The Power of Negative Feedback from an Artificial Agent to Promote Energy Saving Behavior

Cees Midden and Jaap Ham

Eindhoven University of Technology Department of Human-Technology Interaction PO Box 513, 5600 MB Eindhoven +31 40 247 3446

Abstract. In this paper we analyze the role of negative feedback as provided by artificial agents. We examine the hypothesis that negative feedback offers substantial potential to enhance persuasive interventions aimed to change behavior. This hypothesis is tested based on a review of several studies using the same experimental paradigm that includes a virtual washing machine, in which users have to make choices how to program the washing machine. The studies show how the provision of positive and negative feedback influences these choices under various experimental conditions. Results show that negative feedback can be more effective than positive feedback, also independent of the presence of positive feedback. Negative feedback is in particular effective when the feedback is social instead of factual. Furthermore, the analysis suggests that the effect of negative feedback is enhanced under conditions of task similarity, which stimulate using the feedback for performance improvement. Finally, we show that negative feedback is superior to positive feedback under multiple goals conditions.

Keywords: Persuasive technology, artificial social agents, social evaluation, sustainability.

1 Introduction

The exhaustion of natural resources and the threats of growing CO2-emissions and climate change effects have urged nations worldwide to seek for substantial reductions in energy consumption. Although technological solutions like more efficient systems and devices and the development of renewable energy sources are of great importance, consumer behavior plays a crucial role in bringing down the level of energy consumption. Influencing consumer behavior to promote energy conservation has become an important target of national and international policy efforts. Thereby, the question which instruments should be applied to promote energy conservation behavior has become highly relevant.

Recent reviews (e.g. 1) have evaluated the effects of interventions to promote energy efficient behavior. Raising people's awareness of energy consumption by providing tailored feedback about their energy consumption (for example in kWh) can promote the achievement of behavioral change. The results are mixed though. Weak linkages between specific actions and energy outcomes caused by low feedback

frequencies (e.g. once month) and insufficient specificity of the feedback (e.g. household in general vs. specific person or specific devices) are underlying these findings. Recently, technological solutions have created new opportunities to improve feedback efficacy by embedding it in user-system interactions. That is, energy use is almost always the outcome of an interaction between a user and some energy-using device. Intervening in these specific interactions could improve the quality of feedback substantially. Some evidence supports this claim. McCalley and Midden (2) demonstrated in several studies that interactive forms of feedback could be effective to enhance energy efficient use of devices like a washing machine. By adding an energy meter to the user interface of a washing machine they achieved 18% of energy conservation both in lab and field studies. Basically, their approach entailed giving factual feedback in terms of kWh consumed as a function of programming choices made by the user, like water temperature, spinning speed or the duration of the washing cycle.

This work was followed up by new attempts to increase the persuasiveness of the system through the introduction of social feedback (e.g. 3, 4). We examined whether social feedback can add to the promotion of pro-environmental behaviors such as energy conservation in the home. Social reinforcement has been applied widely in many domains such as child education, therapeutic programs, health behavior and social interaction as a mechanism for behavioral change (2). Social praise and compliments operate as positive incentives. In previous studies we demonstrated how the effectiveness of social reinforcements as delivered by human actors, can successfully be provided by an intelligent system (see 4 for an overview).

Surprisingly, social feedback by smart computer agents has mainly focused on positive social feedback. One of the exceptions is research by Bracken, Jeffres, and Neuendorf (5), which studied the influence of praise or criticism in feedback by a computer on user experiences (e.g., motivation). Many researchers seem to assume that positive feedback is more effective than negative feedback. There are some legitimate reasons for that. In our own research we found that direct feedback that constrains user choices may lead to reactance responses, which are detrimental to message adoption and behavioral change (6). Also negative information may not work, because it does not specify, by itself, how a person should respond. Yet, in the current paper we conjecture that the effects of negative social feedback are underestimated and can contribute substantially to effective interventions to change behavior. Although negative feedback may be less pleasant for the user, there are reasons to expect significant outcomes from negative feedback.

In a more general sense negative events show, almost universally, a higher impact than positive events and bad information is usually processed more intensely than positive information (7). In an evolutionary sense negative information is more valuable for adaptation than positive information. Ignoring danger is in general more threatening for survival than missing a positive opportunity. One could say that negative information has a higher level of diagnosticity (8). One accident may make a system unsafe, while long periods of flawless functioning are necessary to create a feeling of safety. Information diffusion studies show that bad news travels faster through networks than positive news (9). Similar patterns can be observed regarding feelings of trust. Trust is an important social emotion that allow individual to accept risky decisions that may produce positive outcomes, but also negative outcomes.

