Comparative Feedbacks under Incomplete Information

Nicolas Astier*

February 8, 2017

Abstract

Comparative feedbacks, that is messages providing personalized information about how one's behavior compares to that of relevant others, are nowadays widely used - sometimes at a very large scale - in order to change people's behaviors. Yet there are several possible reasons why such feedbacks may affect behaviors, which makes welfare analysis a very challenging task. This paper uses an online experiment to disentangle the role played by incomplete information and the role played by peer pressure. Incomplete information is shown to be a necessary condition for comparative feedbacks to effectively affect participants' choices, suggesting peer pressure was unlikely to play a major role in our specific setting. Nevertheless comparative feedbacks triggered greater changes in behaviors than other informative-only feedbacks. A possible explanation may be found in cognitive costs, as comparative feedbacks more effectively conveyed to participants the idea that it should not be too difficult for them to reach a better outcome.

JEL codes: D12, D83, Q50.

*Toulouse School of Economics. TSE, Université de Toulouse, Manufacture des Tabacs, 21 allée de Brienne, Fr - 31000 Toulouse. Email: nicolas.astier@m4x.org. I am grateful to Yves Le Yaouanq and Nicolas Treich for very insightful discussions that were at the origin of the project, as well as to the numerous friends and colleagues who spent some time playing with and providing feedbacks on different pilot versions of the experiment, and to seminars participants at TSE and CIRED. I also thank Thomas-Olivier Léautier for his guidance and his financial support for running the experiment, and Louis-Gaëtan Giraudet for very helpful comments on an earlier draft of the paper. All remaining errors are mine. Finally, I gratefully acknowledge financial support from the Corps des Ponts, des Eaux et des Forêts.

"Social influences come in two basic categories. The first involves information. If many people do something or think something, their actions and their thoughts convey information about what might be best for you to do or think. The second involves peer pressure. If you care about what other people think about you [...], then you might go along with the crowd to avoid their wrath or curry their favor." (Thaler and Sunstein (2008, p.54))

1 Introduction

Comparative feedbacks, that is messages providing personalized information about how one's behavior compares to that of relevant others, are nowadays used in many different areas. As an example, numerous utilities send such feedbacks to their customers in order to decrease their electricity (Allcott and Mullainathan, 2010) or water (Ferraro et al., 2014) consumption. Some of these programs have nowadays reached a very large scale[1].

In many situations where comparative feedbacks are used, consumers do not have all the information they would need to take optimal decisions. This is typically the case for residential energy or water consumers, who very often lack either detailed and frequent information about their consumption, or the time needed to process such information (Kempton and Layne, 1994). In such settings, there are at least two distinct reasons why people may change their behavior after receiving a comparative feedback:

- **An informative channel**: agents will update their beliefs about *the way their choices map into outcomes* (utility bill, daily comfort, etc.), that is the net utility they would derive privately from choosing a particular action[2].
- **A normative channel**: agents caring about others' perceptions will update their beliefs about *the way their choices map into self or social esteem*, that is how an

[1]For instance, the company Opower tailors comparative feedbacks for more than 50 million consumers around the world.

[2]This channel supposes that agents can infer something from the average behavior in the population. This is for example the case if some agents are endowed with valuable private information.

observer would assess their social "status" from their actions.

Both channels need not be orthogonal (Cialdini and Goldstein, 2004). For example, if esteem is derived from one's ability at a given task previously unknown to the agent, a comparative feedback provides information which is both informative (the additional surplus I can hope to get if I increase effort) and normative (how good I am at the task). Although acknowledged a while ago (Deutsch and Gerard, 1955), this duality remains poorly understood because of the difficulty to disentangle both channels[3].

This paper aims at improving our understanding of the relative importance of both channels. Such an understanding is crucial to the assessment of the welfare impact of comparative feedbacks. Indeed peer pressure sometimes comes at a cost[4] and, in environments where information is highly incomplete, alternative feedback strategies can be contemplated such as messages warning consumers if they exhibit abnormal outcomes (Gleerup et al., 2010). The welfare implications of comparative feedback programs should then be assessed in comparison to other programs aiming at closing the information gap, rather than to a business-as-usual scenario where consumers remain imperfectly informed.

In this paper, we design an online experiment in order to assess, within a controlled environment, how people react to comparative feedbacks in complete and incomplete information environments. In our specific setting, we find that comparative feedbacks had no measurable effect on participants' choices when they were under complete information. Conversely, comparative feedbacks did trigger a significant response under incomplete information.

[3] "*We do not yet know of a clean design to separate observational learning from the conformity effect*" (Cai et al. (2009), footnote 8).

[4] In the model by Bénabou and Tirole (2011) for example, peer pressure is modelled as a zero-sum game which ends up distorting people's choices.

We then investigate, in the incomplete information environment, how people respond to two types of informative-only feedbacks. The first type of feedback informs participants about what would have been their outcome had they made optimal choices. The second type of feedback warns outliers that the outcome they have reached is very far from the optimal outcome.

We find that informative-only feedbacks triggered a smaller change in behaviors than did comparative feedbacks, despite an absence of evidence suggesting peer pressure played a significant role. A possible explanation is that participants who received a comparative feedback exhibit an increased confidence in their ability to reach a better outcome. Indeed, in an incentivized question asked during the experiment, participants reported being more aware of that they could have reached a better outcome after having received a comparative feedback than after having received informative-only feedbacks, despite the fact that the latter conveyed more accurate information. Different kinds of cognitive costs may explain this difference, notably the cost to internalize the information contained in the feedbacks or participants' beliefs about the cognitive cost to incur in order to change their behavior.

The paper is organized as follows. Section 2 provides some background and motivates the experimental set-up. Section 3 describes the experiment whose results are given in section 4 and discussed in section 5. Section 6 concludes.

2 Background

2.1 Literature review

The influence that comparisons with peers have on people's beliefs and choices has been acknowledged in the psychology literature as early as the 1950s and Asch (1952) experiments. For example Festinger (1954) argued that individuals feel a need to regularly assess the relevance of their opinions and the level of their abilities, and that information about others provide a standard of comparison for such self-evaluations. A few decades later, the so-called "social norm theory" (Berkowitz, 2004; Cialdini and Goldstein, 2004) emerged as a way to fight against "undesirable" behaviors, such as college drinking. In a nutshell, the theory states that people usually comply or conform to social norms. However, they often hold biased beliefs about the occurrence of undesirable behaviors, stating that such behaviors are more widespread than they actually are (Prentice and Miller, 1993). Providing accurate information about the actual occurrence of the undesirable behavior should then make people aware of their misperception and change behaviors in the desired direction.

In practice however, a consensus over the effectiveness of comparative feedbacks in changing behaviors was relatively hard to reach from field experiments' results. Take the example of residential energy consumption, where comparative feedbacks were experimented as early as the 1980s. Midden et al. (1983) found that comparative feedbacks were not significantly more effective at inducing energy savings than feedbacks about customers' own historical consumption for electricity, and were ineffective at decreasing natural gas consumption. Academic reviews of early experiments were hence reporting rather pessimistic results: "*comparative feedback [...] has been shown to have positive effects on reducing energy use in the workplace [...] the results are not as clear-cut for households*" (Abrahamse et al., 2005) ; "*none of the twelve studies dealing with normative*

comparisons could demonstrate an effect on consumption so far" (Fischer, 2008). Despite these early negative results,[5] comparative feedbacks reemerged in the late 1990s, as the concept of "social norms marketing" was growing more and more popular. In the 2000s two very influencial studies (Schultz et al., 2007; Nolan et al., 2008) found comparative feedbacks to be effective at encouraging energy conservation. Nowadays, such feedbacks are used at a very large scale, decreasing on average households' energy consumption persistently by about 2% for electricity (Allcott and Mullainathan, 2010; Allcott and Rogers, 2014), and 0.7% for natural gas (Allcott and Kessler, 2015).

Comparative feedbacks have now been tested in a variety of environments, and a closer look at the literature however unveils several limits to their ability to change people's behavior in the desired direction:

- **For individuals with the "desirable" behavior (boomerang effect)**: people who initially do have a "desirable" behavior may revert to a less virtuous behavior (Schultz et al., 2007; Byrne et al., 2016).

- **For individuals with the "undesirable" behavior (excuses, discouragement)**: people receiving many "bad news" may give up trying to improve on their behavior. They may develop strategies allowing them to ignore the feedbacks,[6] or stop the efforts they used to consent performing (Beshears et al., 2011).

- **For all individuals**: for a variety of reasons (bayesian inference, morality, etc), people may dismiss either the relevance (they think the feedbacks provide information irrelevant to their specific situation) or the appropriateness (they do not think the considered activity should be monitored) of comparative feedbacks:

[5] Maybe partly driven by the small sample sizes used in the early field experiments.

[6] For example Siero et al. (1996) noted "*in the case of a continuing bad performance, comparative feedback can also have negative effects [...] people tend to avoid comparisons with others who perform better*".

- **Credibility**: the source sending the feedbacks influences the magnitude of consumers' response (Craig and Mccann, 1978). For example, interviews of British consumers carried by Roberts et al. (2004) suggest they are not very interested in receiving comparative feedbacks because they do not really trust their retailer. Consistently, Allcott (2015) found that comparative feedback campaigns are less effective when they are carried for investor-owned utilities than when they are carried for publicly-owned utilities.
- **Complaints**: in Ferraro and Price (2013), the utility received some angry phone calls from low-consumption customers complaining that they were wasting money on supervising them. In Ayres et al. (2009) some consumers complained about frowning emoticons which were subsequently removed from Opower's Home Energy Reports. In Costa and Kahn (2013), a non negligible percentage of consumers report disliking comparative feedbacks, leading the authors to plainly conclude: "*What works in California may not work in Lubbock, Texas*"[7].

Economists recently made an important contribution which helps to think about these puzzles (Bernheim, 1994; Bénabou and Tirole, 2006, 2011): instead of assuming an exogenous compliance to norms, they argued that people may follow norms for *endogenous* reasons, namely because they care about how others perceive them, that is their social "status". It is thus in their private interest to bias their actions in order to signal a higher (perceived) status. Providing information about how one compares with others may then alter the underlying parameters of this image-seeking game and

[7]Consumers' stated willingness to receive comparative feedbacks is very heterogenous (Allcott and Kessler, 2015), and seems to depend for example on their political opinions (Costa and Kahn, 2013), on their culture (Fischer, 2008), etc... In the case of residential electricity consumption, Fischer (2008) notes: "*For the UK (IEA 2005, p.10) and for Sweden (Sernhed et al. 2003), it is reported that citizens exhibit an interest in comparison with their own previous consumption, but are much less interested in comparisons with other households. On the contrary, the Finnish customers in the study conducted by Haakana et al. (1997) desired normative comparisons, and the Japanese respondents in Ueno et al. (2005, p.1,293) were much more interested in comparisons with others than with own previous consumption*".

influence behaviors.

