ORIGINAL ARTICLE

Watts your usage? A field study of householders' literacy for residential electricity data

Melanie R. Herrmann · Duncan P. Brumby · Tadj Oreszczyn

Abstract Smart metering studies typically focus on quantifying behavior change. However, little is known about how users understand energy information and analyze and interpret feedback from energy data visualizations. To investigate this, we gave 13 participants from nine UK households an electricity power clamp meter. Prior to installing and using the device, we conducted interviews with participants to gauge their understanding of their home electricity consumption and found that participants varied considerably from limited to substantial energy literacy. Two weeks after the clamp meter had been installed, we conducted a contextual inquiry in which we asked participants to explain the web-based time series visualization of their recorded electricity data. We found that the visualization proved unfit: participants relied on memories and suggested likely routines, while widely being unable to reliably identify specific events in the data visualization. In follow-up interviews 3 months later, we found that participants' understanding of their home electricity consumption had hardly changed. Finally, we invited participants to generate ideas how smart electricity feedback could be optimized. They named different forms of disaggregation, higher temporal resolution, and interactivity as design requirements. In summary, these results suggest that people find home energy data very difficult to understand and link to everyday actions and behaviors.

Keywords Electricity feedback · Behavior change · Energy literacy · Data visualization · Disaggregation

M. R. Herrmann (✉) · D. P. Brumby
Interaction Centre, Department of Computer Science, Faculty of Engineering, University College London, 66-72 Gower Street, London WC1E b 6BT, UK
e-mail: melanie.herrmann.14@ucl.ac.uk

T. Oreszczyn
The Bartlett School of Environment, Energy and Resources, Energy Institute, Faculty of the Built Environment, University College London, 14 Upper Woburn Place, London WC1H 0NN, UK

Introduction

To meet the many challenges of climate change, a general and widespread transition to a low-carbon economy will be necessary. Many countries are pursing policies to meet this goal. For example, the UK Climate Change Act 2008 sets a target for an 80% cut in greenhouse gas emissions by 2050 (http://www.legislation.gov.uk). The UK government has also committed to rolling out smart meters to all homes and small businesses by 2020 (http://www.gov.uk). Germany has taken similar measures and will have a large-scale roll out of intelligent metering systems by 2020, with the target of having an 80% adoption rate by this time. Several other countries have also commenced with similar large-scale roll outs of domestic smart meters (e.g., Australia, Canada, Denmark, Finland, France, Iran, Ireland, Italy, Japan, Malta, Netherlands, New Zealand, Norway, Spain, Sweden, UK, and US). One of the core features of a smart meter is

to offer automated meter reading for billing. This replaces conventional meters by allowing consumption data to be automatically sent to utility companies. Many smart meters also communicate with in-home displays (IHDs). This offers a tangible benefit to the end-user, enabling them to more closely monitor their energy consumption. In addition to smart meters, there are now also other devices on the market, such as electricity power clamps, that come with mobile apps or websites that feedback energy usage information. All of these modern smart metering tools and devices share a common assumption: that by enabling the monitoring of domestic energy consumption, people will better understand their consumption and so learn to manage their usage better, save money, and reduce emissions.

Previous research has shown that eco-feedback interventions can lead to changes in behavior, which in turn result in domestic energy savings (Darby 2006; Fischer 2008; Ehrhardt-Martinez 2010; Froehlich 2011). However, across the countries that have introduced smart meter feedback so far, questions have been raised as to how efficient the smart metering feedback is in helping householders reduce their energy consumption. Despite the potential for eco-feedback interventions to help with domestic energy savings, there are a number of challenges and limitations with the current generation of systems (Wallenborn et al. 2011; Pereira et al. 2013). Here, we focus on the user interface. These typically visualize data showing consumption over time. However, most people are not trained in reading complex time series graphs and might therefore find it difficult to extract useful information and insights from them (Galesic and Garcia-Retamero 2011). Also, a large share of the population is thought to be energy illiterate (i.e., they are not familiar with the content of the feedback) (DeWaters and Powers 2011; Brewer 2013). While numerous studies have investigated the effect of eco-feedback on behavior change, it is rare that they explicitly investigate data comprehension, that is, the cognitive sense-making processes (Yun et al. 2010; Fischer et al. 2016). It remains unclear what kind of information is needed to provide an actionable basis for behavior change.

