

Environmental self-identity and the impact of social information on energy use: experimental evidence

Pre-Analysis Plan

Jacopo Bonan, Giovanna d’Adda, Cristina Cattaneo, Massimo Tavoni

Fieldwork locations: Italy
Fieldwork dates: November 2017 to June 2018
Date of Pre-Analysis Plan: January 26, 2018

1 Introduction

This document outlines our pre-analysis plan for a field experiment on the impact of social information on energy consumption, to be conducted with customers of a large Italian electric utility. The document summarizes (i) our experiment and resulting data and (ii) our research questions and the plan of regressions.

In preparation for the experiment, we collected data from a subsample of EGL customers in municipalities throughout Italy from April to June 2017. At the time of writing this plan, we have access to the survey data on individual characteristics and energy use, which we have analyzed to inform the experimental treatments.

We intend to submit this Pre-Analysis Plan to the AEA RCT Registry.

2 Sample and treatment

2.1 Description of the sample

The project is centered around a customer engagement and energy efficiency program conducted with utility's customers, which kicked-off in July 2016. The program targets roughly 450,000 existing customers, making up the current pool of power or dual fuel customers: this group is referred to as the customer base (CB) and is our primary sample.

To be eligible for the program, households must be single-family homes, have at least one to two years of valid pre-experiment energy consumption data, and satisfy some additional technical conditions. In particular, households need to have valid names and email addresses, no negative electricity meter reads, at least one meter read in the last three months, no significant gaps in usage history, exactly one account per customer per location, and a sufficient number of neighbors to construct the neighbor comparisons. Moreover, customers with relatively too low or high electric usage history are not eligible.

A total of 447,376 eligible customers have been initially included in the experimental sample, of which 402,688 and 44,688 have been randomly assigned to the treatment and control group, respectively.

2.2 Structure of the treatment and randomization

The main goal of the program is to increase loyalty, digitalization and engagement of customers. Energy efficiency goals are secondary. The intervention follows the typical design of Opower programs, already described and evaluated by several papers ([Allcott, 2011](#); [Allcott and Rogers, 2014](#)). The program consists of the following elements:

- Home Energy Report (eHER): a treatment group from CB customers receives an eHER every two months via email. The eHER features the following information:
 - Static neighbor comparison: comparison of one's own previous month consumption with that of about 100 similar homes nearby and 20 most efficient similar homes nearby.
 - Feedback on energy consumption (thumbs up).
 - Marketing module: this is a space that can contain season specific messages, or messages aimed at drawing customers' attention to specific features of the Genius suite, such as the energy saving tips or the energy profile. Our experimental treatments feature in the eHER as marketing modules.

- Link to the web portal.
- The web portal: this is available to all customers, regardless of whether they are in the treatment or control group, as long as they are registered to the website. It features:
 - Static neighbor comparison, as in the eHER.
 - Feedback on energy consumption.
 - Dynamic neighbor comparison: overview of one’s own monthly energy consumption, broken down by time of day tariffs, for the past 12 months. Includes comparison with 100 similar homes and 20 most efficient similar homes nearby.
 - Comparison of own consumption over previous month and same month of previous year.
 - Energy saving tips: tips on how to save energy, divided into categories (behavior change, small investment, large investment). Each tip includes an estimate of saving potential.
 - Energy profile: survey of household characteristics and appliances, with the goal of targeting tips, refining the neighbor comparison, and in the future target sales campaigns.

Therefore, the general program relies on an encouragement design: only customers in the treatment group receive the eHER, but the information is available to everyone on the website. The treatment group, selected through a randomization algorithm (minmax t-statistic) which matches customers on the basis of baseline consumption and geographic location (Bruhn and McKenzie, 2009), receives the energy report bimonthly by email, while the control group receives nothing.

Our project involved a set of preliminary activities, which have informed the intervention we implement in the RCT. The RCT aims to understand the mechanisms behind the effect of the program, and is the subject of the present PAP.