A second reason why consumers may react to comparative feedbacks is *incomplete information*. Indeed consumers may not know how their behavior influences their outcome very well,[8] and observing others' behavior may allow them to learn about it. Such incomplete information environments can for example arise because of rational inattention. Starting with Banerjee (1992) and Bikhchandani et al. (1992), a vast litterature on social learning has studied such situations, although this literature mainly focussed on identifying conditions under which crowds may (rationally) herd on a suboptimal action (Eyster and Rabin, 2010, 2012; Eyster et al., 2015).

There exists alternative reasons that could potentially explain people's response to comparative feedbacks. One is anchoring: comparative feedbacks may act as a reference point to which consumers converge. Such an explanation is however of little relevance in environments where consumers' payoffs arise from a complex combination of numerous choices, such as residential energy consumption. Another mechansim is that consumers may have relative preferences. Although this is unclear why, in applications such as residential energy consumption, people would derive a significant pride or shame from their privately-observed energy consumption (Heffetz, 2011),[9] the normative channel discussed above does capture some features of relative preferences. Finally comparative feedbacks may credibly act as cues drawing attention to otherwise ignored parameters relevant to the decision (Allcott and Rogers, 2014; Taubinsky, 2013). Such a "saliency effect" was not found in the context of the field experiment by Cai et al. (2009).

[8] For example the warm glow derived from giving to a charity may depend on the ability of the charity to use the money efficiently, as in Niehaus (2014).

[9] Interestingly Opower initially randomized over "batches" of households expecting spillovers within neighborhoods (Ayres et al., 2009), that is neighbors talking together about received feedbacks. No evidence of such spillovers was found (Costa and Kahn, 2013), suggesting that neighbors do not discuss that much about energy conservation. Nevertheless, the fact that energy conservation is very often framed as public goods gives some credit to the existence of relative preferences (Yoeli et al., 2013).

2.2 Motivation behind our experimental setting

The set up of our experiment is inspired by the use of comparative feedbacks in settings such as residential electricity consumption. Although residential customers consume electricity *services* (lighting, heating, etc...), their are most often billed for *monthly* electricity *quantities*: their consumption is thus aggregated both over a long period of time and over many different appliances. Kempton and Layne (1994) made this point salient in a powerful comparison which motivated our experimental design: "*consider groceries in a hypothetical store totally without price markings, billed via a monthly statement like 'US$527 for 2362 food units in April'* "[10]. Consumers' choices are thus made under highly incomplete information, since they do not know the individual consumption[11] - and thus the price - of the services they consume. Such a complex environment may deter any attempt to learn at all,[12] and it is not very surprising that households have been found to hold biased beliefs about the electricity consumption of individual appliances (Wood and Newborough, 2003; Attari et al., 2010).

In the absence of user-friendly information about per-appliance usage costs, most consumers remain imperfectly informed and rely on heuristics and routines, which may then be sustained by *ex post* wishful thinking. For example, customers may convince

[10]Faruqui et al. (2010) drew a similar comparison: "*Imagine a world in which Joe Smith drives up to the gasoline pump in his large SUV, fills up his truck, and drives away without paying a dime. The gasoline is not free, but Smith won't know how much he purchased or how much he owes until a month later because he has a monthly account with the filling station. When his wife drives up to the pump in the family sedan, she goes through the same procedure; as does their high school senior, who drives up to the pump in her compact coupe. The Smiths get a combined bill a month later and don't know how the charges accumulated. Was it Joe's driving, his wife's driving, or their daughter's driving that accounted for the lion's share of the bill? What makes life even more interesting for the Smiths is that none of their cars have a speedometer or a gasoline gauge. They get no feedback at all on how to manage their gasoline bill*".

[11]To make things worse, electricity consumption is measured with the unfamiliar unit "kilowatt hours" (Kempton and Montgomery, 1982; Roberts et al., 2004). Besides, it is still common that bills are based on *actual* consumption only once a year.

[12]For example Brounen et al. (2013) found after surveying 1721 Dutch households that about half of participants answered "*I have no idea*" to the question "*How much do you pay for your monthly electricity bill?*".

themselves that the cost of the energy-saving measures they could take is prohibitive, finding excuses not to disturb their habits[13]. Changing behaviors then requires to overcome these *ex post* rationalization strategies, something general information provision or public appeal messages typically fail to do[14] (Ferraro et al., 2014).

Heuristics' driven choices create a wide dispersion in individual consumption patterns. Iyer et al. (1998); Zimmermann (2009); Kwac et al. (2014) report that the distribution of daily electricity consumption for households with similar characteristics tend to have a thick right tail. Consistently Armel et al. (2013) note that "*anecdotal reports from plug monitoring companies, disaggregation developers, and researchers overseeing feedback studies suggest that the largest savings often come from surprising places, such as an extra Tivo, a pool pump, a pottery wheel, or an electric towel or floor warmer, inadvertently left on*"; and Leighty and Meier (2011) report that sustainable savings were realized following an energy crisis by the simple fact that "*many "forgotten" or "spare" devices like clocks and televisions in guest rooms and spare refrigerators or freezers remained unplugged*". As a consequence, in incomplete information environments, comparative feedbacks may also partly act as an effective warning device, allowing customers to realize there is something wrong with their consumption pattern,[15] in the spirit of the learning by noticing literature (Hanna et al., 2014).

[13] "*All respondents in general felt that energy must be saved. Enthusiasm for conservation somewhat decreased the more people spoke of themselves or their own situation*" (Midden et al., 1983). In Midden et al. (1983) some consumers who received comparative feedbacks but did not change their behavior argued their comparison group was not relevant to their situation, but "*[i]n many cases these doubts seemed to be rationalizations rather than well-justified arguments*".

[14] Interestingly, one situation where general information provision has been found to be effective is during times of emergency and electricity shortages (Costa, 2013; International Energy Agency, 2011; Leighty and Meier, 2011; Lutzenhiser et al., 2004; Meier, 2009; Reiss and White, 2008). Experiencing electricity shortages may arguably grab consumers' attention and/or help to counteract rationalizations and wishful thinking: "*the more people regarded the energy problem as a political one, the less they saved and the more people experienced a shortage the more they saved*" (Midden et al., 1983).

[15] Consistently high baseline consumers typically realize higher savings (Dolan and Metcalfe, 2013; Ferraro and Price, 2013; Byrne et al., 2016). Besides the order of magnitude of savings (which is discussed in Appendix A) achieved by comparative feedbacks is similar to the order of magnitude of savings in Gleerup et al. (2010), where messages warning about abnormal consumption levels were used.

In order to better understand to what extent comparative feedbacks may change behaviors through information provision, we set up a web-based experiment which replicates some of the main features of residential electricity consumption. We now describe the details of our experiment.

3 Experimental design

3.1 Description of the experiment

We designed an online game inspired by Kempton and Layne (1994) metaphor of a grocery shop without price tags that would charge monthly bills to its consumers. Participants were recruited on Amazon's Mechanical Turk platform. They were asked to take care of a virtual pet called "Noney", choosing every day which food items to give to her[16]. The ludic framework of the experiment was a deliberate attempt to harness participants' intrinsic motivation to play seriously[17].

The goal of the game was to maximize Noney's happiness with a fixed weekly budget. In economics terms, the game consisted in maximizing one's utility by allocating a fixed budget between three different goods:

- A numeraire good (which was framed as water and dry cat food for food items' choices, and as a weekly gift to Noney bought with the money leftover from the weekly budget).

- Two goods that may provide inframarginal utility: milk and fish.

[16] An additional non-instrumental task was performed in order to divert players' attention - as do various real-life tasks - and to encourage players to adopt simplified heuristics.

[17] Participants registered on online platforms very often perform numerous boring tasks in a row, not paying much attention to them. Making the task more "fun" than usual was intended to break this monotonicity so as to increase participants' attention. Participants were invited to leave comments after finishing the experiment. Many of them stressed they did appreciate the framing of the experiment.

Two virtual weeks were played, in order to be able to send feedbacks based on the outcome of the first week. All parameters remained identical in both weeks.

On a given day, participants were informed of the utility derived from consuming milk (resp. fish). The more participants previously gave Noney milk (resp. fish) during the current week, the lower the incremental utility was. An optimal behavior was thus to stop giving milk (resp. fish) to Noney when the incremental utility derived became lower than the price of the food item.

Players were randomized between two environments: one environment where the prices of milk and fish were known - the Complete Information (CI) environment -; and one environment where the prices of milk and fish were unknown - the Incomplete Information (II) environment. The randomization was performed explicitly during the tutorial, so that players who ended up in the II environment were aware that some other players were in a CI environment. Figure 1 displays the demand functions for milk and fish, as well as the prices of the food items. It can be seen that the optimal behavior (under CI) was to feed Noney with milk 5 times, and with fish once, which corresponds to a weekly bill[18] of \$7.1.

At the end of the first week, players were informed of their realized weekly bill. Control groups in CI and II environment received no further feedback. However, one treatment in the CI environment and three treatments in the II environment were implemented at the end of the first week[19]:

[18]The numeraire food items' prices were respectively 1 dime (water) and 4 dimes (dry cat food).

[19]The allocation across treatments was pseudo-random in the sense that data was first collected for both control groups and the information only treatment, and then for the other treatments (doing so comparative and warning feedbacks were based on realized data, making the experiment non-deceptive). In addition, in order to get balanced treatment groups, the allocation to treatments was random at first, and then forced towards treatments with the lowest number of observations. Given participants' arrival on the platform is random, it is unlikely that this procedure created any bias.

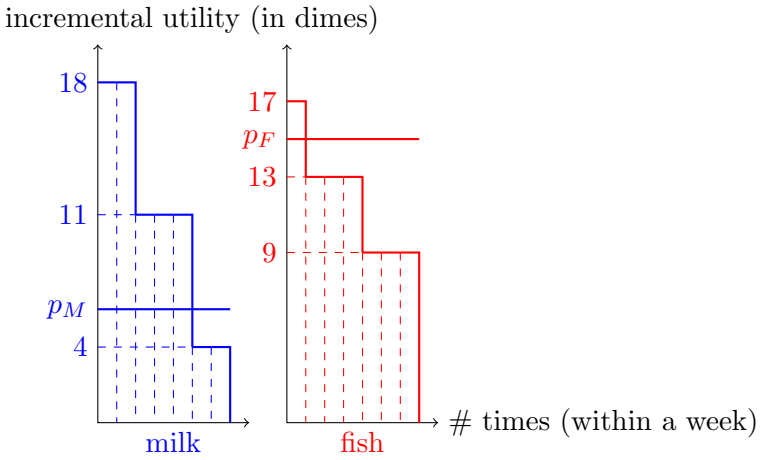


Figure 1: Demand curves for milk and fish

- **Comparative feedback:** players in this group received a feedback telling them how their first-week bill compares to the control groups'.
- **Information only:** players were told the amount of the utility-maximizing bill.
- **Warning to outliers:** players having a first week bill significantly higher than the utility-maximizing one received a feedback warning them about their abnormally high bill (a threshold bill of \$9.9 was chosen, corresponding to the eighth decile of the first week bills observed in control groups).

