0.66]
Robust bias-corrected	−1.130**	−1.077**	−2.264***	−2.349***	−0.0470	0.310
	(0.461)	(0.456)	(0.619)	(0.611)	(0.686)	(0.676)
			[0.002]	[0.001]	[1]	[0.68]
Observations	130,466	130,466	65,091	65,091	65,305	65,305
BW select method	1 m.s.e.	2 m.s.e.	1 m.s.e.	2 m.s.e.	1 m.s.e.	2 m.s.e.
BW above	30.59	35.67	34.50	40.27	27.94	39.15
BW below	30.59	26.95	34.50	28.89	27.94	23.14
Effective number of observations above	31,226	36,752	17,852	21,096	10,698	20,632
Effective number of observations below	23,216	20,908	12,669	11,051	13,900	9,392
Two versus one thumbs up						
Conventional	0.254	−0.564	−0.426	−1.046*	0.491	−0.161
	(0.531)	(0.434)	(0.704)	(0.601)	(0.691)	(0.619)
			[1]	[0.14]	[1]	[0.66]
Robust bias-corrected	0.488	−0.409	−0.118	−0.883	0.731	−0.150
	(0.531)	(0.434)	(0.704)	(0.601)	(0.691)	(0.619)
			[1]	[0.27]	[0.771]	[0.68]
Observations	216,328	216,328	107,729	107,729	108,477	108,477
BW select method	1 m.s.e.	2 m.s.e.	1 m.s.e.	2 m.s.e.	1 m.s.e.	2 m.s.e.
BW above	30.10	76.51	35.93	84.88	33.80	67.97
BW below	30.10	33.90	35.93	37.67	33.80	30.86
Effective number of observations above	32,148	40,038	18,765	22,115	20,015	18,306
Effective number of observations below	35,704	69,226	21,124	36,967	17,843	31,980

The table shows the impact of the injunctive norm, that is, the number of thumbs up, on electricity usage, overall and by normative prime message. The outcome variable of the RD estimations is the average daily energy usage (kWh) in the 3 months after the receipt of the eHER augmented with the normative prime, normalized by the average control group consumption in the post period, around the three versus two thumbs up and two versus one thumbs up. The estimations in columns 1 and 2 are on the whole sample, in columns 3 and 4 are on the sample of customers who received the descriptive norm prime and in columns 5 and 6 are on the sample of those who received the injunctive norm prime. The first row shows the conventional RD estimate for the thumbs-up comparison, whereas the second corrects for bias[48]. BWs were selected to minimize the m.s.e.[48,49]. Odd columns use the same BW on either side of the cutoff, whereas even columns estimate separate BWs. Standard errors are clustered at the customer level in parentheses. FDR (false discovery rate)-adjusted q values are given in square brackets. $^{***}P < 0.01$, $^{**}P < 0.05$, $^{*}P < 0.1$.

threshold within the standard report is smaller than the same effect when the report is augmented with the descriptive prime (coefficient = −2.426, s.e. = 0.619, $P < 0.01$; Table 1, column 3). This confirms that the combination of consistent descriptive and injunctive information boosts the effectiveness of social information in inducing energy conservation.

The RD estimation provides a robust identification of the causal effects, which are, however, only local. We complemented it with an estimation of the heterogeneous impact of the primes by the number of thumbs up that customers receive (equation (3) in Methods). We obtained similar results: although the descriptive prime does not influence consumption on average (coefficient = 0.088, s.e. = 0.149, $P = 0.554$; Table 3, column 1), it led to a negative and statistically significant decrease in consumption among customers who received three thumbs up (coefficient = −0.959, s.e. = 0.284, $P < 0.001$; Table 3, column 4). This negative effect was persistent over 6, 12 and 18 months (Supplementary Note 4 and Supplementary Table 14). Interestingly, the effect of the descriptive prime on the entire group of customers who received three thumbs up is smaller than the RD estimates for the three versus two thumbs-up threshold combined with the descriptive prime. Although these simple heterogeneity results should be taken with caution, as thumbs up are correlated with customers' characteristics (Supplementary Table 10), they can be interpreted in light of our conceptual framework and suggest a potential determinant of the strength of descriptive information. The sample of customers who received three thumbs up includes the most efficient users, who are far from the three versus two thumbs-up threshold. The further customers are from the threshold, the larger the deviation between their own consumption and the average electricity use, and therefore, we argue, the larger the strength of the descriptive norm contained in the standard eHER. In other words, we suggest that in our setting conformity motives become more influential the further away individuals are from the descriptive norm.

This interpretation was confirmed by analysing specifically the effect of the descriptive prime among customers who experienced an upgrade in the injunctive feedback (from two to three thumbs up) relative to the previous report. These customers are likely to overlap with the customers included in the RD estimation, as being close to the three versus two thumbs-up threshold may result in downgrades and upgrades between reports. The effect of the descriptive prime on these customers is in line with the RD estimates for the