In this paper, we investigate how people make sense of their domestic electricity consumption data in the form of time series data visualizations. We first interviewed participants to see if they know how much electricity they used for carrying out everyday actions. As part of this interview, we had participants sketch a graph of what they thought their electricity consumption looked like on a typical day. Second, we provided households with an electricity power clamp meter, the Loop energy saving kit, and recorded their electricity usage. We conducted a contextual inquiry using the think-out-loud method (Lewis and Rieman 1993), where we asked participants to make sense of the eco-feedback provided by the Loop. Third, in order to examine long-term engagement and learning over time, we interviewed our sample in a follow-up study 3 months after the first interview. Fourth, we derived further user requirements by interviewing our participants to assess which eco-feedback characteristics they think would be crucial for maximizing comprehension and learning and how these could be visualized.

Before describing our study's method and results, we are first going to review important related research. For our study, the three most relevant fields of work are the following: behavior change, energy literacy, and graphical literacy.

Related work

Behavioral frameworks

Recent advancements in sensing and tracking technology have brought about opportunities to develop behavior change systems in the field of human-computer interaction (HCI) (Hekler et al. 2013). Hekler et al. advocate for grounding empirical research in theory and also for researchers to refine the theory if necessary based on empirical findings. Behavioral models that describe the process of change are typically conceptual frameworks. They comprise several constructs, such as the fundamental building blocks of a behavioral theory; they are valuable for research because they provide more specific guidance to the design and implementation of behavior change technologies, helping to guide the evaluation process (Hekler et al. 2013).

A prevalent behavior change model is the transtheoretical model (Prochaska et al. 1994). It describes the stages of precontemplation, contemplation, preparation, action, maintenance, and termination. However, this model dates back to a time before ubiquitous computing, smart houses, and

big data. It therefore does not address the interaction with digital tools and does not describe possible comprehension or engagement problems with technology. Previous work found that users do not engage sufficiently with energy monitoring technologies and therefore miss out on the potential that these technologies offer for energy savings (Costanza et al. 2012; Yang et al. 2014). While numerous studies have looked into the effects of eco-feedback on behavior change, it is rare that they explicitly investigate data comprehension, that is, the cognitive sense-making processes (Yun et al. 2010; Fischer et al. 2016).

The stage-based model of personal informatics systems proposed by Li et al. (2010) includes the stages of preparation, collection, integration, reflection, and action. The model is useful as it explains possible barriers in proceeding from one stage to a subsequent stage. It explicitly addresses the collection of data, as well as the user integrating and reflecting on the collected information, which is crucial for the design of behavior change technology. In the integration phase, information is prepared, combined, and transformed (e.g., users have collected the data with pen and paper they need to gather and organize all their notes and transcribe them into a tech-based application). Alternatively, the organization of data that is digital to begin with can be automated by a system. Reflection involves looking at and exploring the information, possibly interacting with visualization. In this stage, users might struggle with understanding the data.

Epstein et al. (2015) expand the stage-based model of personal informatics and propose the lived informatics model of personal informatics. They point out that previous approaches focus too much on behavior change goals and so do not adequately capture the nature of tracking in real life. People might have very different goals for using personal informatics tools. For instance, some people track out of curiosity, without wishing to make changes. Still, reflection and comprehension are central. Furthermore, the model addresses the selection of a tool, the interaction with the tool, and its role in information process.

Both the stage-based model of personal informatics systems (Epstein et al. 2015) and the lived informatics model of personal informatics (Li et al. 2010) are relevant to our research in understanding how eco-feedback technologies can shape users' behavior. We argue that it

is important to further investigate how users make sense of data when interacting with a tool and how digital information visualization can help facilitate learning and understanding about one's own behavior.

Energy literacy

One of the major challenges in eco-feedback is that the majority of the population has poor energy literacy (DeWaters and Powers 2011; Brewer 2013; Rego Teixeira 2014; Bager 2014). DeWaters et al. (2007) claim that "an energy literate person needs to have a basic comprehension of fundamental scientific principles," "a sound knowledge base," and an "understanding of energy concepts." According to DeWaters and Powers (2011), energy literacy encompasses behavioral as well as cognitive (content knowledge) and affective (attitudes, values) characteristics, and "will empower people to make appropriate energy-related choices and embrace changes in the way we harness and consume energy." Brewer's (2013) definition corresponds with the cognitive aspect described by DeWaters and Powers (content knowledge): After Brewer, energy literacy is "the understanding of energy concepts necessary to make informed decisions on energy use at both individual and societal levels." The author gives examples such as knowledge about electricity generation or the awareness that certain appliances use much more power than others, but that some low-consumption appliances (such as the fridge) consume more over time because they are constantly on.