Appendix B shows screenshots of the feedbacks received in the different treatments[20]. Table 1 displays the number of participants in the different treatments:

	Control	Comparative Feedback	Information Only	Warning
CI	50	48		
II	50	51	50	52

Table 1: Number of participants per treatment

[20]The game with the instructions received by participants will be left available online for some time at `http://www.nastier.eu/experiments/noney/index2.php` (password: Itsamilkmansworld).

4 Results

4.1 Strategies to increase data quality

Several screening devices were implemented to make sure participants paid attention to the instructions. First, both the descriptive summary of the experiment and the first page of instruction made it clear that any payment (including the fixed participation fee) was conditionnal on successfully answering a few comprehension questions. Second the first page of instructions was formal and lengthy. It explained that the variable bonus paid to participants was relatively steep with respect to performance (ranging from \$0 to \$3.5, the latter being more than three times the fixed participation fee), and that an average performance translated into a relatively low bonus. Third, the tutorial was quite long so as to being necessary for answering the comprehension questions. Fourth, a first failure to answer comprehension questions correctly obliged players to go back to the beginning of the tutorial. Fifth, a second failure to answer those questions kicked participants out of the game without payment. A single log per participant was allowed so that kicked-out players could not log in again. Finally, the experiment was framed as a ludic online game in an attempt to harness intrinsic motivation. These various strategies did prevent a significant number of prospective participants from enrolling into the experiment, as shown in Table 2.

Giving-up step	Number
First page of instructions	12
After a few pages in the tutorial	15
After a first-failure in comprehension questions	21
After a second-failure in comprehension questions	25
TOTAL	73

Table 2: Number of participants screened out

Hence about 20% of players (73 out of 374) who logged in did not complete the tutorial. Given that 85% of players (258 out of 301) who took part in the experiment provided correct answers to the tutorial questions at their first attempt, it is reasonable to think that the main reason for failing the tutorial questions was a lack of interest. As such it is very likely that the screening strategies implemented increased data quality.

4.2 Comparative Feedbacks under CI

We now look at players' response to comparative feedbacks under complete information.

4.2.1 Validity of randomization

Table 3 provides summary statistics (means and standard deviations) for the first week of the game, that is when instructions were so far identifical for both the control and the treatment groups. The randomization process seems to have reasonably worked based on these statistics.

Treatment	Control group CI0	Treatment group CI1
First-week bill (in $)	7.27 (1.34)	7.70 (1.60)
Time spent on tutorial (in min)	7.08 (2.84)	8.15 (4.00)
Time spent on week 1 (in min)	4.31 (1.49)	4.86 (1.36)
Tutorial passed at 1st attempt (%)	82	93.75
Sample Size	50	48

Table 3: Summary statistics for week 1 (CI environment)

4.2.2 Participants' response to the comparative feedback

Graphical results

Figure 2 shows the histograms of weekly bills for the control group CI0 (left panels), and for the treatment group CI1 who received a comparative feedback at the end of

week 1 (right panels). Week 2 bills are in red, while week 1 bills are in blue. Appendix C.1. provides further graphical representations of the results in terms of distributions in the space of commodity bundles. Graphically it seems that participants did not change their behavior much after receiving the comparative feedback.

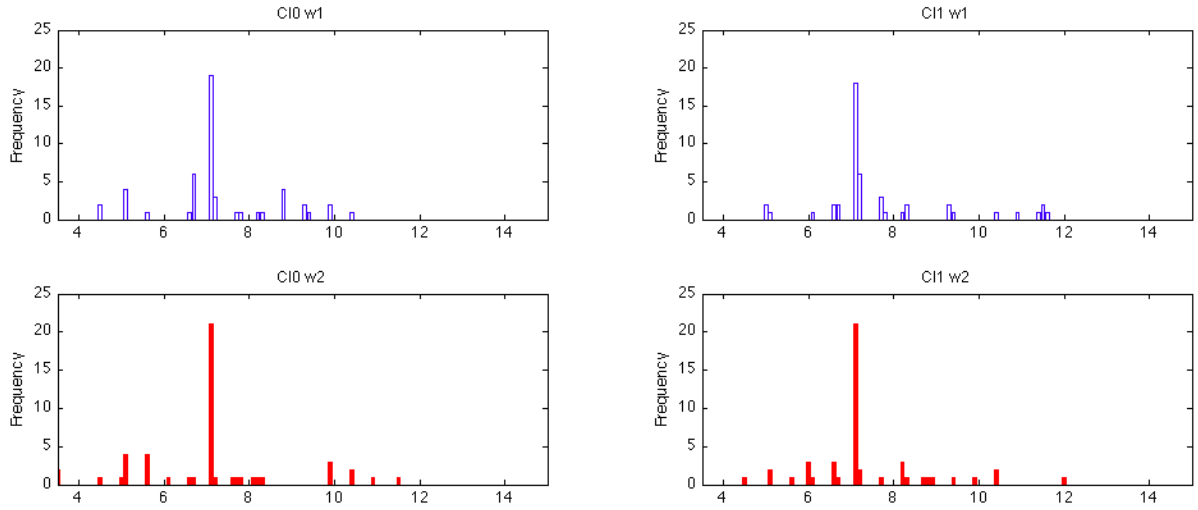


Figure 2: Bills' distributions under complete information for week 1 (top) and week 2 (bottom), with (right) and without (left) a comparative feedback at the end of week 1.

Statistical analysis

We run several specifications of the following regression:

$$Y_i = \alpha + \tau R_i + \beta X_i + \epsilon_i$$

where Y_i is participant i's second-week bill, R_i is an indicator variable equal to 1 if the participant received a comparative feedback and 0 otherwise, and X_i is a vector of control variables. The coefficient of interest is τ which represents the magnitude of the treatment effect (decrease in week 2 bill in \$). Specification (1) includes no control variable (simple comparison of means). Specification (2) includes the first week bill as

a control. Specification (3) adds the total time spent on week 1 as a control. Finally, specification (4) adds the total time spent on the tutorial as a control. The following table summarizes the results (robust standard errors are given in parenthesis).

Specification	(1)	(2)	(3)	(4)
Constant	7.11 (0.25)***	2.99 (0.87)***	2.83 (0.83)***	2.81 (0.81)***
Treatment	0.25 (0.32)	0.01 (0.28)	−0.01 (0.29)	0.00 (0.29)
First week bill		0.57 (0.12)***	0.56 (0.12)***	0.58 (0.12)***
Time first week			0.04 (0.11)	0.10 (0.13)
Time tutorial				−0.05 (0.04)
R^2	0.006	0.285	0.286	0.294

Table 4: Comparative feedback under CI ($^{***}: p < 0.01$, $^{**}: p < 0.05$, $^{*}: p < 0.1$)

The statistical analysis confirms the naked-eye intuition: comparative feedbacks had no impact on participants' choices in a complete-information environment.

A possible explanation may be that the optimal bill ($7.1) happened to lie between the second decile bill ($6.7) and the average bill ($8.1) of the comparison group, so that the feedback conveyed little peer pressure to most players under CI. However, the dispersion in second week bills of participants who played suboptimally does not seem smaller in the treated group than in the control group[21]. Thus a more likely explanation may be that since the bill does not relate to any intrinsic characteristic of the players (there is no horizontal differentiation in the game), participants had no reason to respond to the comparative feedback under CI.

4.3 II environment

We now turn to the treatments under incomplete information. Besides the control group (II0), three treatments were implemented: comparative feedbacks (II1), information-only

[21] See Appendix C.1.2. for a representation of week 2 choices in the space of commodity bundles.

(II2) and warning-to-outlier feedbacks (II3).

4.3.1 Validity of randomization

Treatment	II0	II1	II2	II3
First-week bill (in $)	8.94 (2.31)	7.90 (2.52)	8.09 (1.90)	8.47 (2.42)
Time spent on tutorial (in min)	8.06 (4.09)	7.29 (3.86)	7.27 (3.45)	7.46 (4.89)
Time spent on week 1 (in min)	4.90 (3.77)	4.26 (1.95)	4.51 (1.55)	3.93 (1.86)
Tutorial passed at 1st attempt (%)	90	84.3	88	76.9
Sample Size	50	51	50	52

Table 5: Summary statistics for week 1 (II treatments)

Table 5 provides summary statistics for the first week of the game under II. Not surprisingly, the distribution of first week bills is much more dispersed under II than it was under CI. Because of this higher dispersion in first-week outcomes, our preferred specification will be (2) which includes a control for the first-week bill.

4.3.2 Participants' response to the comparative feedback

Graphical results

Figure 3 shows the histograms[22] of weekly bills for the control group II0 (left panels), and for the treatment group II1 who received a comparative feedback in an incomplete information environment (right panels). Week 2 bills are in red, while week 1 bills are in blue. Graphical intuition now suggests that comparative feedbacks triggered a decrease in week 2 bills.

[22] Again, alternative representations of the distribution of outcomes in the II environment are provided in Appendix C.2.

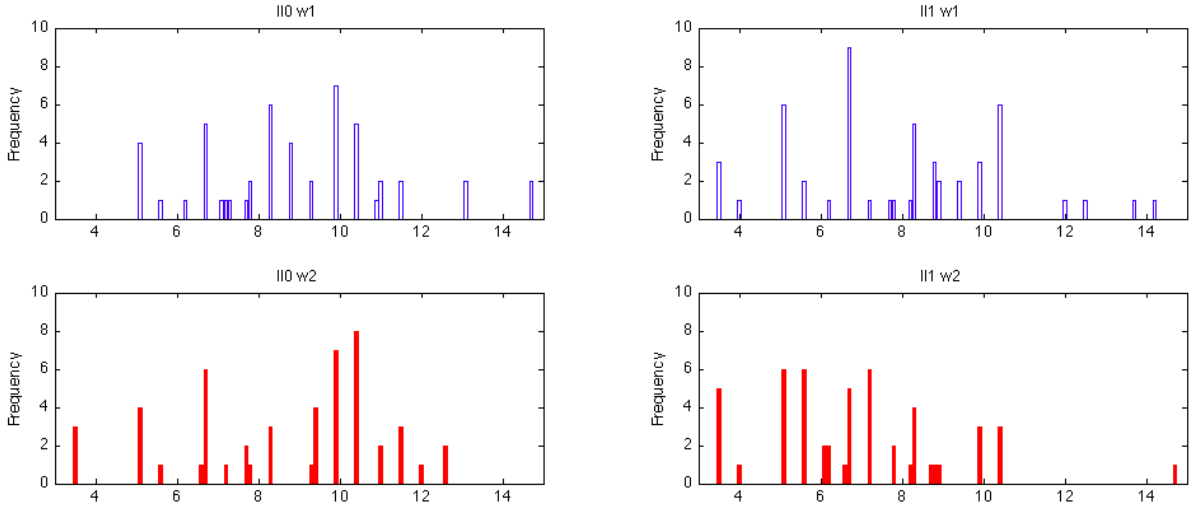


Figure 3: Bills' distributions under incomplete information for week 1 (top) and 2 (bottom), with (right) and without (left) a comparative feedback at the end of week 1.