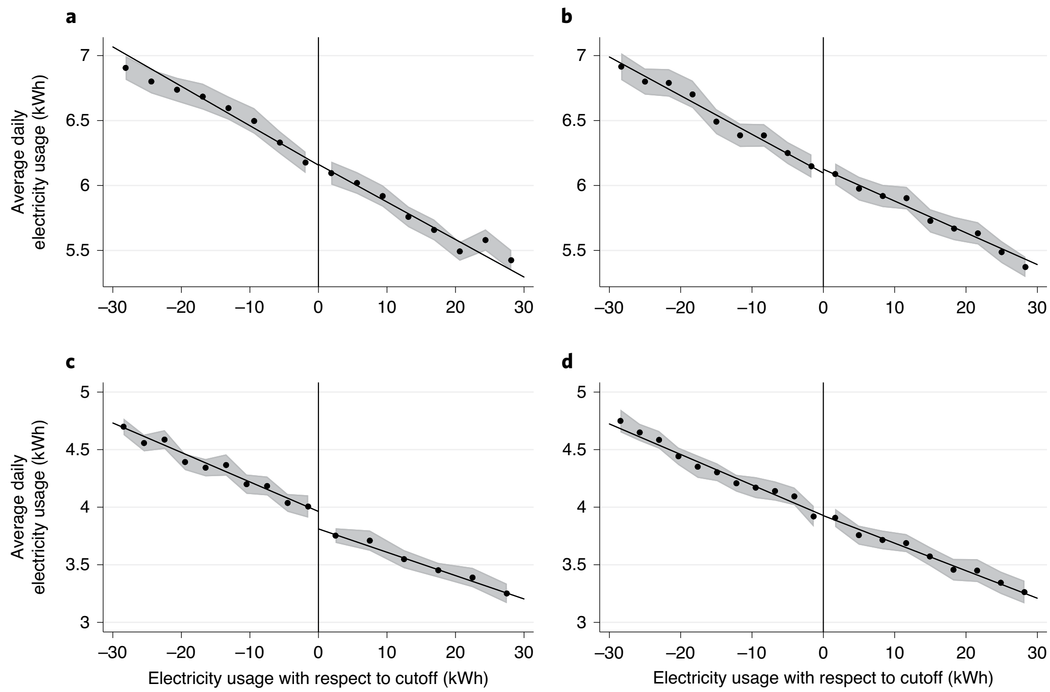

Fig. 4 | Heterogeneous impact of the normative primes at different injunctive feedback cutoffs. Each dot represents the average daily electricity usage (kWh) in the 3 months after the receipt of the eHER augmented with the normative prime, within evenly spaced bins. The solid line represents local linear fit, estimated separately on either side of the cutoff with the shaded area showing 95% confidence intervals. The number of bins is selected through the integrated m.s.e. The running variable (horizontal axis) reports the individual difference between each customer's monthly consumption (kWh) in the period reported in the eHER and the relevant cutoff. The running variable (horizontal axis) reports the individual difference between each customer's monthly consumption (kWh) in the period reported in the eHER and the relevant cutoff. For positive values of the score, customers get an extra thumbs up with respect to those with negative values of the score. **a**, Customers who received the descriptive norm prime around the two versus one thumbs-up cutoff (*n* = 107,729). **b**, Injunctive norm prime around the two versus one thumbs-up cutoff (*n* = 108,477). **c**, Descriptive norm prime around the three versus two thumbs-up cutoff (*n* = 65,091). **d**, Injunctive norm prime around the three versus two thumbs-up cutoff (*n* = 65,305).

Table 2 | Regression discontinuity estimates of the impact of the injunctive norm on electricity usage

	Three versus two thumbs up		Two versus one thumbs up	
	1	2	3	4
Conventional	−0.855**	−0.834**	0.0321	0.00172
	(0.368)	(0.342)	(0.400)	(0.345)
Robust bias-corrected	−0.683*	−0.701**	0.167	0.131
	(0.368)	(0.342)	(0.400)	(0.345)
Observations	134,970	134,970	224,212	224,212
BW select method	1 m.s.e.	2 m.s.e.	1 m.s.e.	2 m.s.e.
BW above	25.07	35.54	26.49	77.43
BW below	25.07	21.87	26.49	24.06
Effective number of observations above	25,013	35,922	28,920	71,838
Effective number of observations below	20,674	17,915	30,963	28,136

This table shows RD estimation of average daily energy usage in the 3 months after the receipt of the eHER in February–March 2018 (that is, the one preceding the eHER augmented with the normative prime), normalized by the average control group consumption in the post period, around the three versus two thumbs-up cutoff (columns 1 and 2) and two versus one thumbs-up cutoff (columns 3 and 4). The first row shows the conventional RD estimate, whereas the second corrects for bias[48]. BWs were selected to minimize the m.s.e.[48,49]. Odd columns use the same BW on either side of the cutoff, whereas even columns estimate separate BWs. Standard errors are clustered at the customer level in parentheses. ***$P < 0.01$, **$P < 0.05$, *$P < 0.1$.

three versus two thumbs-up threshold combined with the descriptive prime (Fig. 5) and larger than the average impact of the descriptive prime on the three thumbs-up subsample. Other factors may contribute to these results, but we note that they are consistent with our argument that, when descriptive and normative expectations diverge, the resulting behaviour is a function of the relative strength of the two types of feedback.

Conclusions

Our findings have implications for the design of social information programmes that rely on the combination of different types of norms to maximize behavioural change. Similar programmes are used in several domains, such as tax compliance[26,27], charitable giving[28] or water conservation[29]. According to our conceptual framework and empirical results, no single type of normative information is more effective in absolute terms. Policymakers should, instead, pay attention to the type of normative feedback they include in their communication, strive to diversify them, avoid conflicting information when it mitigates the desirable effects and exploit it otherwise, be aware of the diminishing returns from additional pieces of social information and of the varying strength of conformity motives across individuals.

Of course, our results may be specific to the context that we studied, and particularly to the formulation of injunctive and descriptive feedbacks that characterize the energy efficiency programme we evaluated and the normative primes we designed. For example, the wording and graphical representation of the injunctive feedback in the eHER of this study differ from those of widely evaluated standard social information programmes[6,17,18,21,30]. Further investigations

Table 3 | Impact of the descriptive versus injunctive messages on electricity usage

	All	One thumbs up	Two thumbs up	Three thumbs up
	1	2	3	4
Post	−0.300***	−0.081	−0.459***	−1.190***
	(0.113)	(0.194)	(0.128)	(0.213)
Descriptive prime × post	0.088	0.225	0.296*	−0.959***
	(0.149)	(0.252)	(0.170)	(0.284)
		[0.142]	[0.09	[0.003]
Constant	102.302***	133.685***	81.618***	50.329***
	(0.044)	(0.076)	(0.052)	(0.080)
Observations	2,783,190	1,358,951	1,002,787	421,452
R-squared	0.109	0.171	0.093	0.044
Number of customers	256,487	125,249	92,407	38,831

The dependent variable is the average daily electricity usage (kWh), main and heterogeneous effects, normalized by average control group consumption in the post period. The reference period for the analysis is October 2017 to August 2018 (3 months impact). All the specifications include customer fixed effects and month by year fixed effects. Standard errors are clustered at the customer level in parentheses and FDR-adjusted *q* values in square brackets. ***$P<0.01$, **$P<0.05$, *$P<0.1$.