It is disputable if users need to understand complex scientific principles or to know how electricity is generated in order to be economical in their household. Instead, we think a more practical definition is needed to help the design of user-friendly solutions that are meant to increase energy literacy, thus shaping energy-saving behavior. We focus here on the cognitive characteristics: if users do not understand the information, they are not able to change even if they wanted to. This leaves us with DeWaters et al. (2007) question of "how much and what types of knowledge and understanding are critical for empowerment?"

Most importantly when providing the user with information, feedback should prioritize "practical" knowledge over "technical" knowledge (Hofman 1980). Álvarez and Vega (2009) and Ellegård and Palm (2011) point out that interventions need to provide the practical mapping of information to everyday

activities to allow users to integrate it into their knowledge structures (i.e., their mental models of energy usage). Gabe-Thomas et al. (2016) have investigated householders' mental models of domestic energy consumption using card-sorting technique. They found few shared ideas about which appliances belong together. Where clusters emerged across participants, they were based on activities and locations in the home, but not on actual energy consumption. They conclude that energy-related communication should be framed to practices within the home and feedback must be tailored to these preexisting perceptions in order to be comprehensible and actionable. Chisik (2011), too, explored users' mental models of electricity using a sketching method. Exploring people's mental models of electricity, electrical infrastructure, and usage per appliance, he found that they did not have clear concepts of electricity and deduced consumption rates from appliances' size and usage duration. We conclude that we need to focus on people's knowledge on how much energy they consume for everyday activities.

Graphical literacy

Residential energy feedback systems normally output the recorded data in graphical format. The Smart Metering Equipment Technical Specifications (SMETS) state that the "IHD shall be designed to enable the information displayed on it to be easily accessed and presented in a form that is clear and easy to understand" (Department of Energy and Climate Change DECC 2012). However, concerns have been raised about how clear and easy to understand energy data is (Roberts and Baker 2003). While visualizations tend to represent energy data as time series either in line or bar charts, people may not necessarily be good at reading graphs (Galesic and Garcia-Retamero 2011). In fact, Boy et al. (2014) have argued that very few studies have investigated how well people can extract information from a graphical representation, such as a line graph or bar chart.

Graphical perception and literacy are defined as "the visual decoding of information encoded on graphs" (Cleveland and McGill 1984). The decoding process requires considerable cognitive effort (Costanza et al. 2012). According to Murugesan et al. (2014), "Visualization of energy consumption is widely considered as an important means to assist the end-users and the energy managers in reducing energy consumption

and bringing about sustainable behavior. However, there are no clear design requirements to develop the energy monitoring visualization."

Graphical representations communicate information often better than textual or numeric representations (Fry 1981; Larkin and Simon 1987), and literature on graphic literacy and information visualization suggests that representations enhance thinking (Fry 1981; Kirsh 2010; Munzner 2014). If not chosen wisely, tough, comprehension is severely constrained (Tufte 1985; Baur et al. 2012; Tong et al. 2015). A principle in graphical design is that the representation should match the physical properties of what they present and the representational structure should preserve the conceptual structure of people's mental model (Pinker 1990; Zhang and Norman 1994; Cheng 2011; Cheng 2014). But if that is not the case, the representation becomes less legible and meaningful to the user (Baur et al. 2012; Tong et al. 2015; Galesic and Garcia-Retamero 2011).

Purpose of current study

We have identified a number of aspects from the literature that require further investigation. First, we want to look closer at the sense-making process in interpreting and learning from residential energy data. Second, we think energy literacy needs to be investigated in terms of relevant use cases (i.e., how much people know about the consumption of everyday activities). Third, we want to investigate how comprehensible and useful typical time series representations of domestic energy usage data are. Our research question is whether people understand how everyday actions map to the presented visualization and if they can identify appliances and events in the data pattern.

Method

Sample

To obtain our sample, we advertised our study by posts on Twitter and Facebook, by leaflets in the university building and by word of mouth, aiming for a sample that would cover a variety of housing and occupant types. We recruited 13 participants (six female) from nine households. Eleven out of 13 participants filled in our demographic online

questionnaire. Mean age of the sample was 40 years (SD = 15, range 25–76). Education varied from less than high school to doctoral degree. Five of the participants lived in terraced houses, three in apartments, and two in semi-detached houses. Most of the participants lived in rented accommodation (ten of the 13 participants). None of the subjects lived alone. Two households had used a smart meter before in previous residences but none had a smart meter in their current property. None of the households were on a "time of use" tariff (where electricity costs different prices at different times of the day).