The saying, trust comes on foot, but leaves on horseback' suggests that it may require a lot of effort to build trust, but a single disappointing experience may breakdown a trustful relationship. This prominence of negative information is also predicted by Prospect theory (10), which suggests that potential negative consequences have a higher impact on decision making than potential positive consequences. Negative information can also be important because it signals a need for change, thereby feeding the self-regulatory system, through which an individual can adapt to changing circumstances (11).

Following this reasoning we included negative feedback in our persuasive feedback studies and expected negative feedback to contribute to the persuasive impact of the feedback. Although this issue was initially not at the core of our attention in most of those studies, we found remarkable results on the distinction between positive and negative feedback. In the current paper we reanalyzed our studies with this specific hypothesis in mind. First, we show in Study 1 and 2 how negative feedback influences behavior change. We show how it worked in combination with positive feedback and separate from positive feedback. Next, we discuss the results of Study 3 showing that negative feedback is in particular effective when the context offers an opportunity for adaptation and learning based on the feedback. Finally we present in Study 4 new data that suggest that positive feedback is in particular effective when a single goal has been activated, but fails under conditions of multiple goals. Negative feedback, by contrast, is not only effective when a single goal has been activated, but retains its effectiveness when multiple goals force the user to make trade-offs.

2 Study 1: Social vs. Factual Feedback and Feedback Valence

In our first study (3) about the persuasive effects of feedback on behavior by a smart social agent we set up an experiment in which subjects received social feedback from an artificial robotic agent while carrying out tasks in which they could conserve energy. More specifically, we tested the effects of social feedback compared to factual feedback, the effect of positive vs. negative effect and thirdly the effect of low vs. high perceived agency as characteristic of the feedback source. The latter effect is not relevant for the current paper and will not be discussed here. Thirty-three participants were randomly assigned to one of three experimental conditions. The experimental procedures have been explained elsewhere in detail (3). In short, participants completed 10 washing trials in which they programmed the washing machine (see Figure 1) by making various choices such as for water temperature and spinning speed. They received the factual feedback from the computer system displaying the amount of electricity (in Kwh) consumed. The feedback was presented through a little led-indicator on the display. In the social feedback condition the feedback was provided by a little robotic agent, which is known an iCat, produced by Philips corporation (ref) in the form of stylized head of a cat that was able to display social expressions by moving lips, eyes, eyelashes and eyebrows and by playing speech files. For all participants, a simulated washing machine, a copy of a current model on the market, was presented., while the social feedback was provided by the iCat and selected from a repertoire of six positive and six negative evaluative expressions like fantastic, good, bad or awful.

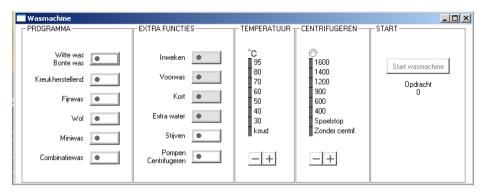


Fig. 1. The virtual washing machine interface

The results supported our expectations. The participants who received social feedback saved more energy than the participants who received the factual feedback. More importantly for the current analysis, findings showed a feedback valence effect. More specifically, these findings showed that the greatest changes to conservation behavior were achieved after negative feedback (compared to positive feedback). In addition we found an interaction indicating that the negative feedback effect was strongest when the social artificial agent, the iCat, provided it.

3 Study 2: Negative Feedback Only

Before drawing firm conclusions, a second study was performed first to replicate these results and second to test whether the negative feedback effect would also occur independent of the positive feedback. Therefor in Study 2 feedback valence was manipulated in a between-subjects design (3). So, while in Study 1 participants could receive both positive and negative feedback, in Study 2 they got either positive or negative feedback. Like in Study 1 factual feedback and social feedback were tested between-subjects. However, we also adapted the factual feedback condition to control for a potential confound between the social nature of the feedback and the evaluative nature of the feedback. On could argue that the social feedback was more effective, not because of the social nature of the feedback, but instead, because it was more directly evaluative compared to the factual feedback. To control for this potential confound we substituted the factual feedback with the led kWh-indicator for one that used colored lighting to provide factual feedback with red light for negative feedback and green light for positive feedback, both at 6 levels of intensity. In the negative feedback condition, factual evaluative feedback was given when participants used more energy than the middle of the scale. This was done by gradually changing the color of the screen from white to red and back within 3 seconds. In addition, an unpleasant sound (a buzzer) was played through the speakers for one second. The strength of the red color and the loudness of the sound depended on the amount of energy consumed by a participant's settings. If e.g. a setting used much energy, the screen changed color to dark red, and a loud buzzer sounded. In positive feedback conditions, factual evaluative feedback was given in an identical fashion, but now the screen color changed from white to green, and the sound played was a pleasant ring. Level of detail of the factual-evaluative feedback was similar to that in social feedback conditions. This setup equalized the evaluative nature of the feedback between the conditions, but kept the distinction between social and non-social feedback. For the remainder the setup was similar to Study 1.

