

Energy UX: Leveraging Multiple Methods to See the Big Picture

Beth Karlin¹✉, Sena Koleva¹, Jason Kaufman¹,
Angela Sanguinetti², Rebecca Ford³, and Colin Chan⁴

¹ See Change Institute, Los Angeles, USA
bkarlin@seechangeinstitute.com

² UC Davis, Davis, USA

³ University of Oxford, Oxford, UK

⁴ Yardi Energy, Vancouver, Canada

Abstract. Engaging the public to decrease their carbon footprint via energy feedback has become a significant topic of both study and practice and understanding how to best leverage technology for this purpose is an ideal question for the field of HCI to address. One common example is Home Energy Reports (HERs) and Business energy reports (BERs), which are paper or electronic reports that display a consumer's energy use alongside various benchmarks and "tips" to help (and persuade) them to save energy. While HERs and BERs show great promise, average savings hover around 1–3% with the potential savings in the average home and/or business closer to 15–20%, leaving potential room for improvement. This paper presents a mixed-methods research framework that is being used to improve BER user experience and energy savings. It blends inductive research methods from the fields of design and HCI with deductive methods drawn from psychology and behavioral economics to develop and test hypotheses and translate findings into real-world application. After introducing the framework, a case study is presented in which these steps are followed over two years of research with one BER product across multiple utility pilots. Implications for both energy feedback specifically as well as suggestions on how this framework can be applied across the broader field of usability are discussed.

Keywords: Energy · Feedback · Usability · Psychology · Multi-disciplinary

1 Introduction

One potent route to changing the way that people use energy, particularly motivating them to conserve, is providing feedback on their energy use [1]. The efficacy of feedback operates under the assumption that most people are uninformed about the relationship between their behaviors and their energy use and, in a larger sense, their impact on the environment; feedback is an attempt to “bridge this ‘environmental literacy gap’” [2]. Energy feedback shows promise for supporting energy reduction (or energy shifting), though the results on the extent of its impact are mixed. While much research has investigated effectiveness of feedback (vs. a control group), the design of feedback interventions that may increase energy savings has received less attention.

Energy feedback falls squarely in the world of HCI research, as it involves the presentation of energy use data (often collected from the smart meter) back to consumers [2]; these communications necessarily must consider the interface with its human consumers. To be effective energy feedback must not only reach and engage consumers, but should also effectively lead them to change their behavior in some way. Persuasive HCI in this domain should thus consider both design elements and the impact of these elements on customer behavior [2]. Different empirical methods reveal different insights into these processes; thus, understanding how to use them together and individually is key to maximize learning.

This paper presents a multi-method framework for effectively designing HCI to engage consumers in energy saving behavior. To illustrate this framework, we detail a case study of approximately two years of collaborative research with Yardi Energy. Yardi offers a diverse range of products in the area of asset management, energy cost analysis, and energy efficiency software for real estate companies and energy utilities. One of these products is their Business Energy Report (BER) program. BERs are paper or electronic reports that communicate computer-generated energy use data to commercial utility customers and encourage them to use less energy, either through capital investments in efficiency upgrades or through behavior change. BERs include features such as benchmarking (e.g., comparisons between customers' energy use and their peers) and energy-saving "tips," (e.g., information on efficient lighting). Our central question throughout this research was: "How can the design elements of BERs be leveraged to most effectively support energy reduction goals?"

2 Background

Feedback, the process of providing information about the result of an action that can be used to reinforce and/or modify future behavior, is considered an important dimension of behavior change [3] and has been used in fields ranging from education [4] to health behavior [5]. Eco-feedback, specifically, refers to feedback that is provided about behavior with the goal of reducing environmental impact [2].

Eco-feedback has received much attention in the energy space over the past decade due to changes in the electricity grid that enable real-time data collection and the processing of large data sets [6], and the proliferation of new feedback products on the market [7]. A recent meta-analysis of energy feedback studies [1] identified several variables that moderate its effectiveness, including frequency, medium, comparison message, duration, and combination with other interventions (e.g. goal, incentive). However, these variables have not been experimentally manipulated in most of these studies, with a focus on testing feedback vs. a no-feedback control group. Further, design variables have not been tested consistently for their role in feedback effectiveness.

A review of past empirical research on eco-feedback identified two primary fields that have conducted studies in this area: psychology and HCI [2]. The authors found that the HCI literature was primarily focused on the visual design of eco-feedback using qualitative methods whereas psychology literature was primarily focused on the effectiveness of eco-feedback interventions using experimental methods; there was

little overlap between the two field in terms of references and citations. They concluded that “perhaps a future goal for HCI should be to initiate collaborations with environmental psychologists” [2].