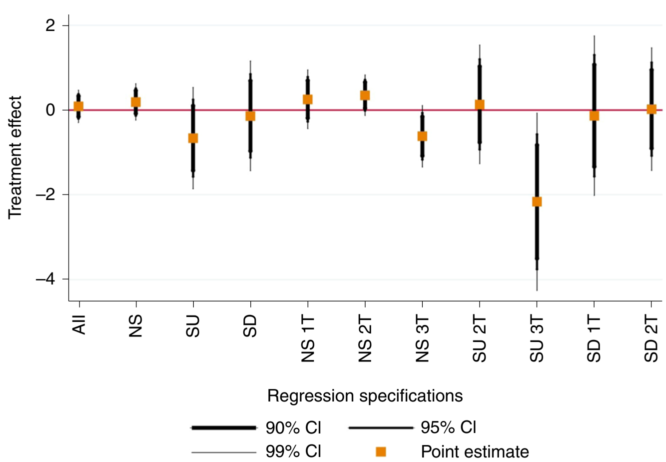

Fig. 5 | Prime impact by up- and downgrades in the injunctive feedback category. The figure plots the coefficient of the descriptive norm × post period on average daily electricity usage (kWh), normalized by the average control group consumption in the post period, estimating equation (3) for the following subsamples. All, whole sample of customers included in the normative prime trial ($n=2{,}783{,}190$); NS, customers who experienced no switch in the number of thumbs in the eHER augmented with the normative prime with respect to the previous one ($n=2{,}334{,}191$); SU, switch up, that is, an improvement in the number of thumbs in the eHER augmented with the normative prime with respect to the previous one ($n=239{,}501$); SD, switch down, that is, a decrease in the number of thumbs in the eHER augmented with the normative prime with respect to the previous one ($n=209{,}498$); NS 1T, no switch and one thumbs up in the eHER augmented with the normative prime ($n=1{,}232{,}602$); NS 2T, no switch and two thumbs up ($n=773{,}165$); NS 3T, no switch and three thumbs up ($n=328{,}424$); SU 2T, switch up and two thumbs up ($n=146{,}473$); SU 3T, switch up and three thumbs up ($n=93{,}028$); SD 1T, switch down and one thumbs up ($n=126{,}349$); SD 2T, switch down and two thumbs up ($n=83{,}149$). CI, confidence interval.

in other behavioural domains and of alternative formulations of descriptive and injunctive feedback are needed to verify the generalizability of our findings.

Finally, although we identify potential determinants of the strength of the social information messages, we are far from formulating a comprehensive theoretical model. Such a model should incorporate other important insights from the social information literature, for instance, on the role of individual descriptive and normative second-order beliefs, from the perceived consensus around the norm or from misperceptions that concern one's own compliance with the norm. Similarly, within our setting, due to lack of data, we can only test a few implications of our conceptual framework and cannot control for the influence of other important determinants of the impact of social information identified in the literature. The predictions from a more comprehensive model should, instead, be subject to a systematic experimental investigation.

Methods

Ethics statement. Ethical approval for the use of the data that support the findings of this study was granted by the Institutional Review Board at Politecnico di Milano (approval number 04/2017). Consent for their administrative data to be used for the research was given by the users as part of the privacy consent statements that they signed with the utility.

Programme details. We evaluated a social information programme designed and implemented by the utility and Opower (acquired by Oracle in 2016). The eHER that constitutes the core feature of the programme differed from the standard Opower HER, evaluated in other works[6,17,18,21,30], under three respects: first, it was delivered by email rather than by post, hence the notation eHER; second, it did not feature a section with energy saving tips—tips could be consulted by interested customers within their personal area on the utility's website, accessible through a clickable link on the eHER; third, the normative feedback was given through thumbs-up symbols accompanied by the expressions 'excellent', 'good' and 'you can do better', rather than through the standard smiley faces coupled with 'great', 'good' and 'above average'. The first two differences are consistent with the objective to foster customers' digital engagement, which the utility primarily wanted to achieve through the programme, whereas the third is the result of focus groups conducted by the utility and Opower to define the design of the eHER.

Our augmented eHER added a simple treatment to this basic set-up in the form of a message at the bottom of the report. We proposed a formulation of the descriptive and injunctive normative messages based on previously collected survey data and collaborated with Opower and the utility to finalize the wording, layout and graphical aspects of the messages. Opower and the utility were responsible for the randomization of the normative messages and the implementation of the test.

In addition to the experimental test discussed in the present study, we manipulated in a similar way the November–December 2017 eHER. The experiment aimed to test the impact of environmental identity on energy conservation. It featured a treatment message that primed individual environmental self-identity and a control message that encouraged energy conservation[22]. Given that both the November–December 2017 and the April–May 2018 primes were randomized, participation in the environmental prime test should not affect the results presented here. Nevertheless, we support this claim through further tests reported below.

Sample. Our sample of analysis is represented by the entire eligible customer base at the time of the start of the programme ($n=464{,}523$). Eligibility criteria were established and verified by the utility and Opower, and included availability of a contact email address and a set of technical requirements, such as living in single family homes, having one year of pretreatment consumption data without missing, negative or abnormally high usage, and having a sufficient number of neighbours—defined as customers who lived within a 10 km radius and were similar in terms of type of housing and any other characteristic available to the utility) for the neighbour comparison. The utility is present over the full national territory and the programme was targeted to all eligible customers regardless of their area of residence (Supplementary Fig. 2 shows the study sample distribution across Italian municipalities). Moreover, to foster energy conservation was not the main goal of the programme. These considerations alleviate concerns of site selection bias[31].