In particular, as part of the preliminary activities, we conducted a survey on a sub-sample of our partner utility’s power and dual fuel customers (CB customers). In order to maximize the level of sample representativeness, we first identified a sub-sample of CB customers. This sub-sample was constructed in such a way to have the same shares of customers with the following characteristics as the reference population: gender of the contract holder, age (defined in five brackets), geographical area of residence (North-West; North-East; Center; South), yearly energy consumption (defined in four brackets). The shares were imposed to be the same for treatment and control group. Based on this stratification, a total of 155,691 customers were contacted via email and invited to participate to an online survey. The survey data were combined with administrative data on treatment assignment and engagement with the program suite to

investigate the impact of the program on:

- Energy consumption
- Determinants of energy consumption: investment and behavior change
- Customer engagement
- Awareness and knowledge
- Other energy or environmental conservation behaviors
- Environmental identity
- Social norms

The heterogeneous effects of the program were tested along several dimensions:

- Baseline endowment of electric appliances
- Pre-treatment energy consumption
- Baseline time invariant household characteristics: composition, level of education, income, ownership vs rent of the house, time living in the neighborhood, size of the house
- Environmental values
- Personality traits: big 5, self-efficacy
- Social capital: trust, civic engagement, cooperation
- Social identity: perceived similarity, identification with neighbors

The main focus of the present project is the question on the mechanisms behind the effect of the program. Starting from the factors found to be significantly correlated with program effectiveness and customer engagement in the analysis conducted on the program and survey data, we designed an RCT, whose treatment makes one of this factor more relevant or salient. In particular, the exploratory analysis revealed that environmental values are significantly correlated with the impact of the eHER within the sample of customers surveyed. This PAP therefore focuses on the treatment we designed to examine the role of environmental values.

In the preliminary analysis, we found that customers with stronger environmental values are more responsive to the treatment. Values are antecedents of environmental preferences, intentions, and behaviour and guide individual principles. While values tend to be stable in everyone's life, environmental values are correlated with environmental self-identity, which, on the contrary, can be primed (van der Werff et al., 2013). Therefore, we designed a treatment message aimed at priming environmental self-identity. The treatment message is included in the November/December eHER. Comparison of outcomes for customers exposed to the environmental self-identity priming relative to control subjects, will reveal the role of environmental values on the program impact.

To identify the most effective messages to prime environmental identity to include in the RCT, we pre-tested a battery of different messages through an online experiment conducted on Prolific Academic. Subjects in the experiment were exposed to one of different messages, priming environmental self-identity, and then were asked to donate to an environmental NGO and to read a set of energy saving tips. The pre-test also included manipulation checks for the primes, so as to make sure that their impact on pro-environmental behavior worked through their effect on environmental self-identity.

The most effective message, in terms of both donation levels and perceived self-identity measures in the pre-test, was one that reminded individuals of the ways in which they already saved energy in their daily lives, boosting their self-image as pro-environmental people. We thus selected this message to be included in the eHER marketing module, followed by a request to find more ways to save energy by consulting the energy saving tips contained in the program portal. The control message simply asks customers to find ways to save energy by reading the tips. In the analysis, we will compare engagement with the eHER and energy saving tips, and energy consumption, between customers exposed to the eHER enriched by the environmental self-identity message, customers exposed to the standard eHER, and control customers receiving no eHER. The assignment of customers in the program to the group receiving the eHER augmented with the environmental identity prime and the control message was done through a randomization algorithm (minmax t-statistic) which matches customers on the basis of baseline consumption and geographic location (Bruhn and McKenzie, 2009). The sample of customers still enrolled in the program and who actually received the eHER containing the experimental messages is 278,252 customers (according to the data we currently dispose). About half of the group (49.9%) received the environmental priming message and the remaining 50,1% received the control one.

3 Research questions and analysis

3.1 Experimental analysis

By drawing attention on energy use, the eHER's may increase the moral cost of consuming energy. One channel, through which this effect can work, is that the eHER may activate biospheric values and strengthen people's focus on benefiting the environment. To shed light on the role of this mechanism, we assess the effect of receiving the standard eHER on engagement and consumption, and the additional effect on these outcomes of augmenting the eHER with a message priming environmental self-identity. The study addresses the following research questions. For each of them the specification, the test of hypothesis and the sample of analysis are indicated.

Research Question 1 *What is the impact of the standard eHER and of the eHER augmented with the environmental self-identity prime on customer engagement and energy consumption?*

To answer this question, we estimate the effect of the receipt of the November-December eHER on digital engagement with the program portal and consumption. The engagement with eHER data will be the primary outcome indicators for our analysis of the impact of our treatment, given that they are messages built in the eHER.

We use the sample of customers enrolled in the program who received the marketing module during the November-December wave for a time window running from November 2016 to April 2018:

$$Engage_{it} = \alpha + \beta_2 Post_t + \beta_3 Prime_i \times Post_t + h_m + g_i + \varepsilon_{it} \quad (1)$$

where $Engage_{it}$ is a set of variables related to digital engagement and $Post_t$ is a dummy which becomes one the month following the receipt of the eHER belonging to the November/December experimental session and $Prime_i$ is a dummy equal to one for treated customers receiving the eHer augmented with the environmental self-identity prime. h_m and g_i are time and individual fixed effects, respectively.