3 Research

The current research presents a framework that integrates methods from the HCI and psychology fields; it progresses from observational methods with a “user-centered” focus to the development and testing of hypotheses, and finally to the optimization of eco-feedback design and the testing of the generated materials in the field. It has three phases: (1) Observe, (2) Test, and (3) Apply (see Fig. 1 below). The first phase, “Observe,” involves customer research and literature review; the objective is to assess the opinions and thoughts of potential users in an inductive fashion and review and integrate relevant literature from across disciplines. This phase serves to generate hypotheses and inform methods to test them in the subsequent “Test” phase. This phase includes user testing and A/B testing to measure both stated and revealed preferences of design features with users. In the final “Apply” phase, insights from testing are used to optimize the design and launch it with actual users in an applied setting.

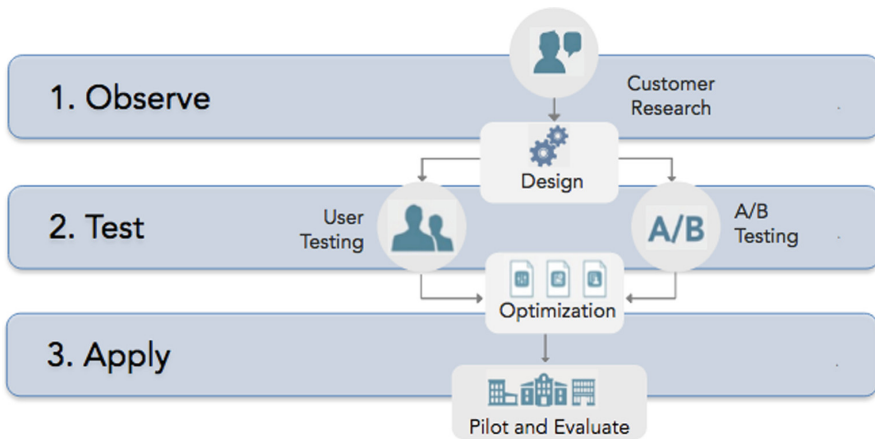


Fig. 1. Research framework

This framework is both multi-disciplinary and iterative. While the benefits of individual elements are recognized by researchers and practitioners alike, they are often used in isolation rather than as part of a holistic, programmatic approach. Thus, the framework’s unique strength lies in how various methods can work together and inform one another. To that end, the process does not stop at the “Apply” phase; one can also evaluate the effectiveness of the final product “in the real world” and use these results in restarting the process. For example, the findings might show that certain materials were more successful for one group of recipients than another, at which point

one can go back to the literature (“Observe”) for clues on relevant group differences and/or redesign and test new materials with these insights in mind (“Test”).

While a variety of research methods can be employed through each phase, certain methods are particularly well-suited to achieve the most common goals of each one. Table 1 summarizes each method, its benefits, and in which phase of the process it can be used. Below this table, we describe each phase in more detail and in the next section present a case study, with resulting findings, from each method.

Table 1. Multidisciplinary methods for studying human-computer interaction

Method	Description/benefits	Observe	Test	Apply
Literature review	Understand research and theory across disciplines	x		
Ethnography	Assess attitudes, feelings, actions, and barriers of potential users	x	x	
Content analysis	Reduce textual data to a set of codes with specific definitions	x	x	
Focus groups	Ask a fixed set of questions to several participants in a group setting	x	x	x
Interviews	Ask a fixed set of questions to one participant in a one-on-one setting	x	x	x
Surveys	Collect both quantitative and qualitative self-reported data at scale	x	x	x
Eye tracking	Understand how users interact visually with a designed image		x	
User testing	Systematically walk a user through interactions with a product		x	
Experimental design	Understand differences in response between multiple versions		x	x

3.1 Observe Phase

The goal of this phase is to remove or reduce preconceptions regarding what may be effective and simply observe, both the users that we plan to target, as well as existing relevant research and theory across disciplines. This step is similar to the first step in the design thinking process – Empathize – which suggests that one should not approach a problem with a potential solution in mind, but instead by “immersing yourself in [the] experiences [...] of the user for whom you are designing” [8]. While there are multiple methods that can be used in this phase (e.g., surveys, focus groups, ethnography), we discuss the way we used literature review and interviews in our case study.