Materials

We provided each household in the study with an electricity-monitoring device: the Loop energy saving kit (available from UK-based technology company Navetas, *See* https://www.your-loop.com). The Loop energy saving kit consists of three hardware items. First, there is a current clamp transmitter that measures the household's electricity consumption. The transmitter must be clipped to an electrical conductor (i.e., one of the electricity meter's cables). It transmits measurements to a receiver by radio transmission. Second, the receiver, which is plugged into the household's Internet router, communicates the household's electricity consumption data back to the server so the occupant can see the collected information online. Third, a power plug that powers the receiver. The customer is guided through the installation process on the your-loop.com website.

A key feature of the Loop energy saving kit for the purposes of this study is that it has a web portal that allows users to look at their household's electricity consumption (available at: www.your-loop.com, *See* Fig. 1). Using the web portal, householders can see their recorded electricity usage graph on a daily, weekly, or monthly basis, in a range of unit options (financial cost as estimate in GBP, energy use as estimate in kilowatt-hours, or environmental impact as estimate in CO_2 emission). We chose the Loop for our study because it is representative of the tools currently available on the market. It is very affordable and easy to install, and it visualizes domestic electricity consumption as time series data.

Methodology

We conducted a field study to investigate how users interpret and learn from home electricity data. Interviews took place partly face-to-face, partly over Skype. The study comprised four parts.

First, we explored participants' everyday knowledge about electricity usage in their home by asking them a couple of questions (*See* "Procedure"). In addition to the verbal responses, we asked them to make a sketch of their electricity consumption over a day. Participants were free to choose the type of graphic and metrics they wanted, and we stressed that drawing skills were not important. Sketching is a method typically used in HCI to inform interface design (Buxton 2007). Like Chisik (2011), we chose this approach within a user study because sketches are a rapid, accessible, and expressive method which reveals the mental model of the subject: visuals and thinking are closely linked and sketches support the thinking process because they are an externalization of internal thought (Tversky et al. 2008; Kirsh 2010; Greenberg et al. 2012; Walny et al. 2015).

Second, we conducted a semi-structured contextual inquiry (Holtzblatt and Jones 1993), using the think-out-loud method (Lewis and Rieman 1993) while looking at the recorded electricity data on the Loop website. Our aim was to determine whether people understand peak periods of electricity usage and explain what was contributing towards this, and also if they could deduce how they might go about reducing energy usage on the basis of this data. The think-out-loud method in the contextual inquiry is a useful tool to investigate how people understand the data. We asked users to actually use the Loop feedback with the task to make sense of the data. Asking them to think-out-loud at the same time grants insights into what they are thinking, what they are trying to do, what questions come up as they explore the data, and how they read and interpret what they see. The semi-structured method offers the required flexibility, as every user might go about approaching the data in their own ways, while still providing enough structure to analyze the qualitative data in the context of the task and compare across the sample. In those interviews that were conducted on Skype, participants would share their screen with the interviewer so we were able to see the website.

Third, we conducted a follow-up interview with ten of the 13 participants 3 months after the initial interview in order to examine long-term engagement (Lazar et al.

2015). The aim was to find out if participants were still using the Loop electricity-monitoring tool and if they had improved their understanding of their domestic electricity usage.

Fourth, we wanted to derive further user requirements. We therefore asked participants which characteristics in electricity metering feedback would be crucial for maximizing comprehension and learning and how these could be visualized. We stressed that they were free to use their imagination.

All interviews were transcribed in the transcription software f5. The transcripts were then imported into the qualitative data analysis software Nvivo and analyzed thematically (Aronson 1994; Braun and Clarke 2006).

Procedure

We sent the Loop energy saving kit to the participating households, including the instruction to set the device up but not log in to the website until after we conducted the interview. All participants later confirmed that they had complied with this instruction. Two weeks after the participants had installed the Loop, we conducted the first of two interviews. These interviews took place throughout March and April 2015. The second follow-up interview took place 3 months later (throughout June and July 2015). Three participants choose not to take part in the follow-up interview, reducing the sample size to ten.