Statistical analysis

Table 6 presents the results of regressions (same specifications as in Table 4).

Specification	(1)	(2)	(3)	(4)
Constant	8.63 (0.34)***	5.56 (0.99)***	6.09 (1.15)***	6.29 (1.29)***
Treatment	−1.72 (0.46)***	−1.36 (0.48)***	−1.43 (0.49)***	−1.45 (0.50)***
First week bill		0.34 (0.10)***	0.33 (0.10)***	0.33 (0.10)***
Time first week			−0.09 (0.06)	−0.08 (0.06)
Time tutorial				−0.02 (0.05)
R^2	0.123	0.236	0.247	0.249

Table 6: Comparative feedback under II (*** : $p < 0.01$, ** : $p < 0.05$, * : $p < 0.1$)

Contrary to what happened under CI, comparative feedbacks did trigger a significant decrease in second-week bills. The comparison of the outcomes for the CI and the II environments thus suggests that participants reacted to the comparative feedback mainly

because of its informative content, rather than because of its normative framing. We now further test this hypothesis by studying the information-only and warning treatments.

4.3.3 Participants' response to the information-only feedback

Graphical results

Figure 4 shows the weekly bills obtained in the control group II0 (left panels), and the treatment group II2 who received a message conveying the value of the utility-maximizing bill (right panels).

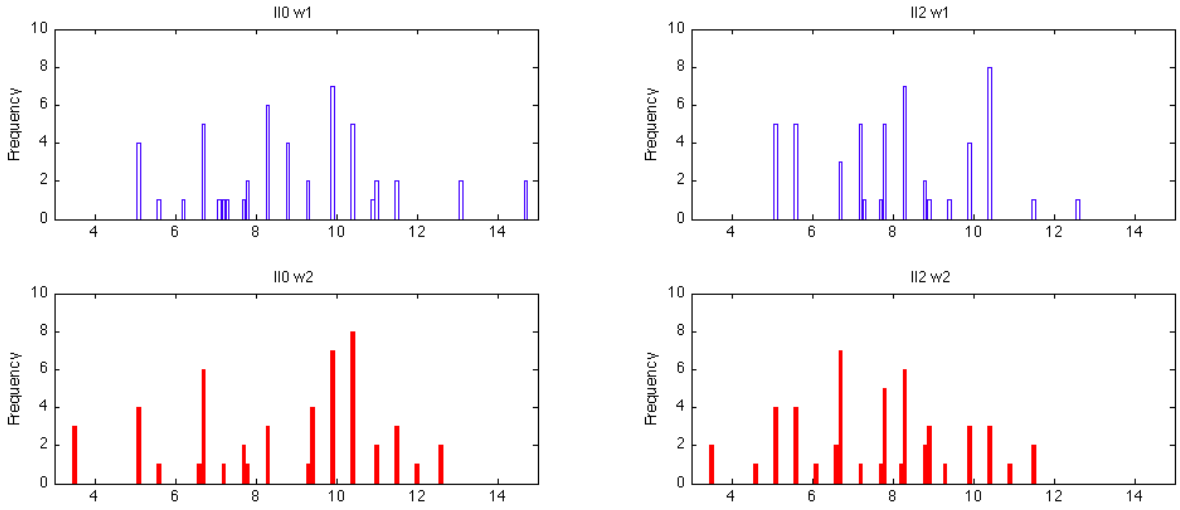


Figure 4: Bills' distributions under incomplete info for week 1 (top) and 2 (bottom) with (right) and without (left) an information-only feedback at the end of week 1.

Statistical analysis

Table 7 presents the results of the regressions.

Specification	(1)	(2)	(3)	(4)
Constant	8.63 (0.34)***	4.61 (0.83)***	5.08 (0.97)***	4.47 (1.14)***
Treatment	−1.02 (0.44)**	−0.63 (0.41)	−0.67 (0.42)	−0.61 (0.43)
First week bill		0.45 (0.08)***	0.44 (0.09)***	0.45 (0.09)***
Time first week			−0.07 (0.05)	−0.09 (0.06)
Time tutorial				0.07 (0.05)
R^2	0.051	0.229	0.237	0.249

Table 7: Informative-only feedback (*** : $p < 0.01$, ** : $p < 0.05$, * : $p < 0.1$)

Interestingly, although the treatment effect has the expected sign, both its magnitude and its significance ($p-$values range from 0.128 to 0.161 for specifications (2) − (4)) are weak compared to what was observed for comparative feedbacks. The information-only feedbacks were thus less effective in inducing participants to decrease their bills.

4.3.4 Participants' response to the warning feedback

Graphical results

Figure 5 shows the weekly bills obtained in the control group II0 (left panels), and the treatment group II3 (right panels). Participants in the II3 group whose first-week bill was higher than \$9.9 received a message warning them that their first week consumption was abnormally high (in practice, 17 out of the 52 participants in this treatment received the warning).

Statistical analysis

Table 8 presents the results of the regressions.

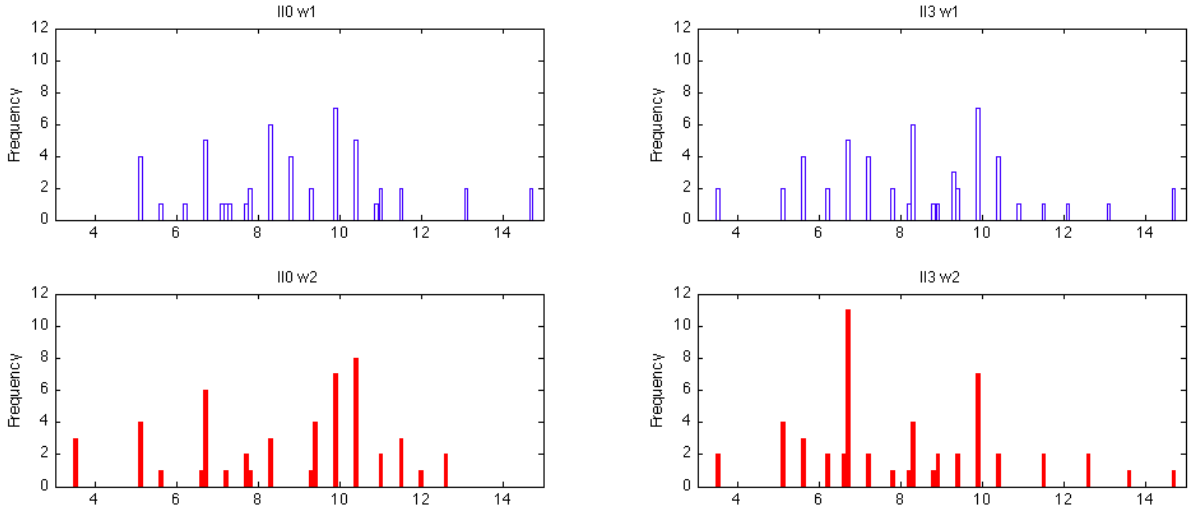


Figure 5: Bills' distributions under incomplete info for week 1 (top) and 2 (bottom) with (right) and without (left) receiving a warning feedback at the end of week 1.

Specification	(1)	(2)	(3)	(4)
Constant	8.63 (0.34)***	6.32 (1.00)***	6.79 (1.11)***	7.27 (1.13)***
Treatment	−0.57 (0.49)	−0.45 (0.48)	−0.54 (0.49)	−0.55 (0.49)
First week bill		0.26 (0.10)**	0.25 (0.10)**	0.27 (0.10)**
Time first week			−0.08 (0.05)*	−0.05 (0.04)
Time tutorial				−0.10 (0.04)**
R^2	0.014	0.076	0.087	0.116

Table 8: Warning feedback (*** : $p < 0.01$, ** : $p < 0.05$, * : $p < 0.1$)

Again, although the treatment effect has the expected sign, both its magnitude and its significance are low compared to what was observed for comparative feedbacks. However, restricting attention to participants with a first-week bill higher than \$9.9 (17 players in II3 and 21 players in II0), that is focussing on participants who received or would have received the warning, the treatment effect on the treated becomes significant:

Specification	(1)	(2)	(3)	(4)
Constant	9.44 (0.46)***	7.90 (2.78)***	7.52 (3.00)**	7.76 (3.25)**
Treatment	−2.30 (0.66)***	−2.30 (0.67)***	−2.31 (0.67)***	−2.26 (0.68)***
First week bill		0.14 (0.24)	0.14 (0.24)	0.13 (0.26)
Time first week			0.08 (0.17)	0.08 (0.17)
Time tutorial				−0.02 (0.04)
R^2	0.256	0.265	0.268	0.269

Table 9: Warning feedback (treatment on the treated)

To summarize, it appears that - in our specific experimental setting - incomplete information was a necessary ingredient for comparative feedbacks to have any influence on participants' choices. The feedback informing participants of the utility-maximizing bill had a smaller and non-significant impact, although in the expected direction. Finally, the warning feedback had a high and significant effect on the treated, but the average treatment effect for the whole group was not significant.

5 Discussion

We now use the more detailed information collected in order to further investigate participants' response to the different feedbacks.

5.1 Heuristics and cognitive costs

5.1.1 Choice patterns

Knowing that the incremental utility from consuming milk and fish is decreasing, an optimal choice pattern (at least weakly) consist in choosing milk and fish first, and then switching to the numeraire good after a few days[23]. Table 10 reports the percentage of

[23]The incremental utility after having consumed a given amount of a good was indeed unknown until this amount was reached, so that there was an option value to consume milk and fish first.

players who exhibit such a "single-switching-point" pattern depending on the environment and the week played.

	CI	II
Week 1	27.6%	14.8%
Week 2	45.9%	22.7%

Table 10: percentage of participants who exhibit a "single-switching-point" pattern

Hence most participants opted for suboptimal heuristics (mostly diversifying choices), especially so under incomplete information.

5.1.2 Detecting suboptimal behavior from first-week bill

At the end of week 1, players were asked the following question regarding their first week choices: "*What do you think are the chances (in %) that you could have increased the number of smiley units you have collected so far by making different food choices?*"[24]

Beliefs were elicited in an incentive-compatible manner. More precisely, we slightly modified the classic BDM procedure (Becker et al., 1964) in order to reward accurate beliefs symmetrically. The BDM procedure indeed rewards accuracy asymmetrically: an accurate guess of 100% will always be rewarded, while an accurate guess of 0% only has a fifty percent chance to be rewarded. This feature may create a feeling of unfairness since an accurate guess may not guarantee a sure payment. We instead used a symmetric elicitation procedure[25].

[24]Smiley units stand for utility in the wording of the game.

[25]This procedure comes from a joint work with Yves Le Yaouanq. It works as follows. Assume participants are asked to state their subjective probability (in %) that the state of the world is 1 (vs. say 0). Once a participant submitted her guess, a number p between 0 and 100 is randomly picked. If the participant's guess is above both p and 50, she receives the reward if the state is 1. If her guess is below p and above 50, we run a lottery were the chances to earn the reward are p percent. Symmetrically, if her guess is below both p and 50, she receives the reward if the state is 0. Finally, if her guess if above p but below 50, a lottery with a p percent chance of winning is run.