To estimate the impact on consumption, following [Allcott and Rogers \(2014\)](#), we consider three periods. Period 0 is the pre-treatment period (July 2015-July 2016), period 1 is the period during which eHER is implemented alone (July 2016-November 2017), period 2 is the post-prime period and follows the delivery of the eHER containing the environmental marketing module (November 2017-April 2018). We denote I_m^p an indicator variable for whether month m is in period p . We estimate the impact on energy consumption

on the whole sample of customers for a time window going from July 2015 to April 2018:

$$EnCons_{itm} = (\tau^1 I_m^1) \times Program_i + (\tau^2 Program_i + \alpha^2 Prime_i) \times I_m^2 + h_m + g_i + \varepsilon_{itm} \quad (2)$$

where $Program_i$ is equal to one for customers enrolled in the engagement program. The first line identifies the main effects of receiving the eHER in Period 0 and 1 before the prime was sent, while the second line identifies the treatment effect for the group of households receiving the eHER augmented with the social identity prime, in the post-prompt period.

Research Question 2 *How do treatment effects vary with respect to baseline household characteristics? Which customers' respond more to the standard and augmented Genius communication?*

This research question is addressed through the analysis of treatment effects heterogeneity along dimensions which are deemed relevant in the literature and of which we have data collected during the first wave of the survey. The outcomes of interest are the same considered in research question 1. The following empirical specification is estimated on the sample of customers included in the first wave of the survey for a time window going from July 2015 to April 2018:

$$Y_{itm} = (\tau^1 I_m^1) \times Program_i + (\tau^2 Program_i + \alpha^2 Prime_i) \times I_m^2 \times X_i + h_m + g_i + \varepsilon_{itm} \quad (3)$$

where X is one of the following dimensions:

- Pre-program energy consumption
- Baseline endowment of electric appliances
- Baseline time invariant household characteristics: gender, age, composition, contract tenure, level of education, household income, time living in the neighborhood, size of the house
- Environmental values
- Geographical area (regional comparison: north - centre - south)

Through this analysis we aim to confirm the results obtained examining the heterogeneity of the standard program message on the basis of the baseline survey data; and to extend them to the evaluation of heterogeneity of the program message augmented by our environmental identity marketing module.

The main findings of the heterogeneity analysis that we conducted using the baseline survey data is that pre-treatment usage along with environmental values are the major sources of heterogeneity. Families with higher pre-consumption respond to the intervention by curbing energy consumption. This is because the scope for energy conservation is larger for high-usage households. The second source of heterogeneity is related to environmental values. Among families with high pre-consumption, those who display high environmental values response more strongly to the program in terms of energy curtailment. This result provides an insight on a possible mechanism driving the effectiveness of the program. Environmental values are an important moderator of the effect of the program on consumption. Since the eHER may increase the moral cost of energy use, such effect may be stronger, the stronger a person's environmental values. Therefore, we expect that further priming environmental self-identity through an additional communication, will produce an additional effect on energy conservation.

Research Question 3 *What are the mechanisms leading to higher treatment effects?*

We conduct a manipulation check to test if priming environmental identity works through the expected mechanism. We test if the communication received through the eHER affects environmental self-identity and customers' willingness to save. If the treatment message effectively activates environmental identity, we expect a positive coefficient of β_2 in the equation below:

$$EI_i = \alpha + \beta_1 Program_i + \beta_2 Program \times Prime_i + \gamma X_i + \varepsilon_i \quad (4)$$

where EI_i measures either environmental self-identity or customers' willingness to save. Environmental self-identity is measured through the answers to the following 3-item question : "Acting pro-environmentally is an important part of who I am", "I am the type of person who acts pro-environmentally ", "I see myself as a pro-environmentally person". We measure willingness to save by the answer to the question: "Do you intend to save energy? Please report your answer on a scale from 1 to 10" . The vector X_i includes customers' characteristics measured during the May-June 2017 survey (household size, respondent gender and age, education, geographic location, house ownership vs rent, tenure). The analysis is conducted on the sample of respondents to the second wave of survey, run in December 2017- January 2018.

Consistent with the heterogeneous treatment effect found with respect to environmental values, we expect the eHER to activate environmental self-identity with the effect being stronger for the augmented eHER.