Results were in support of the conclusions of Study 1. Social feedback proved to be more effective than factual feedback, both with kWh indicator and the colored light condition. The difference between both types of factual feedback was not significant. More important for the current analysis, the main effect of feedback valence was replicated in this between subjects design, demonstrating that the effect of negative feedback occurred without the availability of positive feedback. Furthermore we found the same interaction indicating that the effect of negative feedback was strongest for social feedback compared to factual feedback. The latter single effect proved to be non-significant. Moreover it seemed as if the positive feedback encouraged participants to enhance their energy consumption. Thus, this last finding suggests that while people may feel pleased by a system that offers compliments, this does not ensure that the user may also change her or his behavior. In particular this may happen when multiple goals have been activated. We will address this issue in Study 4.

Together Study 1 and 2 demonstrated the effectiveness of negative feedback and in particular when the feedback was social in nature. These rather surprising effects of negative feedback evoked the question why these effects occurred. This was the reason for setting up Study 3 in which we explored conditions that facilitated the working of negative feedback.

4 Study 3: Learning from Negative Feedback

We argued that an important reason for the effectiveness of negative feedback is that negative feedback is especially useful in situations that allow people to avoid aversive consequences by learning from previous experiences. It has been argued that negative feedback is not effective because, in contrast to positive feedback, it does not learn how to avoid the negative consequences (12). In other words it would not stimulate learning sufficiently. Consistent with this explanation also in other fields similar hypotheses have been tested. Earlier research in the domains of risk management (13) suggested that people seek information about impending dangers only when they see a possibility to avoid negative consequences. In line with these findings, we argue that the effects of negative feedback will be moderated by the extent to which options are available to avoid negative consequences. This could for instance be the case when tasks that people have to perform are relatively similar (as compared to tasks that are dissimilar). In such a task context feedback can provide better options to learn from previous negative experiences because these may offer indications for improvement in follow-up tasks. So, we argued that task similarity would enhance the effects of negative feedback. Therefore, in Study 3 we investigated the hypothesis that the persuasive effect of negative social feedback vs. positive social feedback as provided by an artificial social agent will be enhanced under conditions of high task similarity in comparison to low task similarity (14).

In this study in which 120 participants participated we manipulated in a 2x2 design high vs. low task similarity and positive vs. negative social feedback by the artificial agent. The experimental setup followed the setup and procedures of Study 1 and 2. Task similarity was manipulated by distinguishing a set of washing tasks with high similarity vs. one with low similarity. This distinction was based on a pretest in which we asked 27 participants to sort 16 washing tasks on similarity. The outcomes of the pretest were used to compose a set in which subsequent tasks were rather similar vs. a set in which subsequent tasks were rather dissimilar. So, the 16 tasks were the same in both conditions, but only the order of the tasks was different to induce the desired effect. For example, in the high similarity condition tasks including the same type of material, for instance woolen or cotton wear.

The results confirmed again the main effect of feedback valence showing a higher level of behavior change as a result of negative feedback. Moreover, a significant interaction was found showing that the effects of negative feedback, compared to positive feedback, was enhanced in the task similarity condition, thereby supporting our hypothesis.

In sum, this study showed that negative feedback is able to contribute to behavior change interventions if the right learning conditions are met. In the next section we will argue that negative feedback can also compensate for weaknesses of positive feedback.