II0 control group:

The first-week bill conveys some information. If a participant chose milk (resp. fish) $n_M \in [0, 7]$ times (resp. n_F times) during the week and received a bill B, then she should be able to infer that consistent subjective beliefs $\hat{p}_M$ (resp. $\hat{p}_F$) regarding the price of milk (resp. fish) must satisfy:

$$n_M\hat{p}_M + n_F\hat{p}_F + (7 - n_M) + 4(7 - n_F) = B$$

where the price of the alternative numeraire good to milk (resp. fish) is 1 (resp. 4), as in the experiment. The fact that choices are discrete (the quantity of food items are integers) creates an additional opportunity to learn. More precisely, it can be shown that only choices $(n_M, n_F) \in \{(0,0), (0,1), (1,4), (2,4), (3,4), (4,4), (5,0), (5,1), (5,2), (5,3), (5,4)\}$ can be rationalized by subjective prices' beliefs given the received first week bill. We thus test whether the participants in the II0 control group who made first-week choices that cannot be rationalized were more likely to state they first-week choices were suboptimal than other participants. We find no evidence that this is the case:

	Rationalizable	Non-Rationalizable
Stated proba of suboptimal 1st-week choices (%)	71.8 (12.9)	61.2 (23.4)
Sample Size	15	35

Table 11: Beliefs that first-week choices were suboptimal (group II0)

CI groups

Answers by participants under CI to the question whether they believe their first-week choices were suboptimal or not provides some further evidence that players dedicated heterogenous levels of attention to the experiment. Indeed the only reason for participants under CI not to play optimally is that they may not want to incur the cognitive cost to learn how to play optimally. Such participants should thus be aware that they

chose to rely on heuristics, and should state that they could have made better choices. To test this hypothesis, we run the regression $Y_i = \alpha + \beta Opt_i + \epsilon_i$ where Y_i is the participant i belief that her first week choices were suboptimal, and Opt_i is an indicator variable taking the value 1 if participant i first week choices were utility-maximizing.

Specification	(1)	(2)	(3)	(4)
Constant	69.10 (2.70)***	42.91 (10.13)***	27.60 (13.89)**	27.88 (13.87)**
Optimal	−38.72 (5.46)***	−36.62 (5.53)***	−33.84 (5.68)***	−33.85 (5.67)***
First week bill		3.39 (1.22)***	3.27 (1.25)**	3.04 (1.32)**
Time first week			3.32 (1.95)*	2.54 (2.23)
Time tutorial				0.65 (0.90)
R^2	0.380	0.406	0.429	0.433

Table 12: Awareness of suboptimal choices under CI

(*** : $p < 0.01$, ** : $p < 0.05$, * : $p < 0.1$)

As expected, the regression strongly suggests that participants who played suboptimally under CI were aware of it, highlighting the role played by cognitive costs.

II treatment groups

The different feedbacks gave participants under II clear evidence on whether their first-week choices were optimal or not. It is thus natural to wonder to what extent they internalized the information provided. To answer this question, we regress participants' stated belief about their ability to make better choices on a dummy variable for the different treatments under II using specification (2).

Treatment	CF	Info-Only	Warning (all)	Warning (treated)
Constant	43.71 (10.15)***	40.02 (11.01)***	41.22 (9.84)***	95.00 (22.52)***
Treat. Dummy	13.40 (4.08)***	8.11 (4.52)*	4.18 (4.50)	11.94 (6.80)*
First week bill	2.31 (1.05)**	2.72 (1.15)**	2.59 (1.02)**	−2.43 (2.05)
R^2	0.132	0.081	0.073	0.121

Table 13: Treatments effectiveness in raising awareness

(*** : $p < 0.01$, ** : $p < 0.05$, * : $p < 0.1$)

Hence participants' elicited beliefs are consistent with the hypothesis that comparative feedbacks tend to be more effective than purely informative feedbacks in making people aware of their suboptimal behaviors, and confident in their ability to reach a better outcome. Different kinds of cognitive costs may explain this result. First, the information contained in comparative feedbacks may be easier to process, given its graphical presentation. However Table 14 shows that participants seem to have spent enough time on the feedback page to read its content carefully (see Appendix B for screenshots of the feedback page).

Treatment	Mean (Std deviation)
Control group CI0	9.7 (7.2)
CF treatment CI1	35.5 (23.1)
Control group II0	12.2 (8.4)
CF treatment II1	34.8 (16.1)
Information-only treatment II2	33.8 (24.9)
Warning treatment II3 (all)	19.6 (17.2)
Warning treatment II3 (treated)	25.1 (17.2)

Table 14: Time spent on the feedback page (in seconds)

Second, comparative feedbacks may allow to learn about cognitive transaction costs. Indeed the fact that many other players have different bills does not only suggest that one's current behavior is not optimal, it also provides convincing evidence that adopting a behavior closer to the norm is not too difficult ("if others do it, I should be able to do it too"). As a consequence, comparative feedbacks' effectiveness in our experiment may be partly explained by factors beyond the mere information they provide about realized consumption[26].

5.1.3 End of game price guesses

At the end of the game, participants under II were asked their best guesses for the prices of milk and fish. These guesses were strongly incentivized: close enough bets would typically increase participants' payoff by \$2-\$2.5, an amount roughly equivalent to the total average payment received by players who provided wrong price guesses for their participation in the *whole* experiment. In other words, participants were offered a chance to earn in a few seconds an amount of money they had just spent (on average) 16 minutes and 43 seconds to earn.

Besides, since participants had received two weekly bills by the time they were asked to guess prices (plus eventually some informative feedbacks), they were in a position where they had, in theory, all the needed information to retrieve the actual prices[27]. It turned out that *none of the* 203 *participants under II provided accurate guesses for both prices*. This result illustrates that (infrequent) bills aggregating consumption both across time and services make it very difficult for consumers to learn over time.

[26] A similar phenomenon is found in Houde (2014) regarding energy labels. Using micro-data of the US refrigerator market, the author finds that some consumers have a large willingess-to-pay for the Energy Star label, well beyond the energy savings they provide, meaning the treatment effect of the label is not solely driven by their informative content regarding the product.

[27] Unless they made identical aggregate choices in both weeks, which was the case only for 30 participants (14.8%) under II, out of which 6 received the information-only feedback, and 9 received the comparative feedback.

5.2 Welfare

In what preceeds, we measured the effectiveness of the different feedbacks by their impact on participants' second-week bills. However, bills are a rough proxy for welfare. One may thus wonder if receiving a comparative feedback actually made participants better off. Because *all* players had the *same* preferences in our experiment, measuring "welfare" is an easy task. Indeed, participants' welfare just consists in their score at the game. To study the impact feedbacks had on participants' "welfare", we can thus regress consumers' score in week 2 on a treatment dummy. Results are reported in the following table (using specification (2)).

Treatment	CF	Info-Only	Warning (all)	Warning (treated)
Constant	1.58 (0.45)***	1.62 (0.32)***	1.92 (0.39)***	2.40 (0.84)***
Treat. Dummy	−0.21 (0.20)	0.01 (0.18)	−0.14 (0.16)	−0.45 (0.24)*
First week bill	0.11 (0.05)**	0.11 (0.03)***	0.08 (0.04)*	0.04 (0.07)
R^2	0.108	0.072	0.056	0.096

Table 15: Impact of treatments on second-week consumer surplus

(*** : $p < 0.01$, ** : $p < 0.05$, * : $p < 0.1$)

Interestingly none of the feedbacks provided had a significant impact on "welfare" and, if anything, any impact is more likely to have been *negative*: participants tend to have overreacted to the feedbacks provided, leaving some inframarginal utility on the table. Of course, in the absence of external validity, we are in no position to infer something about the welfare impact of such feedbacks in real-life applications. However, our "welfare" analysis illustrates why feedback programs should be evaluated - to the extent that it is feasible - in terms of their impact on social welfare, as attempted in Allcott and Kessler (2015), and not solely in terms of their impact on easily measurable metrics.

Besides, some cognitive costs, such as peer pressure, may be of significant magnitude in some applications. Indeed, while some survey-based studies suggest that a service warning outlier customers of their abnormal consumption would be welcomed,[28] we saw in section 2.1 that experiments involving comparative feedbacks sometimes raise critics and complaints. Information-only feedbacks are thus a better counterfactual to evaluate comparative feedbacks campaigns that a business-as-usual scenario.

6 Conclusion

This paper uses an online experiment to disentangle the role played by incomplete information and the role played by peer pressure in people's response to comparative feedbacks. Incomplete information is shown to be a necessary condition for comparative feedbacks to influence participants choices, suggesting peer pressure was unlikely to play a major role in our specific setting. Participants' response to comparative feedbacks within our experiment is thus most likely explained by learning.

The experiment replicated some stylized features of the choice environmnent faced by residential energy consumers. In such environments, consumers receive (infrequent) bills aggregating consumption both across time and services. The results of our experiment illustrate that such environments make it very difficult for consumers to learn over time: most participants chose to rely on heuristics, and none was able to accurately retrieve the unknown prices despite high incentives and enough information to do so.

Given such a widespread reliance on heuristics, an effective feedback must both catch consumers' attention and convince them it is worth their while changing their behaviors.

[28] For example Sernhed et al. (2003) report from surveys in Sweden that nearly 90% of consumers state wanting to be warned if their energy consumption suddenly increases; and Kempton and Layne (1994) report that the '*most frequently reported purpose of comparing bill information was to check on unusual bills*'.

Interestingly, comparative feedbacks seem to work better for that purpose than purely informative feedbacks.

A possible explanation for the higher effectiveness of comparative feedbakcs is that they also convey credible information about how difficult it may be to improve on the current outcome ("proof by example"). Comparative feedbacks may thus be used as a temporary tool to catch the attention of households with abnormally high consumption levels, rather than as a regular feedback strategy,[29] especially given the potential welfare losses arising from peer-pressure distorted choices.

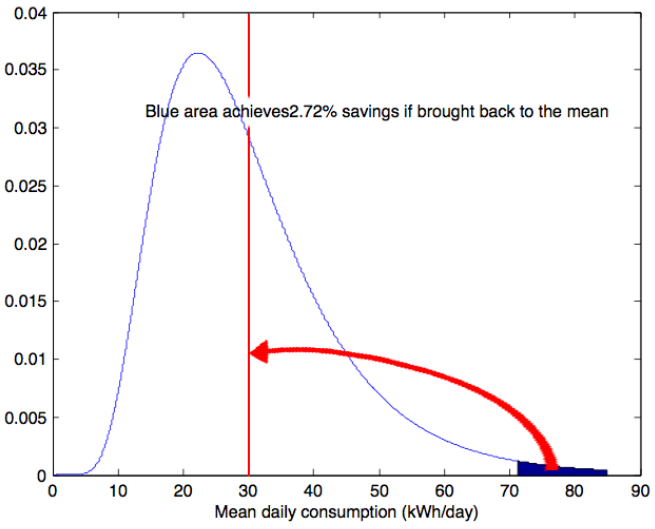
[29]Interestingly, Allcott and Rogers (2014) reached a similar conclusion from a cost-benefit analysis perspective.