Eligible customers were randomly assigned to the treatment ($n=418{,}178$) and control ($n=46{,}345$) groups by the utility through the minmax *t*-statistic algorithm, depending on baseline consumption and geographical location[32]. The small relative size of the control group was determined by the utility in collaboration with Opower, with the goal to minimize the number of customers who did not receive the programme, but avoid issues of statistical power in the evaluation of its impact. The experimental design could eventually capture a minimum detectable effect with 90% power and 5% significance of about 0.36%. As for heterogeneous effects by pretreatment usage, the minimum detectable effects ranged from 0.8 to 1.2%. These effect sizes are relatively small with respect to those found in the literature[19,33–35].

The programme was rolled-out in three waves: July (39% of treated customers), October (33%) and December 2016 (28%). After that, customers received reports bimonthly. Conditional on being assigned to the treated group, 91% of customers received at least one eHER, on average 8.8 over the first 24 months of the programme. In April and May 2018, the report augmented with the experimental prime was sent to 256,487 programme participants, randomly assigned to receive the descriptive ($n = 127,899$) or injunctive ($n = 128,588$) message. The sample of customers who received the augmented eHER was smaller than the entire sample of programme participants at that time ($n = 348,131$)—as the implementation of the programme was not under the research team's control, we cannot document the reasons for such a discrepancy. We did, however, examine individual correlates of the receipt of the augmented eHER (Supplementary Note 5 and Supplementary Table 15).

About 21.5% of customers left the dataset between the launch of the programme and August 2018, primarily due to termination of the contract with the utility. We checked that the attrition was non-differential by treatment status using a two-sided t-test (Supplementary Table 1). In Supplementary Table 1, we also report a similar test as to whether attrition in the 3 months after the delivery of the augmented eHER, equal to about 5.8%, correlated with the experimental prime treatment. We explored individual correlates of attrition at different points in time using regression analysis (Supplementary Table 16). These allowed us to assess the extent to which attrition was a threat to the internal and external validity.

Data. We had access to historical electricity consumption data from July 2015 to December 2019 for all the customers. These data were provided by the utility, after being verified by the electricity distributor, and constituted our main outcome variable. Similarly to other works, we computed the average daily electricity usage in a month from the total monthly consumption and normalized it by dividing by the average post-period control group consumption and then multiplying by 100 (ref. [17]). We also computed the average daily pretreatment electricity usage as the mean of the average daily consumption in a month between July 2015 and June 2016.

We also had access to information on the treated customers' engagement with the reports and on the reports' contents. These data were provided by Opower. We knew when a eHER was sent and when a customer opened or clicked on a report, although we did not know which one. By opening the reports, customers were able to view the neighbour comparison; by clicking on it, customers were directed to their personal page on the utility's website, where further information, such as energy saving tips and bills, was available. On average, 64 and 30% of the customers opened and clicked on an eHER, respectively, at least once over the two years after the programme launch. As for the eHER augmented with the normative prime, 55 and 9.7% of customers opened and clicked at least once in the two months after its receipt, respectively.

The reports' content data include customers' relative performance within the reference group and the type of feedback they received within each eHER, in terms of number of thumbs up. In general, considering all the reports in the first 24 months, customers received 20% of reports with three thumbs up, 34% with two thumbs up and 45% with one thumbs up, consistent with the definition of the three feedback categories. Overall, 38% of customers received the same feedback throughout the period, and the remaining experienced some change. In the majority of cases, customers experienced both upgrades and downgrades. Specifically to the eHER augmented with the randomized prime, 15.1% of customers received three thumbs up, 36.1% received two thumbs up and 48.8% one thumbs up. As time-invariant controls, we used dummies for the main geographical areas in Italy, that is, the northeast, northwest, central, south and the islands, and the population of the municipality in which the customers lived, obtained by matching the contract municipality data with data on municipalities' characteristics[36]. We missed information on the geographical location of 5,675 customers (about 1.2% of the sample), equally distributed by treatment status ($P = 0.273$). This reduced the sample size to 459,088 customers whenever the analysis featured geographical controls.

Finally, we used data from an online survey conducted with a representative subsample of about 3,000 utility customers in April 2017 to inform the design of the norm primes. In particular, questions on actual pro-environmental behaviour—such as turning off the lights when leaving a room and hanging the clothes to dry instead of using dryer—were used to design the descriptive prime, whereas questions on personal values related to energy conservation—such as whether the customer feels personally responsible to try to save energy, whether the customer would act according to the customer's principles if energy was saved and whether the customer feels morally obliged to save energy—informed the design of the injunctive prime. The survey questions used to elicit personal values and norms were taken from established survey instruments, such as the World Values Survey, and from published studies in environmental psychology[37,38]. More details on the survey can be found in Bonan et al.[22].

Balance. We checked that the Opower treatment and control group, and the samples of customers assigned to the descriptive or injunctive prime, were balanced across the observable characteristics through two-sided t-tests (Supplementary Table 1). In addition, we tested for the randomization balance across the subgroups identified by the combination of the April–May 2018 normative prime treatments and the November–December 2017 environmental prime treatments, mentioned above (Supplementary Table 2). This was done through an F-test of joint significance of subtreatment dummies regressed on the observable characteristics. We confirmed that balance generally holds.