Research Question 4 *Which feature of the program, in its digital engagement dimension, is most conducive to higher treatment effects on consumption?*

In order to answer this question, we investigate the intensity of treatment exposure by analyzing the data on customers' online behavior. To this end, we use data on engagement with eHER for the treatment group and consider the following dimensions:

- Open email
- Click through
- Interaction with tips (number of clicks, categorize by type: behaviour (change over time), small investment (dummy for having done at least one), large investment
- Opt-out

These variables proxy the level of engagement with the program and on intensity of exposure. This allows us to calculate the Local Average Treatment Effect (LATE) and see which types of online activity are more conducive to higher impacts on energy usage. Being the delivery of eHER and augmented eHER randomized, it can be used as instrument for customers' online activities. Hence, the first step is estimated as follows:

$$Engage_{it} = \beta_1 Prime_i \times Post_{it} + \beta_2 Post_{it} + h_t + g_i + \varepsilon_{it} \quad (5)$$

where β_1 provides the effect of being sent eHER on the extent of engagement activities. $Post_{it}$ is a dummy becoming one when the eHER belonging to the November-December wave is sent to the particular customer. We use Cragg-Donald F statistic to check for weak instruments (Stock and Yogo (2005)). We then estimate the following structural form using IV:

$$Y_{it} = \gamma_1 Engage_{it} \times Post_{it} + \gamma_2 Post_{it} + h_t + g_i + \varepsilon_{it} \quad (6)$$

where γ_1 represents the LATE on the population of compliers on the energy consumption, Y_{it} . The exercise is carried out on Genius customers which have taken part to the environmental marketing module test.

4 Data

We have two data sources: (i) administrative data from our partner utility, the former recording customer characteristics, monthly consumption, activity on the website; the latter recording when communication was received, and whether customers opened or clicked through the message; (ii) pre and post-treatment online surveys;

The first source of data is administrative records for all customers in the sample. Such data include:

- Engagement with the eHER and tips: open, click, unsubscribe from eHER, interact with tips
- Energy consumption

The second source of data is survey data, that we collected in two waves. Wave 1 was conducted between April and June 2017. The final sample that completed the survey consists in 4,535 customers, of which 3,720 and 815 from the treatment and control group, respectively. This corresponds to a redemption rate of about 3%. Wave 2 is currently in field (December 2017-January 2018) and is conducted on the same subset of customers from the CB sample that we interviewed in wave one. We expect to be able to track about 1200 customers. The aim of this second wave is both to collect panel data from a sample of treatment and control customers; and to serve as manipulation check for our environmental self-identity treatment. The survey features questions on energy saving behaviors and investment, on socioeconomic and demographic characteristics, environmental values, environmental identity, social cohesion and identity, social norms, personality and social preferences.

The following tables provide detailed information on all variables used in the analysis. For each variable, a short description, information on frequency and time span, use in the analysis and source are provided.

Table 1: Outcome variables

| Name, family | Description | Frequency - time span | Source | Hypothesis |
|---|--|--|---------------|-------------------|
| Consumption | Average daily consumption in the month. | Monthly - From July 2015 to April 2018 | Utility | H1, H2, H4 |
| Open the eHER, Engagement | A dummy equal to one if the customer opened the eHer at least once in the month | Monthly - From Nov 2016 to April 2018 | Utility | H1, H4 |
| Click the eHer, Engagement | A dummy equal to one if the customer clicked the eHer at least once in the month | Monthly - From Nov 2016 to April 2018 | Utility | H1, H4 |
| Visit tips, Engagement | A dummy equal to one if the customer visited at least one tip at least once in the month | Monthly - From Nov 2016 to April 2018 | Utility | H1, H4 |
| Opt-out, Engagement | A dummy equal to one after the month in which the customer opted out of the program | Monthly - From Nov 2016 to April 2018 | Utility | H1, H4 |
| Portal registration, Engagement | A dummy equal to one after the portal registration date and equal to 0 before or no registration | Monthly - From Nov 2016 to April 2018 | Utility | H1, H4 |
| Environmental self-identity, Manipulation check | The standardized score provided to the statement "Acting pro-environmentally is an important part of who I am" | Measured on Dec 2017-Jan 2018 | Survey wave 2 | H3 |
| Willingness to reduce energy consumption, Manipulation check | The standardized score provided to the statement "Are you willing to reduce your energy consumption?" | Measured on Dec 2017-Jan 2018 | Survey wave 2 | H3 |