5 Study 4: Serving a Single Goal vs. Multiple Goals

Without doubt, lack of clarity about the role of set goals has been one of the major reasons why results of feedback have been mixed in the past (15). It was often presumed that energy feedback would allow users to see where they stood in relation to their energy consumption goal and thus would automatically seek to reach it. However, residents may or may not have had an active goal to save energy, and of those that did, many may have given this goal a lower priority than other goals, such as comfort or convenience. In fact, the earlier studies that evidenced some success of energy feedback either assigned a specific energy saving goal to subjects and requested various forms of commitment to the assigned goal (2, 16, 17) or were likely to have unknowingly triggered an energy saving goal in the subjects prior to the experimental treatment (e.g. 18). Those that did not engage the user in some form of goal setting showed no success of lowering consumption through feedback. Possibly others were likely to have unknowingly triggered an energy saving goal in the subjects prior to the experimental treatment (18).

The relevance of goal setting has also been demonstrated for persuasive technology using smart agents. The study by (2) illustrated that goal setting can be a highly effective means of assuring response to feedback in the washing machine paradigm as also used in the Studies 1-3 of the current paper.

In spite of the significance of these results, we should wonder to what extent these results match real life circumstances. As far as we know, studies that investigated goal setting in relation to the effectiveness of feedback, regarded only single goals, that is a goal to conserve energy was set, or activated, or not. In real life however, people usually have multiple goals when making decisions to conserve energy. Even more, energy conservation is usually not the prime goal for using appliances. Most of the time, actions will connect to other major goals like heating the home, cooling foodstuffs or cleaning the laundry. One may wonder how the presence of these other goals could affect the pursuit of a conservation goal? In general, it seems likely that if a user has to divide his or her attention between multiple goals, each goal would receive less attention. In particular, this will be the case if goals are not compatible, which forces a user to make a trade-off between the achievement of both goals. We suggest that for many users the prime need for using energy will dominate the goal to conserve energy. In others words when people have to make a choice, often energy conservation will not prevail. Actually, this may be a reason why adoption of sustainable technology in general is hampered. Producers of sustainable technology improve a product attribute of secondary significance for the consumer, the exception being the case that resource conservation is among the dominant goals of action (15).

If consumers would tend to weigh their prime goal of use higher than the conservation goals, what could be the effect of feedback? Most likely, the user will be more alert on the achievement of the prime goal rather than the conservation goal. That is why we hypothesized that positive feedback on the conservation goal could generate a response that turns attention away from the conservation goal and towards the prime goal. After all, positive feedback may suggest that the conservation goal has been achieved to a reasonable extent. Most likely, negative feedback will not induce that conclusion to the same extent.

Note, that in all three preceding studies that we described in the current paper actually multiple goals were present. We asked our participants to perform washing tasks to clean the laundry in proper way. In addition we asked them to conserve as much energy as possible. Thus, our participant had to make the tradeoff between the two goals. This may be one of the explanations why negative feedback works. One might even reason that in most studies on energy feedback this trade-off of multiple goals has played a role in the decisions of users. In most cases, however, this decision aspect was not articulated in the research. This could also be part of the explanation why the results of feedback studies have been mixed.

On the basis of these considerations on the relationship between multiple goals and feedback, we hypothesized that in a multiple goal situation in which a conservation goal competes with a goal of use, negative feedback may exert stronger effects on the achievement of conservation goals compared to positive feedback (Hypothesis 1). Furthermore we expected that participants aiming for a conservation goal while a (consumption) goal was also activated, would consume more electricity than participants who are aiming for a conservation goal only. (Hypothesis 2). Finally, we expected to replicate the finding of higher savings as a result of negative feedback compared to positive feedback (Hypothesis 3).

5.1 Method

Participants and design. 73 participants (35 men and 38 women), all students at Eindhoven University of Technology, were randomly assigned to one of four experimental conditions: 2 (feedback condition: single goal activated vs. multiple goals activated) x 2 (feedback type: positive vs. negative). In addition, each participant completed 10 washing trials, which composed the 10 levels of our third independent variable.

Materials and procedure. The experimental procedure was identical to that of Study 1, except that only social-evaluative feedback was given and goal setting was manipulated. The expressions and behavior of the iCat were similar to the previous studies.

Participants were asked to complete the washing trials on the simulated washing machine panel that was presented on the screen. In the multiple goal condition two goals were assigned: clean your clothes and save energy. In the single goal condition they were only requested to save energy. Each participant completed a practice washing trial and the ten real washing trials, which were identical for all of the participants. For each trial, participants were instructed to complete a specific type of washing task (e.g., "wash four very dirty jeans").

After the participants completed all trials, they answered several demographic questions, were debriefed and thanked for their participation.