Impact of the programme. We evaluated the impact of the programme by estimating the following model:

$$y_{it} = \beta_1 \mathrm{Post}_{it} + \beta_2 \mathrm{Program}_i \times \mathrm{Post}_{it} + h_t + g_i + \varepsilon_{it} \tag{1}$$

where y_{it} is the customer's i normalized average daily consumption in month t. Program is a treatment indicator, Post is a dummy variable which becomes 1 when customers receive the first eHER. The coefficient β_1 captures the effect of any time variant factors affecting consumption after the start of the programme, while β_2 isolates the impact of the programme on treated customers. $\mathcal{E}_{it}$ is the error term. Given the staggered phase-in of the programme, to allow the identification through difference-in-differences, we randomly assigned control customers to the three programme start waves, with the same proportions as treated customers. This implies that Post becomes 1 at the beginning of each wave for the same share of treated and control customers. The regression also included month-by-year fixed effects, h_t, and household fixed effects, g_i. Standard errors were clustered at the level of household to allow for the presence of within-customer correlation over time in the error term[39]. The average treatment effects can be interpreted as the percentage change (Supplementary Table 3).

We examined the differential response to the eHER depending on pretreatment electricity use by interacting it with the Program × Post dummy. We expressed electricity usage with a continuous variable and with dummies for the quartiles. To allow for differential post-treatment trends, we also interacted the Post dummy with pretreatment usage. To adjust for multiple hypothesis testing, in the subgroup analysis we computed the sharpened two-stage q values (FDR-adjusted q values)[40]. We analysed ex post power by calculating the minimum detectable effects with a 90% power and 5% significance[41]. As a robustness check, we estimated the main and heterogeneous effects of the programme on the subsample of customers for whom information on geographical location was available (Supplementary Table 4).

We evaluated the main and heterogeneous impacts of social information over the first and the first two years of the programme. In a further analysis, we extended the treatment period until December 2019 and explored in greater detail the persistence of the treatment effects (Supplementary Tables 5–7).

As in other works[6,17], our treatment effects are intent-to-treat estimates computed on the full sample of eligible customers, regardless of whether they opted out of the programme or did not open the reports. We kept customers who did not receive or read reports in the analysis to maintain the balance between the treatment and control group and avoid selection issues affecting our results. By doing this, we were likely to underestimate the effect of the programme on the group of customers initially assigned to receive the eHER and who actually saw the treatment communication. In an additional analysis, we examined the role of the engagement with the programme on treatment impacts by instrumenting opening the eHER and clicking on it with the treatment status, and reporting the local average treatment effect (Supplementary Table 8). We did this in cross-section. The outcome variable was the normalized average electricity usage in the 24 months after the launch of the programme. The specifications also included pretreatment electricity usage.

We further exploited our data on the content of the reports, specifically on the number of thumbs up received within each eHER, to examine heterogeneous effects of the treatment depending on whether the customers were upgraded or downgraded with respect to the previous report. We did it by restricting the analysis to the months when the eHERs were sent and by focusing on normalized daily electricity usage in the 3 months after each report. Specifications include customer and month-by-year fixed effects (Supplementary Table 9).

Regression discontinuity estimation. We used RD to estimate the impact of changes in the injunctive feedback included in the neighbour comparison. We justified this approach by showing that customers who received one, two or three thumbs up within the April–May 2018 eHER were different in many respects, which may also correlate with the impact of the treatment (Supplementary Table 10). The RD approach allowed us to eliminate the influence of confounding factors on the estimated effect of changing the feedback category, as it focused on customers for whom the number of thumbs up is random. The price we paid for the improved identification of the effects is that the impacts estimated through RD are local, specific to a neighbourhood of the thresholds.

In our RD framework, the running variables, X_{i1} and X_{i2}, are the customer's i monthly electricity usage (kWh). The cutoffs, c_{i1} and c_{i2}, are the electricity usage of the 15th percentile and the overall average electricity usage among the neighbours, respectively. The assignment variable T_{i1} takes the value of 1 when the customer's usage lies above c_{i1} ($X_{i1} > c_{i1}$) and 0 otherwise. Similarly, T_{21} takes the value of 1 when the customer's usage lies between c_{i1} and c_{i2} and 0 otherwise. We estimated the equations:

$$Y_i = \lambda_0^1 + \eta^1 T_{i1} + f(X_{i1} - c_{i1}) + \varepsilon_1^1 \tag{2A}$$

$$Y_i = \lambda_0^2 + \eta^2 T_{i2} + f(X_{i2} - c_{i2}) + \varepsilon_1^2 \tag{2B}$$

where Y_i is the customer's i-normalized average daily electricity usage in the 3 months after the receipt of the report that contained the normative prime. Model (2A) is estimated from customers who received either three or two thumbs up (for example, with $X_{i1} > c_{i1}$ or $c_{i1} < X_{i1} < c_{i2}$), whereas model (2B) includes those who received either two or one thumbs up (for example, $X_{i1} < c_{i1}$). The parameter λ_0 is a constant, while η captures the effect of obtaining an extra thumbs up. ε_1^1 and ε_1^2 are the error terms. Our main specification used a non-parametric approach (local polynomial point estimation), which amounts to fitting two linear regressions on customers respectively close to the left and to the right of the cutoff[42,43]. This approach used only observations within a BW from the cutoff. The selection of the BW was data driven and aims to minimize the m.s.e. of the local polynomial RD estimator. We computed the point estimates using both one common BW on both sides of the cutoff and two distinct BWs. We used a linear polynomial for $f(.)$, which represents a good trade-off between simplicity, precision and stability[44]. Observations within the BW were weighted through a triangular kernel function, which assigned zero weight to observations outside the BW and positive weights within it, which decreased symmetrically and linearly as the distance from the cutoff increased. The triangular kernel had desirable asymptotic optimality properties[44]. We show point estimates under conventional (from a parametric least-squares estimation) and bias-corrected robust standard errors, which deliver a valid inference under the m.s.e.-optimal BW selection[43] (Table 1).