In the first part of the first interview, we asked participants the following three open-ended questions:

1. Which electrical devices in your household do you believe consume most electricity?
2. In a metric of your choice, can you please estimate how much electricity your household appliances consume?
3. How do you think your electricity consumption look over a day? Please, make a sketch of your electricity consumption over a day.

In the second part of the first interview, we would ask participants to log in to their Loop account. The task in the contextual inquiry was to verbalize what information they see in the graph (Fig. 1) and to explain which appliances or activities have led to the displayed patterns. Our semi-structured interview guide contained a number of questions that we used to nudge participants if they stopped thinking out loud:

1. Please think-out-loud.
2. What do you see?
3. Can you please interpret the graph you see?
4. Can you tell me what the peaks are?
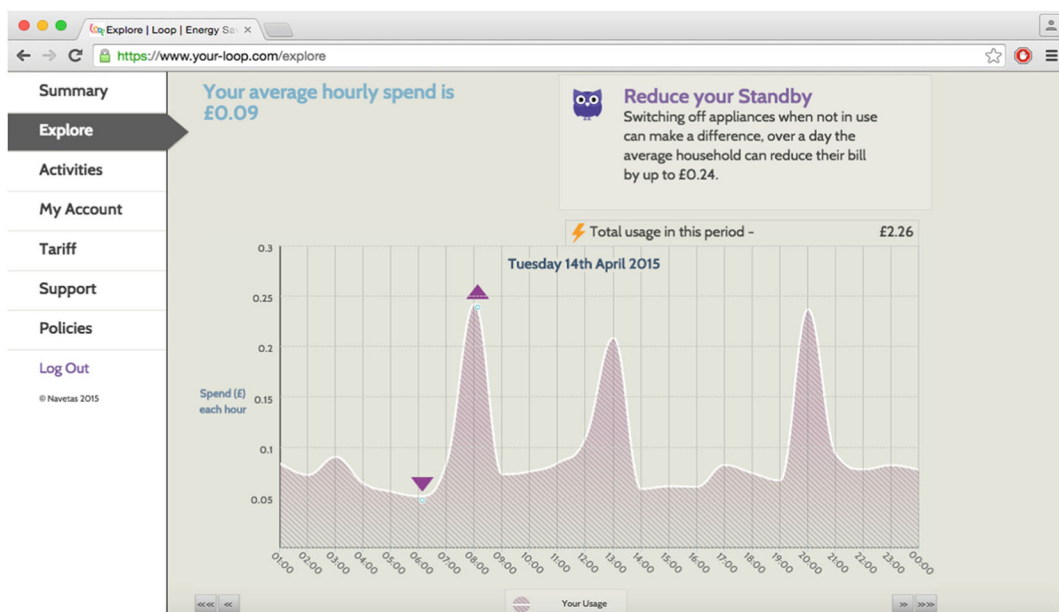5. Can you identify what you did in this moment?

Fig. 1 The historic data feedback view of the Loop

6. Can you identify appliances in your graph?

In the first part of the second interview 3 months later, we asked them the following questions:

1. Since our first interview, have you logged in to the Loop website again?
2. If no,

   a. Can you say why you haven't looked at the data again?
   b. What could have motivated you to look at the data again?

3. If yes,

   a. How often did you look?
   b. When did you look? Please describe the situation.
   c. Why did you look?
   d. What did you learn?

      i. How did you learn it?
   e. Have you gained a better understanding of your data?

      i. If not, why do you think that is?
      ii. If yes, what do you understand better now?

In the second part of the second interview, we instructed participants to imagine that they were involved in designing the "perfect" data feedback. We stressed that they were welcome to use their imagination without considering technical feasibility.

We asked them the following questions:

1. What would your smart meter be like?
2. Which functions would be important?
3. How would it help people to learn?

Results

The following section presents the findings sorted by the four parts of our field study. To help with the overview of the interview data, Table 1 lists the $N = 13$ participants (P1–P13) from the nine participating households (H1–H9).

A priori energy literacy

This section reports the findings from the responses that were given by participants before looking at the recorded data in the first part of the first interview. They gave a variety of answers to the first two open-ended interview questions about which household appliances consume the most electricity and how much. Responses varied from the washing machine (mentioned four times), the fridge (mentioned three times), the shower and the oven (mentioned two times, respectively), the tumble dryer, leaving the lights running or devices plugged in, an electric fireplace, the TV, the computer, and kettle (mentioned by only one participant each).