Appendices

A Order of magnitude of potential average savings

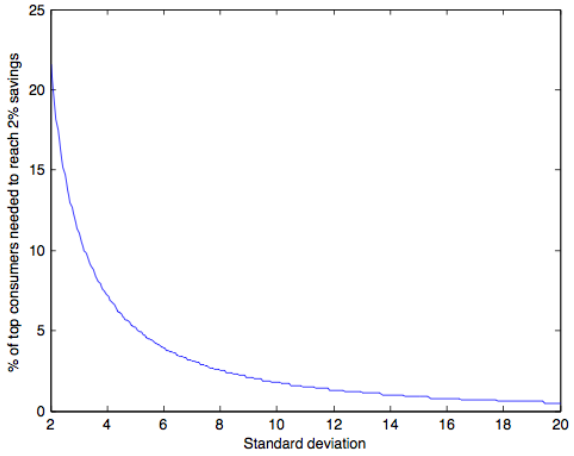
In order to assess the order of magnitude of potential savings from correcting outliers' consumptions, we calibrated a log-normal distribution using the descriptive statistics reported in Opower's field experiments. If consumers' daily consumptions distribution follow a lognormal distribution of mean 30 kWh/day and standard deviation 14kWh/day, bringing the consumption of the 1.5% biggest consumers back to the group mean is enough to achieve *total average* savings of 2.7%, an order of magnitude consistent with those reported in comparative feedbacks experiments.

Study	Mean consumption (kWh/day)	Std Dev	Savings
Ayres et al. (2009)	31.9	-	1.2-2.4%
Costa and Kahn (2013)	30.8	14.7	1.7-2.4%
Allcott and Rogers (2014)	30.3	13.5	2-3%



Of course, the whole group standard deviation is likely to be much higher than the one of comparison groups of similar households, from which actual comparative feedbacks

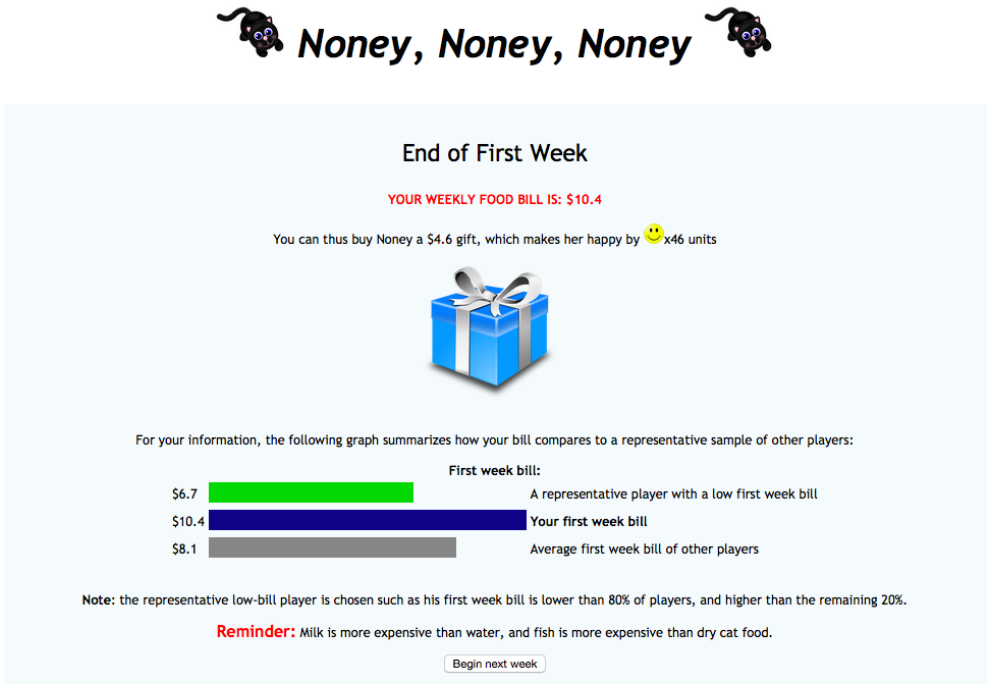
are derived. The following curve shows what fraction of the top consumption customers needs to decrease their consumption to 30 kWh/day in order to reach 2% *total* average savings, when the initial distribution of daily consumptions follows a lognormal of mean 30 kWh/day and a standard deviation given on the x-axis.



B Received feedbacks

The following screenshots display the feedbacks received by the different treatment groups. Note that for the information-only and warning feedbacks, the warning sign displayed was flashing, making it very difficult for participants not to notice the feedback.

B.1 Comparative feedback



B.2 Information about the utility-maximizing bill

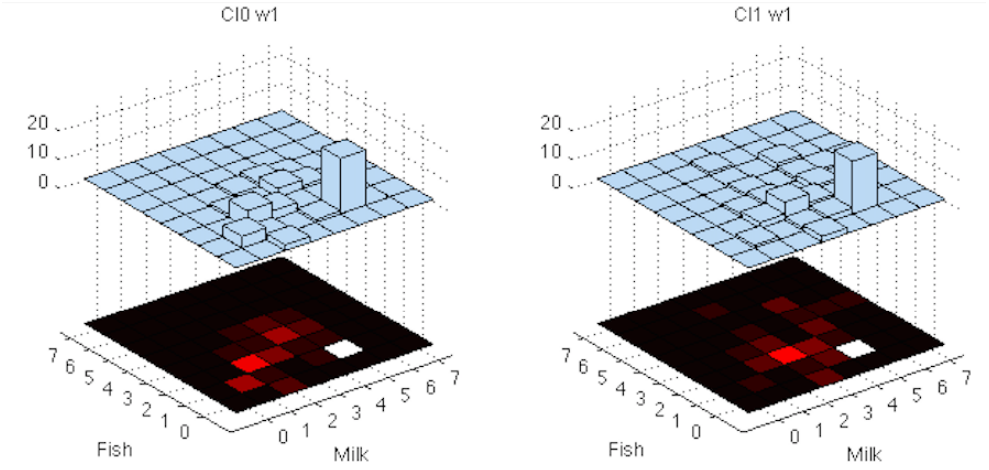
B.3 Warning in case of abnormally high bill

C Further graphs of obtained outcomes

In this appendix, the obtained distributions are drawn in the space (n_M, n_F) (number of times milk and fish were chosen during the week). Migration patterns from week 1 to week 2 are also depicted.

C.1 CI environment

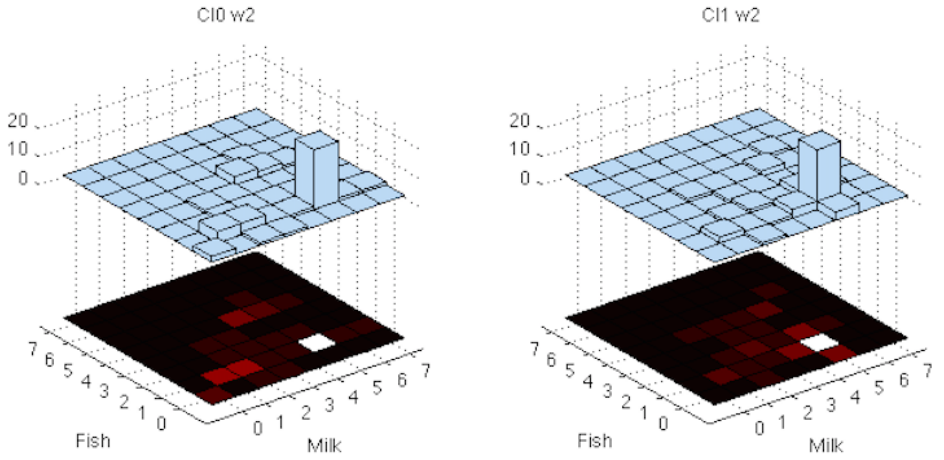
C.1.1 Week 1



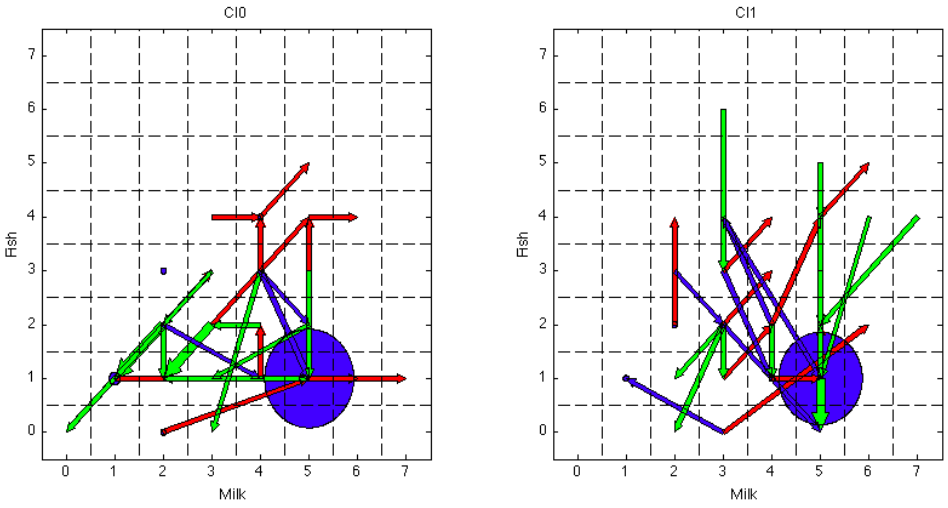
Since the actual choices made by participants are how much milk and fish to consume (rather than which bill to receive), a representation of the distribution of food item choices helps to better understand their actual behavior. On the above graph, the x and y axis indicate how many times milk and fish were chosen during week 1, while the z axis indicates how many participants exhibited a given pattern. Both a histogram and a heat map are given. The left panel represents the control group, while the right panel depicts the comparative feedback group.

C.1.2 Week 2

The following graph shows the same information for week 2 outcomes (still in the CI environment).