Identification within RD design relies on the assumption that customers have no full control over the running variable and hence are unable to manipulate their position in the distribution of electricity usage within their reference group. Manipulation should not be a threat to our identification for several reasons. First, customers do not know the algorithm for the selection of the reference group and assignment of thumbs up. Second, although people may be able to control their own electricity usage, they have no control on their neighbours' consumption, which determines the unpredictability of the consumption level associated with the different cutoffs every round. Third, people do not gain any direct benefit from an extra thumbs up, apart from individual satisfaction, and hence the incentives to game the system should be minimal. We tested this identification assumption by checking for discontinuities in the density of the running variable at the thresholds, using a non-parametric density estimator based on local polynomial techniques[45]. We did the same with the pretreatment covariates, following a procedure suggested in the literature[46] (Supplementary Table 11).

We also performed an RD estimation on the probability that customers opened or clicked in the April–May 2018 eHER. This analysis aimed to support the claim that the effects we found on consumption were due to the neighbour comparison, rather than to discontinuous changes in unobserved determinants of open rates or the likelihood of viewing electricity saving tips at the thresholds (Supplementary Table 12).

We repeated the test of the RD identifying assumption and the RD estimation using a standard eHER, delivered in February–March 2018, to rule out that our results are due to the presence of the normative primes (Supplementary Note 6 and Supplementary Tables 17 and 2) and RD effects over the following 6 and 12 months (Supplementary Table 13).

Impact of the normative primes. We repeated the RD estimation separately on the subsamples of customers randomly allocated to the descriptive and injunctive primes to obtain local marginal effects of moving across the injunctive feedback categories by prime treatment. We adjusted for multiple comparisons[40].

We also assessed the impact of receiving the eHER augmented by the descriptive versus injunctive prime, by estimating the following model:

$$y_{it} = \alpha_1 \mathrm{Post}_{it} + \alpha_2 \mathrm{Descriptive\_norm}_i \times \mathrm{Post}_{it} + h_t + g_i + \mu_{it} \qquad (3)$$

Descriptive_norm takes the value of 1 for the descriptive prime and 0 for the injunctive prime. Post becomes one when customers receive the eHER that contains the prime (roughly half in April and half in May 2018). The coefficient α_1 captures the change in consumption after receiving the eHER for customers assigned to receive the injunctive prime, while α_2 reveals any differential effect of receiving the descriptive prime. Standard errors are clustered at the household level (Table 3). The regression was run on the sample of treated customers at that time ($n = 256{,}487$) and considered the period that spanned from October 2017 to August 2018, but we also explored the persistence of the effects until December 2019, that is, 18 months after the prime (Supplementary Table 14). We also estimated equation (3) for the subsamples of customers who received one, two or three thumbs up, overall and separately for customers upgraded or downgraded with respect to the previous report, to examine the heterogeneous impact of the augmented eHER.

We tested that these results were not due to interaction effects with the November–December 2017 environmental identity test by adding to equation (3) an interaction term between Post, Descriptive_norm and a dummy equal to 1 if a customer was assigned to the environmental prime treatment (Supplementary Table 18).

Preregistration. We analysed the impact of the standard and augmented eHER using non-experimental RD techniques. In terms of individual level data, our analysis relied only on customers' consumption, location and content of the eHER. For these reasons, the analysis we present was not prespecified in a pre-analysis plan: detailed prespecification is warranted when, within the impact evaluation of field experiments, subgroup analysis is expected to be important and there is the possibility to cherry-pick the dimensions of heterogeneity to focus on or that a party to the study has a vested interest[47]. These conditions did not apply to our case.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The data that support the findings of this study are proprietary data of the energy company and cannot be shared publicly. To inquire about access to the proprietary data, please contact M.T.

Code availability

The replication code is available on Open Science Framework at https://osf.io/wz8gb/.

Received: 30 April 2020; Accepted: 28 September 2020;
Published online: 2 November 2020

References

1. Frey, B. S. & Meier, S. Social comparisons and pro-social behavior: testing 'conditional coop-eration' in a field experiment. *Am. Econ. Rev.* **94**, 1717–1722 (2004).
2. Shang, J. & Croson, R. A field experiment in charitable contribution: the impact of social information on the voluntary provision of public goods. *Econ. J.* **119**, 1422–1439 (2009).
3. Gerber, A. S. & Rogers, T. Descriptive social norms and motivation to vote: everybody's voting and so should you. *J. Polit.* **71**, 178–191 (2009).
4. Beshears, J., Choi, J. J., Laibson, D., Madrian, B. C. & Milkman, K. L. The effect of providing peer information on retirement savings decisions. *J. Finance* **70**, 1161–1201 (2015).
5. Allcott, H., Mullainathan, S. & Taubinsky, D. *Externalizing the Internality* (New York Univ., 2011).
6. Allcott, H. & Rogers, T. The short-run and long-run effects of behavioral interventions: experimental evidence from energy conservation. *Am. Econ. Rev.* **104**, 3003–3037 (2014).
7. Ayres, I., Raseman, S. & Shih, A. Evidence from two large field experiments that peer comparison feedback can reduce residential energy usage. *J. Law Econ. Organ.* **29**, 992–1022 (2013).
8. Brent, D. A., Cook, J. H. & Olsen, S. Social comparisons, household water use, and participation in utility conservation programs: evidence from three randomized trials. *J. Assoc. Environ. Resour. Econ.* **2**, 597–627 (2015).
9. Ferraro, P. J. & Price, M. K. Using nonpecuniary strategies to influence behavior: evidence from a large-scale field experiment. *Rev. Econ. Stat.* **95**, 64–73 (2013).
10. Ferraro, P. J., Miranda, J. J. & Price, M. K. The persistence of treatment effects with norm-based policy instruments: evidence from a randomized environmental policy experiment. *Am. Econ. Rev.* **101**, 318–322 (2011).
11. Ferraro, P. J. & Miranda, J. J. Heterogeneous treatment effects and mechanisms in information-based environmental policies: evidence from a large-scale field experiment. *Resour. Energy Econ.* **35**, 356–379 (2013).
12. Jaime Torres, M. M. & Carlsson, F. Direct and spillover effects of a social information campaign on residential water-savings. *J. Environ. Econ. Manag.* **92**, 222–243 (2018).
13. Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J. & Griskevicius, V. The constructive, destructive, and reconstructive power of social norms. *Psychol. Sci.* **18**, 429–434 (2007).
14. Bicchieri, C. & Xiao, E. Do the right thing: but only if others do so. *J. Behav. Decis. Mak.* **22**, 191–208 (2009).
15. Meisel, M. K. & Goodie, A. S. Descriptive and injunctive social norms' interactive role in gambling behavior. *Psychol. Addict. Behav.* **28**, 592–598 (2014).
16. Biccheri, C. & Dimant, E. Nudging with care: the risks and benefits of social information. Public Choice https://doi.org/10.1007/s11127-019-00684-6 (2019).
17. Allcott, H. Social norms and energy conservation. *J. Public Econ.* **95**, 1082–1095 (2011).
18. Jachimowicz, J. M., Hauser, O. P., O'Brien, J. D., Sherman, E. & Galinsky, A. D. The critical role of second-order normative beliefs in predicting energy conservation. *Nat. Hum. Behav.* **2**, 757–764 (2018).
19. Andor, M. A., Gerster, A., Peters, J. & Schmidt, C. M. Social norms and energy conservation beyond the US. *J. Environ. Econ. Manag.* **103**, 102351 (2010).
20. Byrne, D. P., Nauze, A. L. & Martin, L. A. Tell me something I don't already know: informedness and the impact of information programs. *Rev. Econ. Stat.* **100**, 510–527 (2018).