Table 2: Treatments and dimensions of heterogeneity

| Name, family | Description | Frequency-time span | Source | Hypothesis |
|--|--|----------------------------|---------------|-------------------|
| Genius, Treatment | Dummy equal to 1 for customers in the Genius program and 0 otherwise | Time invariant - | Utility | H1, H2, H3 |
| Prime, Treatment | Dummy equal to 1 for customers receiving the eHer augmented with environmental self-identity prime and 0 otherwise | Time invariant | Utility | H1, H2, H3 |
| Pre-treatment consumption, heterogeneity | Average daily consumption over the period July 2015-June 2016 | Time invariant | Utility | H2 |
| Household size, heterogeneity | N. of household members at the time of the first round of the survey | Time invariant | Survey-wave1 | H2 |
| Female respondent, heterogeneity | Dummy equal to one if the survey respondent is female and 0 if male | Time invariant | Survey-wave1 | H2 |
| Respondent age, heterogeneity | Age of the respondent | Time invariant | Survey-wave1 | H2 |
| Household geographical location, heterogeneity | Set of four dummies for geographical areas: North West, North East, Centre, South and Islands. | Time invariant | Survey-wave1 | H2 |
| Household income quartiles, heterogeneity | Set of four dummies for quartiles of the monthly household income | Time invariant | Survey-wave1 | H2 |
| Respondent education level, heterogeneity | Set of five dummies for completed education level: primary, secondary, high school, undergraduate, MA/PhD | Time invariant | Survey-wave1 | H2 |
| House ownership, heterogeneity | Dummy equal to one if the house is owned and zero if rented | Time invariant | Survey- wave1 | H2 |
| House tenure, heterogeneity | Dummy equal to one if the customer has been living in the house for less than 5 years and 0 otherwise | Time invariant | Survey-wave1 | H2 |
| House size, heterogeneity | The dimension of the house expressed in squared metres. | Time invariant | Survey-wave1 | H2 |

Table 2: Treatments and dimensions of heterogeneity, cont'

| Name, family | Description | Frequency-time span | Source | Hypothesis |
|---|---|---------------------|--------------|------------|
| Endowment of electric appliances, heterogeneity | The sum of dummies for the ownership of fridge, freezer, dishwasher, dryer, AC, n. of TV | Time invariant | Survey-wave1 | H2 |
| Environmental values, heterogeneity | Environmental values standardized index. The scores provided to the statements "For this person respecting nature and being in armony with other species is important" and "For this person protecting the environment, preserving nature and conserving resources is important" averaged and standardized (demeaned and divided by the standard deviation) | Time invariant | Survey-wave1 | H2 |

4.1 Sample Balance

For each source of data we aim to conduct balance tests across treatment groups. First, we check that customers randomly allocated to the environmental self-identity treatment are similar to those in the marketing module control group (i.e. customers treated in the Genius program who do not receive the environmental self-identity message) and in the Genius control group (i.e. customers who do not receive the eHER at all) along both the outcome and control variables measured in the pre-treatment period. In case of variables observed over several period of time, yearly pre-treatment averages will be calculated. We denote these variables as y_{i0} and for each of them we estimate the following equation:

$$y_{i0} = \beta_0 + \beta_1 \text{Program}_i + \beta_2 \text{Prime}_i \times \text{Program}_i + \varepsilon_{i0} \quad (7)$$

where β_1 provides the difference in variable y between customers assigned to Genius and the ones assigned to the control group, while β_2 yields the difference between customers receiving the eHER augmented with the self-identity prime and those receiving a control marketing module. We will then report a p -value from a joint test of the following null hypothesis:

$$H_0 : \beta_1 = \beta_2 = 0 \quad (8)$$

References

- Allcott, H. (2011). Social norms and energy conservation. *Journal of Public Economics*, 95(9):1082 – 1095. Special Issue: The Role of Firms in Tax Systems.
- Allcott, H. and Rogers, T. (2014). The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation. *American Economic Review*, 104(10):3003–37.
- Bruhn, M. and McKenzie, D. (2009). In pursuit of balance: Randomization in practice in development field experiments. *American Economic Journal: Applied Economics*, 1(4):200–232.
- Stock, J. and Yogo, M. (2005). *Testing for Weak Instruments in Linear IV Regression*, pages 80–108. Cambridge University Press, New York.
- van der Werff, E., Steg, L., and Keizer, K. (2013). The value of environmental self-identity: The relationship between biospheric values, environmental self-identity and environmental preferences, intentions and behaviour. *Journal of Environmental Psychology*, 34:55 – 63.