Literature Reviews. Perhaps the broadest method for establishing a general understanding of a topic or theme is to examine past theory and research on the subject. For example, in brainstorming new ways of framing BER content, we conducted a literature review on lesser-utilized types of persuasion, such as Omega strategies. Prior work in eco-feedback has largely been rooted in Alpha strategies, which focus on increasing

motivation to move towards a goal (e.g. by increasing the trustworthiness of the message source). In contrast, Omega strategies promote change by reducing the motivation to move away from the goal (e.g. addressing barriers). Reviewing the literature thus provided us an expanded pool of strategies for testing and improving BER content that we would have missed without taking this step.

Another topic in our literature review that led to testable hypotheses was the use of images in sustainability communication. We located studies that showed that visual content aids in information recall [9], that people prefer vertical bar charts for historical benchmarking of energy usage and horizontal bars for social benchmarking [10], and that images of people dominate visual representations of climate change and so could fit more closely with customers' existing mental representations on the topic [11]. Based on this "observation" of the literature, we hypothesized, and subsequently found, that the images in BERs attract readers' attention and that images of people in BER tips could increase their appeal.

Interviews. Interviews can be used to assess user beliefs, needs, behaviors, or barriers to action. They can be conducted in-person or remotely (e.g., by phone) and can be free form (following a natural conversation with very little pre-determined content), structured (following a script of set questions), or semi-structured (set questions with allowances for deviation or deeper probing). Interviews can provide an in-depth understanding of the characteristics and mental processes of individual participants. In 2016, Yardi's data science team identified a set of small-medium business (SMB) customers who were using heating, ventilation, and air conditioning (HVAC) while their businesses were closed. Before jumping into designing an intervention, we wanted to get a better sense of these businesses' perceptions of their heating and cooling use. We conducted phone interviews with a random selection of these pre-identified customers. Rather than asking them why they used their HVAC during closed business hours (which might prematurely steer the conversation or make them defensive) the interview protocol instead assessed their awareness, attitudes, and actions surrounding their HVAC use through general questions (e.g., "How is your air conditioning system at your place of business controlled?").

This design was critical as it allowed us to discover that, despite the fact that Yardi had specifically identified these businesses as using HVAC systems during closed hours, 100% of the interviewees claimed that they turned their cooling either down or completely off during those times. This finding revealed a vast disconnect between how SMB owners think they are using energy and how they are actually using energy. Participants agreed that HVAC use during closed hours is useless and expressed interest in HVAC efficiency to cut business costs. Thus, while we started this research to understand why SMB owners were using their HVAC systems during closed hours, we came to a very different conclusion: that many SMB owners are running their HVAC systems during closed hours unknowingly. This fundamentally altered our approach; for example, instead of suggesting to SMB owners that they should not use their HVAC systems during closed hours, we needed to first help them realize that they were unknowingly doing so.

3.2 Test Phase

While the methods of the “Observe” phase can provide an unbiased understanding and help to develop hypotheses about how to design interventions, the methods of the “Test” phase can test these hypotheses using actual prototypes and controlled experiments. Through iteration, these hypotheses can be refined and used to design and optimize operational prototypes. The two “Test” phase methods we discuss in our case study are usability testing with eye tracking and experimental design (aka A/B testing).

Usability Testing with Eye Tracking. Eye-tracking uses specialized computer hardware to track the eye movements of a user as (s)he looks at material on a screen or page. Eye-tracking is best used when it is thought that someone’s gaze direction and eye movements reveal something that is otherwise outside conscious awareness, and thus cannot be assessed via self-report. While other forms of testing measure what a person says would catch their attention, eye-tracking shows where people are looking. Heat maps are one type of output generated by this method and show where and how long someone looks at different parts of the material on a screen.

The efficacy and utility of eye-tracking can be increased by combining it with other methods of usability testing, such as a think-aloud procedure in which participants speak their thoughts and feelings aloud as they view and interact with the material. Depending on the software used, think-aloud audio and eye-tracking heat maps can be synchronized. What participants self-report with the think-aloud, when triangulated with where their eyes are fixating, can provide valuable insight above and beyond what each of these methods can offer alone.

Based on our review of the literature on sustainability communication, we predicted that images would disproportionately attract users’ attention, and particularly images of people engaging in efficiency behaviors. To test this, we conducted an eye-tracking study with four SMB owners while they examined a sample BER. As the heat maps below show, we found evidence to support our hypothesis: customers’ gazes were indeed more drawn to big numbers, infographics, and pictures relative to the text on the report pages (Fig. 2). This suggested that one route to increasing the appeal and effectiveness of the BER would be to optimize these visual elements.