21. Costa, D. L. & Kahn, M. E. Energy conservation 'nudges' and environmentalist ideology: evidence from a randomized residential electricity field experiment. *J. Eur. Econ. Assoc.* **11**, 680–702 (2013).
22. Bonan, J., Cattaneo, C., D'Adda, G. & Tavoni, M. *Can We Make Social Information Programs More Effective? The Role of Identity and Values* Working Paper 19-21 (Resources for the Future, 2019).
23. Bursztyn, L., González, A. L. & Yanagizawa-Drott, D. *Misperceived Social Norms: Female Labor Force Participation in Saudi Arabia* Working Paper 24736 (National Bureau of Economic Research, 2018); https://doi.org/10.3386/w24736
24. D'Adda, G., Dufwenberg, M., Passarelli, F. & Tabellini, G. *Partial Norms* SSRN Scholarly Paper ID 3362021 (Social Science Research Network, 2019).
25. List, J. A., Metcalfe, R. D., Price, M. K. & Rundhammer, F. *Harnessing Policy Complementarities to Conserve Energy: Evidence from a Natural Field Experiment* Working Paper 23355 (National Bureau of Economic Research, 2017); https://doi.org/10.3386/w23355.00000
26. Hallsworth, M., List, J. A., Metcalfe, R. D. & Vlaev, I. The behavioralist as tax collector: using natural field experiments to enhance tax compliance. *J. Public Econ.* **148**, 14–31 (2017).
27. Bobek, D. D., Hageman, A. M. & Kelliher, C. F. Analyzing the role of social norms in tax compliance behavior. *J. Bus. Ethics* **115**, 451–468 (2013).
28. Krupka, E. L. & Croson, R. T. A. The differential impact of social norms cues on charitable contributions. *J. Econ. Behav. Organ.* **128**, 149–158 (2016).
29. Bahnot, S. Isolating the effect of injunctive norms on conservation behavior: new evidence from a field experiment in California. Organ. Behav. Hum. Decis. Process https://doi.org/10.1016/j.obhdp.2018.11.002 (2018).
30. Allcott, H. & Taubinsky, D. Evaluating behaviorally motivated policy: experimental evidence from the lightbulb market. *Am. Econ. Rev.* **105**, 2501–2538 (2015).
31. Allcott, H. Site selection bias in program evaluation. *Q. J. Econ.* **130**, 1117–1165 (2015).
32. Bruhn, M. & McKenzie, D. In pursuit of balance: randomization in practice in development field experiments. *Am. Econ. J. Appl. Econ.* **1**, 200–232 (2009).
33. Andor, M. A. & Fels, K. M. Behavioral economics and energy conservation—a systematic review of non-price interventions and their causal effects. *Ecol. Econ.* **148**, 178–210 (2018).
34. Farrow, K., Grolleau, G. & Ibanez, L. Social norms and pro-environmental behavior: a review of the evidence. *Ecol. Econ.* **140**, 1–13 (2017).
35. Delmas, M. A., Fischlein, M. & Asensio, O. I. Information strategies and energy conservation behavior: a meta-analysis of experimental studies from 1975 to 2012. *Energy Policy* **61**, 729–739 (2013).
36. Guiso, L., Sapienza, P. & Zingales, L. Long-term persistence. *J. Eur. Econ. Assoc.* **14**, 1401–1436 (2016).
37. Steg, L., Perlaviciute, G., van der Werff, E. & Lurvink, J. The significance of hedonic values for environmentally relevant attitudes, preferences, and actions. *Environ. Behav.* **46**, 163–192 (2014).
38. van der Werff, E., Steg, L. & Keizer, K. The value of environmental self-identity: the relationship between biospheric values, environmental self-identity and environmental preferences, intentions and behaviour. *J. Environ. Psychol.* **34**, 55–63 (2013).
39. Bertrand, M., Duflo, E. & Mullainathan, S. How much should we trust differences-in-differences estimates? *Q. J. Econ.* **119**, 249–275 (2004).
40. Benjamini, Y., Krieger, A. M. & Yekutieli, D. Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* **93**, 491–507 (2006).
41. Haushofer, J. & Shapiro, J. The short-term impact of unconditional cash transfers to the poor: experimental evidence from Kenya. *Q. J. Econ.* **131**, 1973–2042 (2016).
42. Imbens, G. W. & Lemieux, T. Regression discontinuity designs: a guide to practice. *J. Econom.* **142**, 615–635 (2008).
43. Calonico, S., Cattaneo, M. D. & Titiunik, R. Robust nonparametric confidence intervals for regression- discontinuity designs. *Econometrica* **82**, 2295–2326 (2014).
44. Cattaneo, M. D., Idrobo, N. & Titiunik, R. *A Practical Introduction to Regression Discontinuity Designs: Foundations* (Cambridge University Press, 2019).
45. Cattaneo, M. D., Jansson, M. & Ma, X. Simple local polynomial density estimators. *J. Am. Stat. Assoc.* **0**, 1–7 (2019).
46. Canay, I. A. & Kamat, V. Approximate permutation tests and induced order statistics in the regression discontinuity design. *Rev. Econ. Stud.* **85**, 1577–1608 (2018).
47. Banerjee, A. et al. *In Praise of Moderation: Suggestions for the Scope and Use of Pre-analysis Plans for RCTs in Economics* Working Paper w26993 (National Bureau of Economic Research, 2020).
48. Calonico, S., Cattaneo, M. D. & Titiunik, R. Robust nonparametric confidence intervals for regression–discontinuity designs. *Econometrica* **82**, 2295–2326 (2014).
49. Calonico, S., Cattaneo, M. D., Farrell, M. H. & Titiunik, R. Regression discontinuity designs using covariates. *Rev. Econ. Stat.* **101**, 442–451 (2019).