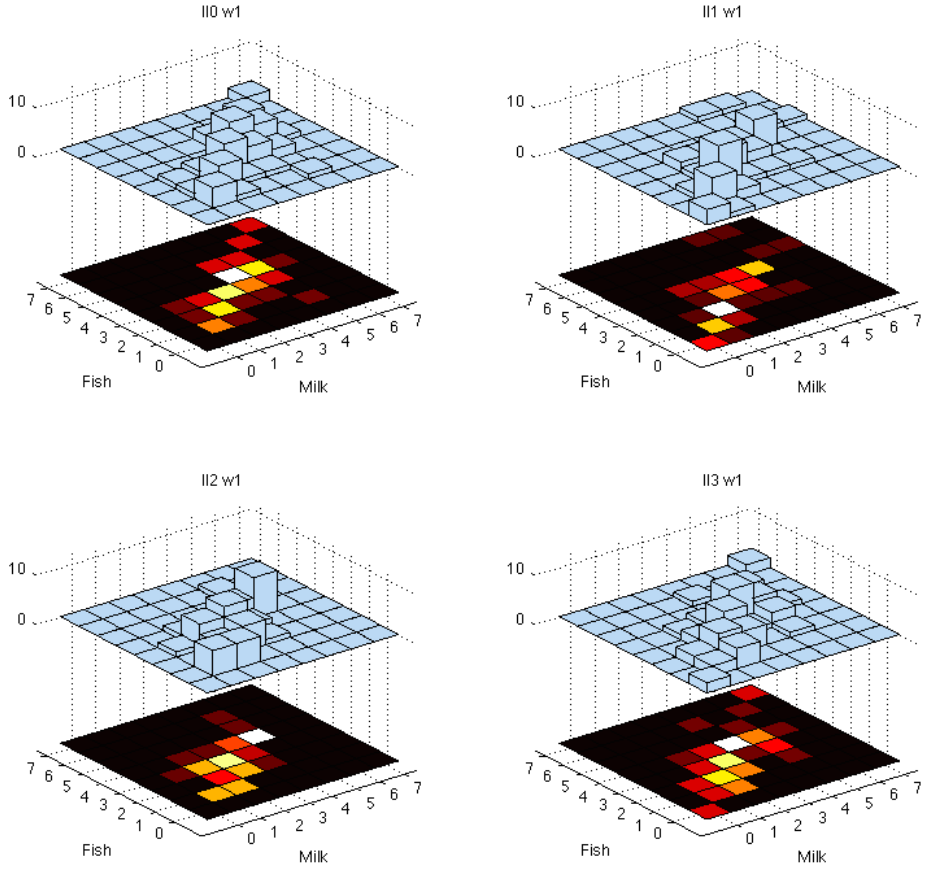
C.1.3 Migration



The above graph shows the changes in consumers' choices between week 1 and week 2. A circle indicates that choices remained the same, while an arrow link week 1 choice to week 2 choice. The bigger the geometric form, the more participants exhibited this pattern. A green (resp. red) arrow means that the consumption of *both* goods weakly decreased (resp. increased). Blue arrows depict ambiguous patterns. The figure shows that most participants made the utility-maximization choice $(n_M, n_F) = (5, 1)$ and stuck to it, whether or not they received a feedback.

C.2 II environment

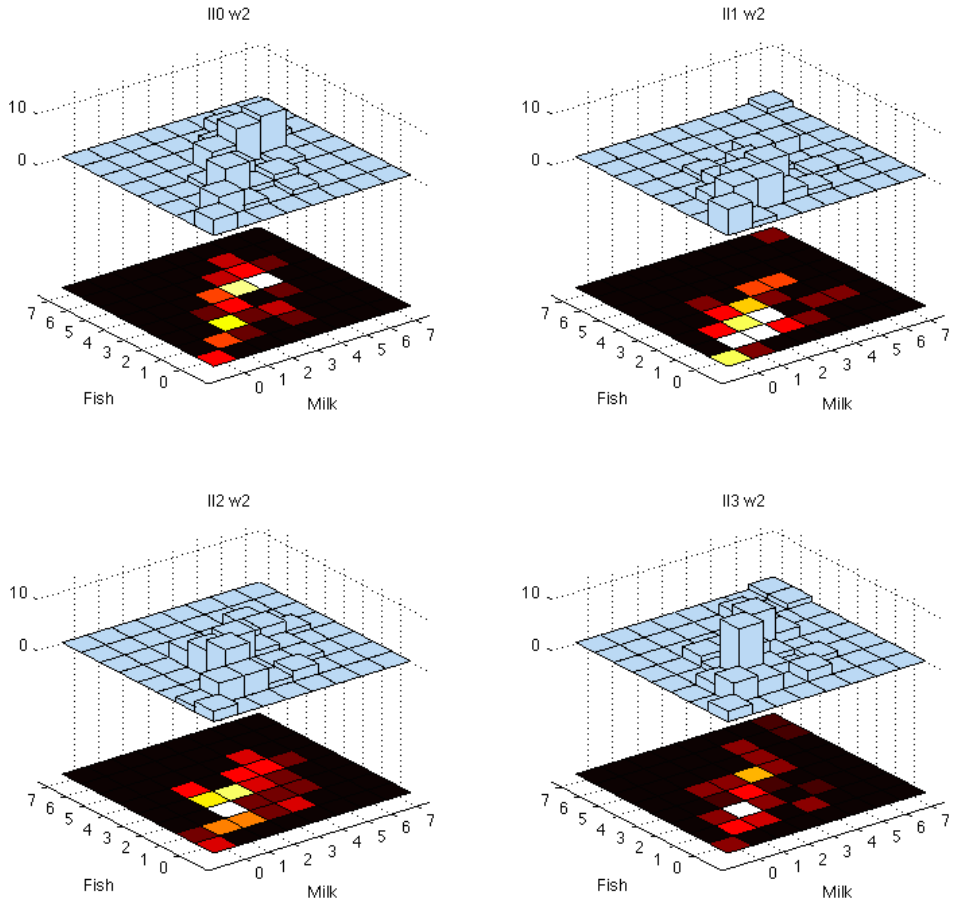
C.2.1 Week 1



Same picture in the II environment for the first week. The top-left panel is the control, top-right the CF treatment, bottom-left the information-only and bottom-right the warning treatment. We observe a greater dispersion in first week outcomes.

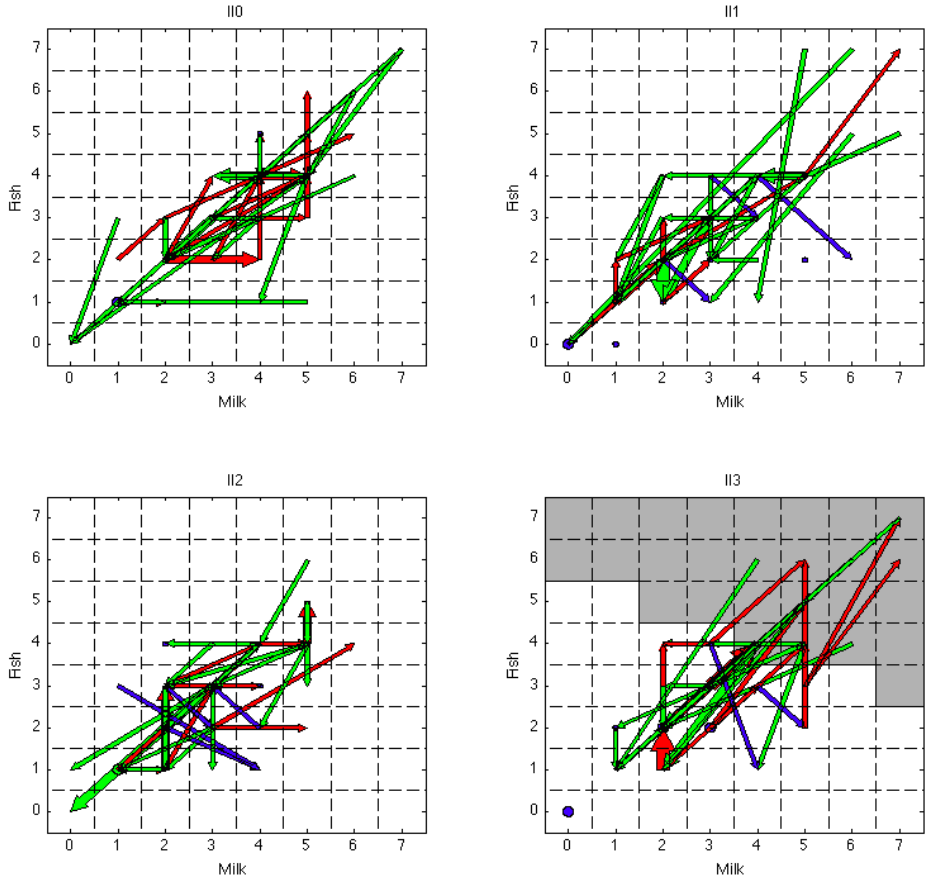
C.2.2 Week 2

We now turn to the second week of the II environment.



C.2.3 Migration

Migration patterns exhibit a significant regression to the mean in all treatments. For the warning treatment (bottom-right), the gray area corresponds to first week choices of consumers who actually received a warning (the other consumers ended-up being in a similar environment as the control group).


II0
II1
II2
II3
Fish
Milk
0
1
2
3
4
5
6
7

References

Wokje Abrahamse, Linda Steg, Charles Vlek, and Talib Rothengatter. A review of intervention studies aimed at household energy conservation. *Journal of Environmental Psychology*, 25(3):273–291, September 2005. ISSN 02724944. doi: 10.1016/j.jenvp.2005.08.002. URL `http://linkinghub.elsevier.com/retrieve/pii/S027249440500054X`.

Hunt Allcott. Site Selection Bias in Program Evaluation. *The Quarterly Journal of Economics*, 130(3):1117–1165, 2015. ISSN 0033-5533. doi: 10.1093/qje/qjv015.Advance.

Hunt Allcott and Judd B Kessler. The Welfare Effects of Nudges: a Case Study of Energy Use Social Comparisons. 2015.

Hunt Allcott and Sendhil Mullainathan. Behavior and Energy Policy. *Science*, 327: 1204–1205, 2010.

Hunt Allcott and Todd Rogers. The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation. *American Economic Review*, 2014.

Carrie Armel, Abhay Gupta, Gireesh Shrimali, and Adrian Albert. Is disaggregation the holy grail of energy efficiency? The case of electricity. *Energy Policy*, 52:213–234, January 2013. ISSN 03014215. doi: 10.1016/j.enpol.2012.08.062. URL `http://linkinghub.elsevier.com/retrieve/pii/S0301421512007446`.

Solomon E Asch. Group forces in the modification and distortion of judgments. 1952.

Shahzeen Z Attari, Michael L Dekay, Cliff I Davidson, and Wändi de Bruin Bruine. Public perceptions of energy consumption and savings. *Proceedings of the National Academy of Sciences*, 107(37):16054–59, 2010. doi: 10.1073/pnas.1001509107/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1001509107.

Ian Ayres, Sophie Raseman, and Alice Shih. Evidence from Two Large Field Experiments that Peer Comparison Feedback Can Reduce Residential Energy Usage Evidence from Two Large Field Experiments that Peer Comparison Feedback Can Reduce Residential Energy Usage. *October*, No. 15386(5):1–35, 2009. ISSN 8756-6222, 1465-7341. doi: 10.2139/ssrn.1434950. URL `http://www.nber.org/papers/w15386`.

Abhijit V Banerjee. A Simple Model of Herd Behavior. *The Quarterly Journal of Economics*, 107(3):797–817, 1992.

Gordon M Becker, Morris H DeGroot, and Jacob Marschak. Measuring utility by a single-response sequential method. *Behavioral science*, 9(3):226–232, 1964.

Roland Bénabou and Jean Tirole. Incentives and prosocial behavior. *American Economic Review*, 96(5):1652–1678, 2006. ISSN 1098-6596. doi: 10.1017/CBO9781107415324.004.