To be able to distinguish the effects of positive and negative feedback we calculated an index based on total number of actions of users in the user interface. So, we not only included the final choices per trial, but all the preceding programming choices. As explained, these were all followed by social feedback, either positive or negative. The index subtracted for each action the following choice, in terms of energy consumption effect, from the current choice, thereby indicating whether the feedback resulted in a higher or lower energy consumption score for the next following choice. This procedure was similar to those in the previous studies.

5.2 Results

The electricity consumption score was analyzed using a 2 (goal: single vs. multiple) × 2 (feedback type: negative vs. positive) mixed model analysis, controlling for multiple observations per participant. Analyses confirmed our expectations regarding Hypothesis 1 and 3. The main effect of the goal factor did not reach significance (Single goal M = -0.11, SD = 0.01; Multiple goal M = -.09, SD = 0.01), F(1, 59,342) = 2,186, p = 0.14). More specifically, these analyses indicated that participants who had been provided with positive social feedback saved less energy (M = .-.05, SD = .009) than participants who had been provided with negative (social) feedback (M = -.16, SD =.01), F (1, 557,676) = 80,586, p < .0001. Secondly, these analyses confirmed the interaction between feedback valence \times # of goals, F(1, 557,676) = 4,863, p = .0.028 (see Figure 2). Specific comparison within the mixed model analysis (using Bonferroni adjustments for multiple comparisons) showed that the simple effect of positive feedback significantly reduced under multiple goal conditions compared to single goal. This effect did not occur for negative feedback. Negative feedback was effective in saving energy under single goal conditions but remained so under multiple goal conditions.

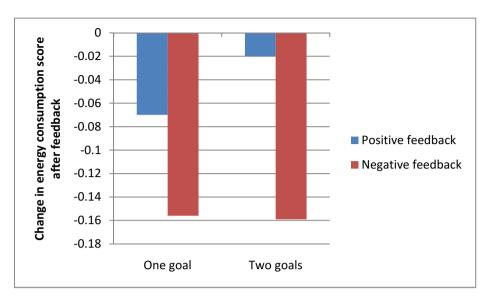


Fig. 2. Change in energy consumption score after positive vs. negative feedback under one vs. two goal conditions in Study 4

6 Discussion

In general, our findings demonstrated that people are sensitive to social feedback as provided by an artificial agent. Apparently, persuasive artificial agents are able to create behavior change among human users. In contrast to earlier work that focused primarily on praise, our feedback interventions included both positive and negative feedback (Study 1). The effect of social feedback on energy conservation should therefore be considered as the joint effect of both positive as well as negative feedback. Interestingly however, our analysis on the level of single programming acts within trials, suggested that the direct influence of negative feedback on following programming choices was greater than the effect of the positive feedback, Furthermore Study 2 demonstrated convincingly that providing negative feedback only is more effective than providing positive feedback only. This finding suggests that the effectiveness of negative feedback is not dependent on the presence of positive feedback. It is however related to the social nature of the feedback. The effects of negative feedback were especially noticeable for social feedback and less for factual feedback. Apparently, the social nature of the feedback amplifies the difference between positive and negative feedback. Results of Study 3 indicated that negative feedback was effective especially when tasks were relatively similar—as is the case of many of the daily routine tasks people perform. Enhanced task similarity creates a context in which the feedback can immediately be employed for improving performance. In this sense negative feedback contributes to self-regulatory action by the user. This function is weaker when the context does not offer opportunities for performance improvement based on the negative feedback. This process may have influenced the

effectiveness of negative feedback in most of our studies using the washing machine paradigm. This paradigm offers a structured environment in which users are requested to make specific choices while using a well-defined interface. In this context it becomes easily clear for the user how to enhance performance when this is indicated by the feedback. In ill-structured environments, it could be necessary to offer the user additional cues how to improve performance.

The effects of positive feedback were relatively minor in our studies. Moreover, it seemed as if the positive feedback encouraged participants to enhance their energy consumption. However, this was not generally the case. Thus, while people may feel pleased by a system that offers compliments to a user, an effect that has also been observed in other studies (e.g. 19), this does not ensure behavior change. In particular, this may hold if a user has other goals to achieve, like in our case cleaning the laundry. In making the trade-off the user may decide to stop focusing on the goal that already received positive feedback. Possibly this effect is also influenced by the goal hierarchy. On might expect that the higher goal will be pursued to a greater extent when a trade-off has to be made. Future research could further examine this factor.