Experimental Design (A/B Testing). While qualitative research can be a great way to generate hypotheses (e.g., the idea that people struggle to remember shutting off large devices when leaving on vacation), it is not as effective for testing hypotheses, especially ones involving complex relationships (e.g., the idea that one message frame will more effectively change a certain behavior among younger, but not older, customers). A/B testing using online samples allows such questions to be answered quickly, cheaply, and at scale. Within an experiment, participants are randomly assigned to view and respond to one of several versions of a design or product. While another option is to show participants all the options at once and ask them to pick their favorite, this measures individuals’ stated preferences. Studies show that stated preferences may or may not match actual preferences in an applied setting [12]. By randomly assigning participants to view just one version, an experimental design allows assessment of customers’ actual, or revealed preferences. If additional questions are added, it is also

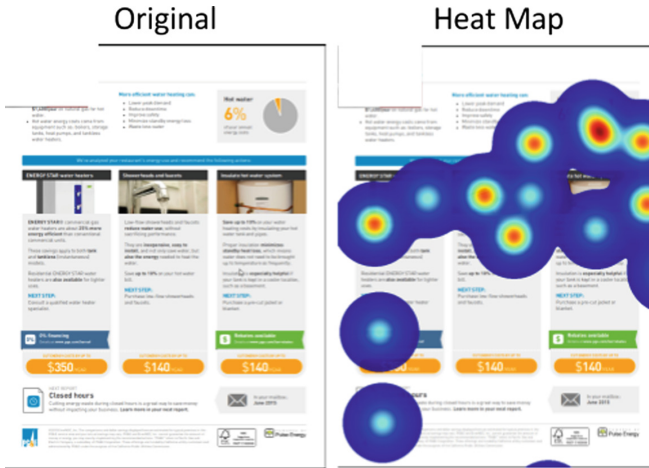


Fig. 2. Eye-tracking heatmaps on BER “tips”.

possible to look for moderating relationships among variables (e.g., do property renters and owners respond differently to an image?).

Over the course of this research, we conducted a series of experimental tests using Amazon’s Mechanical Turk (mTurk), which is an online marketplace in which people participate in short online surveys in exchange for small monetary compensation [13]. In an average study, our sample included ~15% business owners, with a mean age of ~35 years and a 50/50 gender split between male and female. Participants were randomly assigned to one of several versions of a BER image (minimum 50 participants per condition) and asked about their behavioral intention (i.e., self-reported likelihood of engaging in a proposed action) based on the materials they were shown. We also assessed the usability of the material using a scale measuring trust, satisfaction, engagement, and ease of use [14].

In a series of experimental studies based on the literature and our usability findings, we examined the hypothesis that images of people in BER tips would increase behavioral intention to engage in them. The first of these tests compared a single tip with an image of a person with a tip with an image with an inanimate object, and found that people in tip images had a highly significant effect on increasing behavioral intention, supporting our hypothesis. Our next test compared two groups of three tips: one with three images without people, and one in which one of three images had a person in it. In this second study, we found that tips without images of people might actually wash out the effect we had previously observed from a tip with an image of a person, since in this test there was no significant difference in behavioral intention. Our third test re-tested this phenomenon across a spectrum of three tip groups: one in which none of the images had people in them, one in which one of the three images had people, one in which two of the three images had people, and one in which all three images had people. The group with no images of people was rated significantly less useful than tip groups with images of people as a whole; however, this test also showed that groups with a mix of images with and without people outperformed the group

where all three images had people. Collectively, the findings of these three tests do not necessarily reveal that images of people are always the best, but they do suggest that images of people influence the effectiveness of the BER.

Beyond these findings on images, we have tested hypotheses regarding benchmarking language, tip length, rebate banners, and framing business identities, to name a few. Some findings include the fact that greater discrepancies in benchmarking are more engaging and more likely to lead to a change in behavior and that including banners with dollar savings amounts increase the usability and trust of a BER.

3.3 Apply Phase

The third and final step of our research process takes place once materials have been developed and optimized for use with real customers. While the goals of the Observe and Test phases are to, respectively, generate and optimize content with in relation to a particular variable (e.g., images in a report), this final step allows us to combine results from multiple tests of different design elements and measure how this overall design affects actual user behavior. Two methods in this phase we will discuss from our case study are field experiments and field surveys.