Acknowledgements

The research leading to these results received funding from the European Research Council under the European Union's Seventh Framework Program (FP7/2007-2013)/ ERC grant agreement no. 336155—project COBHAM, 'The role of consumer behaviour and heterogeneity in the integrated assessment of energy and climate policies'.

Author contributions

J.B., C.C., G.D. and M.T. conceived and designed the experiments. J.B. analysed the data. J.B. and G.D. contributed the analysis tools. J.B., C.C., G.D. and M.T. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41560-020-00719-z.

Correspondence and requests for materials should be addressed to G.d.

nature research

Corresponding author(s): Giovanna d'Adda

Last updated by author(s): May 5, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used for data collection.
Data analysis	Stata 15 was used for all statistical analysis and for the production of graphs.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data that support the findings of this study were used under license for the current study. Because of the privacy rules in Europe and the non-disclosure agreement with company, the data are not publicly available. Data are however available from the authors upon reasonable request and with permission of the company. In addition, to ensure that the results can be replicated, the code is available upon request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	The study uses quantitative data generated by a field experiment (Randomized Control Trial): the dataset is a panel including monthly electricity consumption and details on the content of email communication sent to a sample of utility customers.
Research sample	Our sample of analysis is represented by 464,763 customers eligible for the the social information program and with non-missing electricity consumption over the entire period of analysis. We have access to this sample thanks to an existing agreement with a large Italian energy utility. Since eligibility requirements include availability of an email contact, we imagine that the sample is not fully representative of the entire customer base of the utility nor of the entire population of energy users in Italy. The administrative data we use is restricted to the RCT sample, and only contains information on customers' consumption and area of residence: they thus do not allow us to conduct formal tests of representativeness. However, the size of the sample guarantees that our results are generalizable to important, and growing, segments of the population.
Sampling strategy	The sample used in the study is the entire population of customers targeted by the utility's social information program at the time the program was launched. The size of the sample, eligibility criteria and randomization into treatments were all determined by the utility and the private consultants that designed and implemented the social information program. Where the research team had any input on the sampling, i.e., in the design of the normative prime experiment, we requested that treatment and control groups be of the same sizes, in order to maximise statistical power.
Data collection	The data used in the study are exclusively administrative data collected directly by the partner utility, and transferred to the research team through secure data sharing protocols.
Timing	Consumption data are available from July 2015 to August 2018. Data on the content of program communication are available for the duration of the program: July 2016 to August 2018.
Data exclusions	No data were excluded from the analysis.
Non-participation	Non-participation in the study occurrs when users terminate their contract with the utility. Even though attrition is non-differential by treatment, we keep attriters in the sample in order to avoid any selection bias affecting the results.
Randomization	Randomization was implemented by the utility through an algorithm (minmax t-statistic), which conducts 1000 randomizations and selects the most balanced draw, along baseline consumption and geographic location (Bruhn and McKenzie 2009).

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study	
☒	☐	Antibodies
☒	☐	Eukaryotic cell lines
☒	☐	Palaeontology and archaeology
☒	☐	Animals and other organisms
☐	☒	Human research participants
☒	☐	Clinical data
☒	☐	Dual use research of concern

Methods

n/a	Involved in the study	
☒	☐	ChIP-seq
☒	☐	Flow cytometry
☒	☐	MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	We only have access to information on the location of utility customers, at the regional level, and on their consumption before the start of the program. The geographical distribution of users is as follows: 29 per cent from North-West Italy, 12.2

per cent from the North East, 31.2 per cent from the Center, and the remaining from the South and Islands. Average baseline consumption is 6.3 kWh.

Recruitment	Inclusion in the sample is based exclusively on technical eligibility criteria set by the utility: households must have a valid name and email address as of June 2016, live in single-family homes, have at least one year of valid pre-experiment energy consumption data, have no negative electricity meter reads, at least one meter read in the previous three months, no significant gaps nor extreme peaks in usage history, and exactly one account per customer per location. While these criteria may result in the study sample not being perfectly representative of the full sample of utility customers, no self-selection in the study is present.
Ethics oversight	We obtained ethics approval for the study from Milan Polytechnic's Ethics Committee, with approval number 4/2017.

Note that full information on the approval of the study protocol must also be provided in the manuscript.