Roland Bénabou and Jean Tirole. Laws and Norms. *Working Paper*, pages 1–44, 2011. URL `http://idei.fr/doc/wp/2010/dessi_071210.pdf`.

Alan D. Berkowitz. An Overview of the Social Norms Approach. 2004.

B Douglas Bernheim. A Theory of Conformity. *Journal of Political Economy*, 102(5):841–77, October 1994. URL `http://ideas.repec.org/a/ucp/jpolec/v102y1994i5p841-77.html`.

John Beshears, James J. Choi, David Laibson, Brigitte C. Madrian, and Katherine L. Milkman. The Effect of Providing Peer Information on Retirement Savings Decisions. NBER Working Papers 17345, National Bureau of Economic Research, Inc, August 2011. URL `http://ideas.repec.org/p/nbr/nberwo/17345.html`.

Sushil Bikhchandani, David Hirshleifer, and Ivo Welch. A Theory of Fads, Fashion, Cus-

tom, and Cultural Change as Informational Cascades. *Journal of Political Economy*, 100(5):992, 1992. ISSN 0022-3808. doi: 10.1086/261849.

Dirk Brounen, Nils Kok, and John M Quigley. Energy literacy, awareness, and conservation behavior of residential households. *Energy Economics*, 38:42–50, 2013.

David P Byrne, Andrea La Nauze, and Leslie A Martin. Tell Me Something I Don't Already Know : Informedness and the Impact of Information Programs. 2016.

Hongbin Cai, Yuyu Chen, and Hanming Fang. Observational learning: Evidence from a randomized natural field experiment. *American Economic Review*, 99(3):864–882, 2009. ISSN 00028282. doi: 10.1257/aer.99.3.864.

Robert B Cialdini and Noah J Goldstein. Social influence: compliance and conformity. *Annual review of psychology*, 55(1974):591–621, 2004. ISSN 0066-4308. doi: 10.1146/annurev.psych.55.090902.142015.

Dora L. Costa and Matthew E. Kahn. Energy conservation "nudges" and environmentalist ideology: Evidence from a randomized residential electricity field experiment. *Journal of the European Economic Association*, 11(May 2010):680–702, 2013. ISSN 15424766. doi: 10.1111/jeea.12011.

Francisco J M Costa. Can Rationing Affect Long Run Behavior? Evidence from Brazil. 2013.

C Samuel Craig and John M Mccann. Assessing Communication Effects on Energy Conservation. *Journal of Consumer Research*, 5:82–88, 1978.

M Deutsch and H B Gerard. A study of normative and informational social influences upon individual judgement. *The Journal of Abnormal and Social Psychology*, 51(3): 629–636, 1955. ISSN 0096-851X. doi: 10.1037/h0046408.

Paul Dolan and Robert Metcalfe. Neighbors, Knowledge, and Nuggets: Two Natural Field Experiments on the Role of Incentives on Energy Conservation. *Centre for Economic Peformance Discussion Papers*, (1222), 2013. URL `http://ideas.repec.org/p/cep/cepdps/dp1222.html`.

Erik Eyster and Matthew Rabin. Naïve Herding in Rich-Information Settings. *American Economic Journal: Microeconomics 2*, 2(November):221–243, 2010.

Erik Eyster and Matthew Rabin. The Limits to Imitation in Rational Observational Learning. 2012.

Erik Eyster, Matthew Rabin, and Georg Weizsäcker. An Experiment on Social Mislearning. 2015.

Ahmad Faruqui, Sanem Sergici, and Ahmed Sharif. The impact of informational feedback on energy consumption - A survey of the experimental evidence. *Energy*, 35(4):1598–1608, April 2010. ISSN 03605442. doi: 10.1016/j.energy.2009.07.042. URL `http://linkinghub.elsevier.com/retrieve/pii/S0360544209003387`.

Paul J Ferraro and Michael K Price. Using nonpecuniary strategies to influence behavior: evidence from a large-scale field experiment. *The Review of Economics and Statistics*, 95(March):64–73, 2013.

Paul J Ferraro, Juan Jose Miranda, and Michael K Price. The Persistence of Treatment Effects with Norm-Based Policy Instruments : Evidence from a Randomized Environmental Policy Experiment The Persistence of Treatment Effects with Norm-Based Policy Instruments : Evidence from a Randomized Environmental Policy. *American Economic Review: Papers & Proceedings*, 101:318–322, 2014. ISSN 0002-8282. doi: 10.1257/aer.101.3.318.

L. Festinger. A Theory of Social Comparison Processes. *Human Relations*, 7:117–140, 1954. ISSN 0018-7267. doi: 10.1177/001872675400700202.

Corinna Fischer. Feedback on household electricity consumption: a tool for saving energy? *Energy Efficiency*, 1(1):79–104, May 2008. ISSN 1570-646X. doi: 10.1007/s12053-008-9009-7. URL `http://link.springer.com/10.1007/s12053-008-9009-7`.

Maria Gleerup, Anders Larsen, Soren Leth-Petersen, and Mikael Togeby. The Effect of Feedback by Text Message (SMS) and Email on Household Electricity Consumption: Experimental Evidence. *The Energy Journal*, 31(3):113–132, 2010. ISSN 01956574. doi: 10.5547/ISSN0195-6574-EJ-Vol31-No3-6.

Rema Hanna, Sendhil Mullainathan, and Joshua Schwartzstein. Learning through Noticing: Theory and Evidence from a Field Experiment. *The Quarterly Journal of Economics*, pages 1311–1353, 2014.

Ori Heffetz. A Test of Conspicuous Consumption: Visibility and Income Elasticities. *Review of Economics and Statistics*, 93(4):1101–1117, 2011. ISSN 0034-6535. doi: 10.1162/REST_a_00116.

Sébastien Houde. How consumers respond to environmental certification and the value of energy information. 2014.

International Energy Agency. Saving Electricity in a Hurry. Technical report, 2011.

Maithili Iyer, Willett Kempton, and Christopher Payne. Comparison Groups as a Tool for Evaluating Energy Efficiency Programs: An Analysis of ENERGY STAR Billing Comparison Groups, 1998.

Willett Kempton and Linda Layne. The consumer's energy analysis environment. *Energy Policy*, 22(10):857–866, 1994.

Willett Kempton and Laura Montgomery. Folk quantification of energy. *Energy*, 7: 817–827, 1982. ISSN 03605442. doi: 10.1016/0360-5442(82)90030-5.

Jungsuk Kwac, June Flora, and Ram Rajagopal. Household energy consumption segmentation using hourly data. *IEEE Transactions on Smart Grid*, 5(1):420–430, 2014. ISSN 19493053. doi: 10.1109/TSG.2013.2278477.

Wayne Leighty and Alan Meier. Accelerated electricity conservation in Juneau, Alaska: A study of household activities that reduced demand 25%. *Energy Policy*, 39(5): 2299–2309, 2011. ISSN 03014215. doi: 10.1016/j.enpol.2011.01.041. URL `http://dx.doi.org/10.1016/j.enpol.2011.01.041`.

Loren Lutzenhiser, James Woods, Susan Lutzenhiser, Rick Kundle, and Sylvia Bender. Lasting impressions: Conservation and the 2001 California energy crisis. In *Proceedings of the 2004 ACEEE Summer Study on Energy Efficiency in Buildings*, volume 7, pages 229–240, 2004.

Alan Meier. How one city cut its electricity use over 30% in six weeks. In *Proceedings of ECEEE 2009 Summer Study*, pages 1687–1691, 2009.

Cees J. H. Midden, Joanne F. Meter, Mieneke H. Weenig, and Henk J. A. Zieverink. Using feedback, reinforcement and information to reduce energy consumption in households: A field-experiment. *Journal of Economic Psychology*, 3(1):65–86, 1983. URL `http://ideas.repec.org/a/eee/joepsy/v3y1983i1p65-86.html`.

Paul Niehaus. A Theory of Good Intentions. 2014.

Jessica M Nolan, P Wesley Schultz, Robert B Cialdini, Noah J Goldstein, and Vladas Griskevicius. Normative social influence is underdetected. *Personality & social psychology bulletin*, 34(7):913–23, July 2008. ISSN 0146-1672. doi: 10.1177/0146167208316691. URL `http://www.ncbi.nlm.nih.gov/pubmed/18550863`.

Deborah Prentice and Dale Miller. Pluralistic Ignorance and Alcohol Use on Campus: Some Consequences of Misperceiving the Social Norm. *Journal of personality and social psychology*, 64(2):243–256, 1993.

Peter Reiss and Matthew White. What changes energy consumption? Prices and public pressures. *The RAND Journal of Economics*, 39(3):636–663, September 2008. ISSN 07416261. doi: 10.1111/j.1756-2171.2008.00032.x. URL `http://doi.wiley.com/10.1111/j.1756-2171.2008.00032.x`.

Simon Roberts, William Baker, and Robin Sadler. Consumer Preferences for Improving Energy Consumption Feedback. Technical Report May, 2004.

P Wesley Schultz, Jessica M Nolan, Robert B Cialdini, Noah J Goldstein, and Vladas Griskevicius. The Constructive, Destructive, and Reconstructive Power of Social Norms. *Psychological Science*, 18(5):429–434, 2007.

Kerstin Sernhed, Jurek Pyrko, and Juozas Abaravicius. Bill me this way ! – customer preferences regarding electricity bills in Sweden. In *ECEEE Summer Study - Time to turn down energy demand*, pages 1147–1150, 2003.

Frans W. Siero, Arnold B. Bakker, Gerda B. Dekker, and Marcel T. C. van Den Burg. Changing organizational energy consumption behaviour through comparative feedback. *Journal of Environmental Psychology*, 16:235–246, 1996. ISSN 02724944. doi: 10.1006/jevp.1996.0019. URL `http://linkinghub.elsevier.com/retrieve/pii/S0272494496900195`.

Dmitry Taubinsky. From Intentions to Actions: A Model and Experimental Evidence of Inattentive Choice. *Working Paper*, 2013.

Richard Thaler and Cass Sunstein. *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New York: Penguin Books, 2008.

G. Wood and M. Newborough. Dynamic energy-consumption indicators for domestic appliances: environment, behaviour and design. *Energy and Buildings*, 35(8):821–841, September 2003. ISSN 03787788. doi: 10.1016/S0378-7788(02)00241-4. URL `http://linkinghub.elsevier.com/retrieve/pii/S0378778802002414`.

Erez Yoeli, Moshe Hoffman, David G. Rand, and Martin A. Nowak. Powering up with indirect reciprocity in a large-scale field experiment. *Proceedings of the National Academy of Sciences*, 110(Supplement_2):10424–10429, 2013. ISSN 0027-8424. doi: 10.1073/pnas.1301210110. URL http://www.pnas.org/cgi/doi/10.1073/pnas.1301210110.

Jean Paul Zimmermann. End-use metering campaign in 400 households In Sweden Assessment of the Potential Electricity Savings. Technical report, 2009. URL http://www.energimyndigheten.se/ContentPages/54730465.pdf.