In our introduction we have discussed the debate about negative feedback and negative information in various research areas. Our results support the notion that negative feedback can be effective in informing people about their performance or about the need to change actions. One might wonder whether this effect is influenced by the artificial nature of the feedback sources in our research compared to human sources in many other studies. Possibly, people can more easily accept negative information from a non-human source than from a human source. Although our reactance studies showed that also artificial agents could evoke reactance, it seems plausible that tendencies to respond defensively are weaker when the feedback originates from an artificial agent. This is one of the challenging questions that lie ahead in exploring and optimizing the persuasive power of artificial agents.

References

- Abrahamse, W., Steg, L., Vlek, C., Rothengatter, T.A.: review of intervention studies aimed at household energy conservation. Journal of Environmental Psychology 25, 273–291 (2005)
- McCalley, L.T., Midden, C.J.H.: Energy conservation through product-integrated feedback: The roles of goal-setting and social orientation. J. Econ. Psychol. 23, 589–603 (2002)
- 3. Midden, C., Ham, J.: Using negative and positive social feedback from a robotic agent to save energy. In: Conference Proceedings of Persuasive 2009, Claremont, USA. pp. article no.12. Springer, Heidelberg (2009)
- Midden, C.J.H., Ham, J.R.C.: Persuasive Technology to promote pro-environmental behaviour, in. In: Steg, L., van den Berg, A.E., de Groot, J.I.M. (eds.) Environmental Psychology an Introduction, West-Sussex / UK (2013)
- Bracken, C.C., Jeffres, L.W., Neuendorf, K.A.: Criticism or praise: The impact of verbal versus text-only computer feedback on social presence, intrinsic motivation, and recall. CyberPsychology and Behavior 7, 349–357 (2004)

- Roubroeks, M., Ham, J., Midden, C.: When artificial social agents try to persuade people: The role of social agency on the occurrence of psychological reactance. International Journal of Social Robotics 3, 155–165 (2011)
- 7. Rozin, P., Royzman, E.: Negativity bias, negativity dominance, and contagion. Personal SocPsychol Rev. 5, 296–320 (2001)
- 8. Slovic, P., Finucane, M.L., Peters, E., MacGregor, D.G.: Risk as analysis and risk as feelings: Some thoughts about affect, reason, risk, and rationality. Risk Analysis 24, 311–322 (2004)
- Naveed, N., Gottron, T., Kunegis, J., Che Alhadi, A.: Bad news travels fast, A contentbased analysis of interestingness on twitter. In: Proceedings of the ACM Web-Sci. (2011)
- Kahneman, D., Tversky, A.: Prospect theory: An analysis of decision under risk. Econometrica 47(2), 263–291 (1979)
- Baumeister, R.F., Bratlavsky, E., Finkenauer, C., Vohs, K.D.: Bad is stronger than good. Rev. Gen. Psychol. 5, 323–370 (2001)
- 12. Skinner, B.: Utopia through the control of human behavior. In: Rich, J.M. (ed.) Readings in the Philosophy of Education. Wadsworth, Belmont (1972)
- Rogers, R.W., Prentice-Dunn, S.: Protection motivation theory, Gochman, D.S. (ed.).
 In: Handbook of Health Behavior Research 1: Personal and Social Determinants, ch. 6.
 Plenum Press, New York (1997)
- 14. Ham, J., Midden, C.: A persuasive robotic agent to save energy: The influence of social feedback, feedback valence and task similarity on energy conservation behavior. In: Conference Proceedings of Social Robotics, Singapore (2010)
- Midden, C.J.H., Kaiser, F.G., L McCalley, L.T.: Technology's Four Roles in Understanding Individuals' Conservation of Natural Resources. Journal of Social Issues 63(1), 155–174 (2007)
- Becker, L.J., Seligman, C.: Reducing air conditioning waste by signaling it is cool outside. Personality and Social Psychology Bulletin 4, 412–415 (1978)
- 17. van Houwelingen, J.H., van Raaij, W.F.: The effect of goal-setting and daily electronic feedback on in-home energy use. Journal of Consumer Research 16, 98–105 (1989)
- 18. Seligman, C., Darley, J.M.: Feedback as a means of decreasing residential energy consumption. Journal of Applied Psychology 62, 363–368 (1977)
- Fogg, B.J.: Persuasive technology: Using computers to change what we think and do. Morgan Kaufmann, Amsterdam (2003)