Field Experiments. The basics of experimental testing in the field are similar to those in an online setting, as described in the Test phase; respondents are randomly exposed to one of a number of conditions or versions of a treatment (or a non-treatment control group) with comparisons between outcomes measured. While pre-testing only allows for a comparison of reported attitudes or preferences (e.g., behavioral intention), field testing allows for a comparison of responses in an applied setting. While they are not as fast or inexpensive as online pre-testing, field experiments provide greater external validity that the product will work in practice. In one experiment, we randomly assigned BER recipients to receive one of two differently framed BER Welcome Letters and measured the percentage that returned a tear-off survey at the bottom of the letter. Half of the participants received a letter with a “corporate” tone and the other half received a letter with a “personal” tone (Fig. 3).

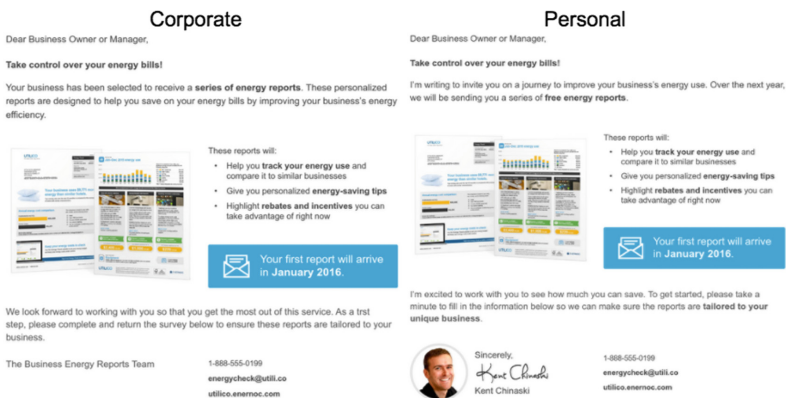


Fig. 3. Welcome letter AB test.

Businesses who received the “corporate” version of the letter returned 7.2% of the tear-off surveys, compared to 6.4% who received the “personal” version; this difference was statistically significant. While this finding was surprising (e.g., we had hypothesized the personal tone would outperform the corporate one), we identified a possible confounding variable in the first paragraph (“Your business has been selected” in the corporate vs. “I’m writing to invite you” in the personal). We reasoned that feeling “selected” might have signaled more personalization than being simply “invited” which resulted in the higher response rate. We later compared these two sentences in an online test and confirmed this hypothesis.

Field Surveys. Field surveys can capture insights from actual users of a design or product in an applied setting (vs. user testing, which is typically conducted with potential users who are recruited to participate). They can include a combination of quantitative data (e.g., “On a scale of 1–5, how satisfied are you with the Business Energy Reports mailed to your organization”) and qualitative data (e.g., “What, if any, behaviors have you engaged in as a result of receiving these reports?”). Tear-off surveys are one approach for programs like Business Energy Reports, in which the survey is attached to the report and customers are invited to complete, detach, and return the tear-off section by mail. Often, the survey also includes a web address where the survey can be submitted online instead of by mail.

Multiple field surveys from large-scale utility pilots were deployed during our two years of research with Yardi Energy. Synthesizing findings from 6 tear-off surveys across different utilities revealed both high satisfaction with and self-reported behavior change resulting from the reports. Sixty percent of respondents across surveys rated their satisfaction as either “4” or “5” on a scale of 1 to 5 where 5 indicated the highest level of satisfaction (mean = 3.7/5, st. dev. = 1.3). We also gained a much clearer sense of the actions that were being taken by report recipients. The most popular behavior was enacting a recommended tip (ranging from 18–40%), visiting an online portal (11–24%), and researching available rebates (8–22% of respondents).

4 Discussion

In order to effectively utilize the smart grid and its potential for engaging consumers to save energy, interested governments and energy utilities must identify and deploy the most effective strategies to process and present energy data to customers in a way that is clear, engaging, motivational, and trustworthy. Looking across theory and methods from the disciplines of HCI and psychology enables the benefits of both to be leveraged and synthesized to improve programs and products. In this paper, we have described a case study that walked through a multi-method framework and deployed dozens of individual research “studies” to examine and refine Business Energy Reports (BERs).

This process allows for a clear path from understanding customers to testing hypotheses and deploying in the field. Exploring theory and customer perceptions through literature review and interviews can help generate hypotheses rooted in observation. Eye-tracking and experimental pre-testing can then be used to test

hypotheses and iteratively optimize design before field implementation. Once in the field, A/B testing and tear-off surveys can be used to measure and validate insights. Collectively, these methods can work together to streamline and improve the ways humans interact with computer-generated data, and, in this case, to most effectively use that data to design energy feedback aimed at motivating behavior change.

There are several key takeaways that highlight the benefit of triangulating findings across the methods embedded in this framework. Our unbiased observation of customers and the literature allow us to generate hypotheses that relate these observations to the specific elements of BERs. After targeting businesses whose smart meter data revealed that they were most likely using their HVAC systems during closed business hours, we found that these customers were unaware (or unwilling to admit) that they were in fact using their HVAC systems then. We were only able to discover this because we kept our initial research observational by simply asking how respondents used their HVAC systems in general. This disconnect between what people do and what they think they do clearly has implications for how we might design an intervention, such that customers are first carefully convinced that they do engage in this behavior.

Likewise, literature review suggested that images and benchmarks are important and that loss frames are the best way to present motivational messaging [15]. Relating these observations to actual BER elements allowed us to develop specific hypotheses about the best way to present them (e.g., images of people and loss framing will increase the behavioral intention of BER recipients). Methods used in the Testing phase led to findings that at times agreed with the existing literature and at other times deviated from it. For example, consistent with the literature, we found that it is as important to consider BER imagery as BER text and that the language and construction of the benchmarks impacts readers differently (e.g., business owners prefer social benchmarks to goal-based benchmarks, and larger discrepancies in benchmarking more effectively engage customers). In contrast, we also found that gain frames sometimes outperformed loss frames (contrary to many assumptions in behavioral science). These insights allowed us to revise and optimize BER content. Lastly, field testing in the Apply phase allowed us to measure the impact of these reports in their entirety (as opposed to refining just their individual components) on actual customers, as they engage with them.

While we reviewed this process and these methods in a very specific context – Business Energy Reports designed to inspire greater energy efficient behaviors – the framework and methods we have discussed can certainly be adapted and applied in a number of situations in which humans are interacting with computer-generated data.

References

1. Karlin, B., Ford, R., Zinger, J.: The effects of feedback on energy conservation: a meta-analysis. *Psychol. Bull.* **141**(6), 1205–1227 (2015)
2. Froehlich, J., Findlater, L., Landay, J.: The design of eco-feedback technology. In: *Proceedings of the 28th International Conference on Human Factors in Computing Systems - CHI 2010* (2010)

3. Bandura, A.: Principles of Behavior Modification. Hold, Rinehart & Winston, New York (1969)
4. Bridgeman, B.: Effects of test score feedback on immediately subsequent test performance. *J. Educ. Psychol.* **66**(1), 62–66 (1974)
5. Becoña, E., Vázquez, F.L.: Effectiveness of personalized written feedback through a mail intervention for smoking cessation: a randomized-controlled trial in Spanish smokers. *J. Consult. Clin. Psychol.* **69**(1), 33–40 (2001)
6. Chopra, A.: Modeling a green energy challenge after a blue button. White House Office of Science and Technology Policy (2011). <http://www.whitehouse.gov/blog/2011/09/15/modeling-green-energy-challenge-after-blue-button>
7. Karlin, B., Ford, R., Squiers, C.: Energy feedback technology: A review and taxonomy of products and platforms. *Energy Eff.* **7**(3), 377–399 (2014)
8. A Design Thinking Process. ME 113. N.p. (2012). Web, 01 Mar 2017
9. Smith, N., Joffe, H.: How the public engages with global warming: a social representations approach. *Publ. Underst. Sci.* **22**(1), 16–32 (2012)
10. Fischer, C.: Feedback on household electricity consumption: a tool for saving energy? *Energy Eff.* **1**(1), 79–104 (2008)
11. O’Neill, S.J., Boykoff, M., Niemeyer, S., Day, S.A.: On the use of imagery for climate change engagement. *Glob. Environ. Change* **23**(2), 413–421 (2013)
12. Nolan, J.M., Schultz, P.W., Cialdini, R.B., Goldstein, N.J., Griskevicius, V.: Normative social influence is underdetected. *Pers. Soc. Psychol. Bull.* **34**(7), 913–923 (2008). doi:10.1177/0146167208316691
13. Karlin, B., Ford, R.: The usability perception scale (UPscale): a measure for evaluating feedback displays. In: Marcus, A. (ed.) Proceedings of the 2013 Human Computer Interaction (HCI) Conference. Springer, Heidelberg (2013)
14. Kahneman, D., Tversky, A.: Prospect theory: an analysis of decision under risk. *Econometrica* **47**(2), 263 (1979)