Not only did participants differ in their opinion as to which household devices used the most electricity, participants also reported low confidence in their responses. Only P12 and P13 (H9) were confident that their electric shower consumes the most electricity. More than half of the sample was unable to quantify electricity usage in any way. For example:

Table 1 Households 1–9, showing that P1 and P2 live together in H1. P3, P4, and P5 live together in H2. P12 and P13 live together in H9. None of the participants lives alone but not all household members were interviewed (i.e., participants) in the study. The household and participant numbering was assigned in chronological order to the households signing up to our study. P4*, P5*, and P10* dropped out after the first interview study and did not take part in the follow-up

H1: P1, P2
H2: P3, P4*, P5*
H3: P6
H4: P7
H5: P8
H6: P9
H7: P10*
H8: P11
H9: P12, P13

P1: *I know what is kilowatts and watts, but that doesn't mean anything to me.*

P8 considered a monetary measure to gauge consumption: *I guess that it [oven] might cost 2 Pounds an hour?*

P11 said the *only measure that I can say to gage is the amount if time that is on. So for measuring, I measure by ... you know, so ... the devices that are on the most.*

To quantify how much appliances draw when on, three participants spontaneously used watts or kilowatts as a unit of power. P3 replied by saying his computer screen uses between 15 and 20 W, the Internet router 7 W, the electric shower 1500 W, and a lamp in the kitchen with two bulbs 60 W. P13 guessed 100 W for the baseline consumption, 800 W for elevated baseline with lights, computer and the like being switched on, and 8 kW for maximum consumption. P9 remembered that the baseline consumption in his old flat was around 300 W and therefore guessed it should be around 500 W in his new place because it is bigger.

We used sketching to reflect users' mental ideas of how much electricity they used over a typical day. All participants opted for a solution with the timeline on the horizontal axis. P12 tried to use the image of a clock in the first place, but realizing that a 12-h clock "*does not work for a 24 h day,*" she concluded that is "*probably easier than to use a graph.*"

The sketches varied greatly in sophistication and detail. P3, P8, and P13 drew staircase-shaped graphs with square waveforms; P1 drew triangle waveforms while the remaining nine participants chose smooth line graphs with sine waveform. The labeling of the *x*-axis ranged from equidistant time steps in numerical scaling to semantic anchors such as "*morning, midday, afternoon,* and *evening,*" coinciding with participants daily routines. For the *y*-axis, P1, P6, and P11 did not use labels whatsoever. P2 noted down "*consumption*"; P7 "*more elec[tricity]*" by the axis. P 3, P4, and P5 chose "*kWhrs*" for their shared sketch; P12 opted for kW. P8,

P9, and P13 added numeric values to their kW-scales. Figure 2 shows three representative examples of the sketches that are further described in the following paragraph.

Figure 2a by P11 shows a smooth sine waveform-like graph from midnight to midnight. It indicates when the household members are getting up and ready for the day and when they return home in the afternoon and evening. When being asked about the height of the consumption depicted in her graph, P1, who drew Fig. 2b, explained: "*that's not so much a measure of how much, but more kind of going up when we are here.*" Figure 2c by P13 is most sophisticated, with square waveform and labels on both *x*- and *y*-axes. P13 depicts clean-cut-on and offsets of device usage in his staircase-shaped graph, he quantifies how much electricity is consumed by the distinct activities, and he annotates the graph to explain the peaks in the data pattern.

Participants' comments while sketching revealed that what they drew was in accordance with their daily routines. The leading motive in participants' narrative is the time of day and the practice that is typically performed at that time. The curve would rise when getting up in the morning, those with electric water heating might mention taking showers, and others making breakfast, including boiling the kettle, toasting bread, and making coffee. The curves would drop where all household members had to go to work or children to school. The curve would then rise again in the afternoon when children came home from school, adults returning from work, taking up evening routines such as cooking, charging the phone, and watching TV. P6 assumed that "*at the weekend it's probably high all the time*" since the family was at home. Daily and regular routines would be mentioned more often than less regularly occurring activities such as washing laundry.

Contextual inquiry

In the second part of the first interview, we would ask the participants to log in to their Loop account and to think-out-loud (i.e., to verbalize what information they see in the graph and to explain which appliances or activities have led to the displayed

Fig. 2 Participants' sketches of daily electricity consumption: **a**, **b**, and **c**. **a** Timeline with sine waveform by P11. **b** Timeline with triangle waveform by P1. **c** Timeline with square waveform and scaling by P13
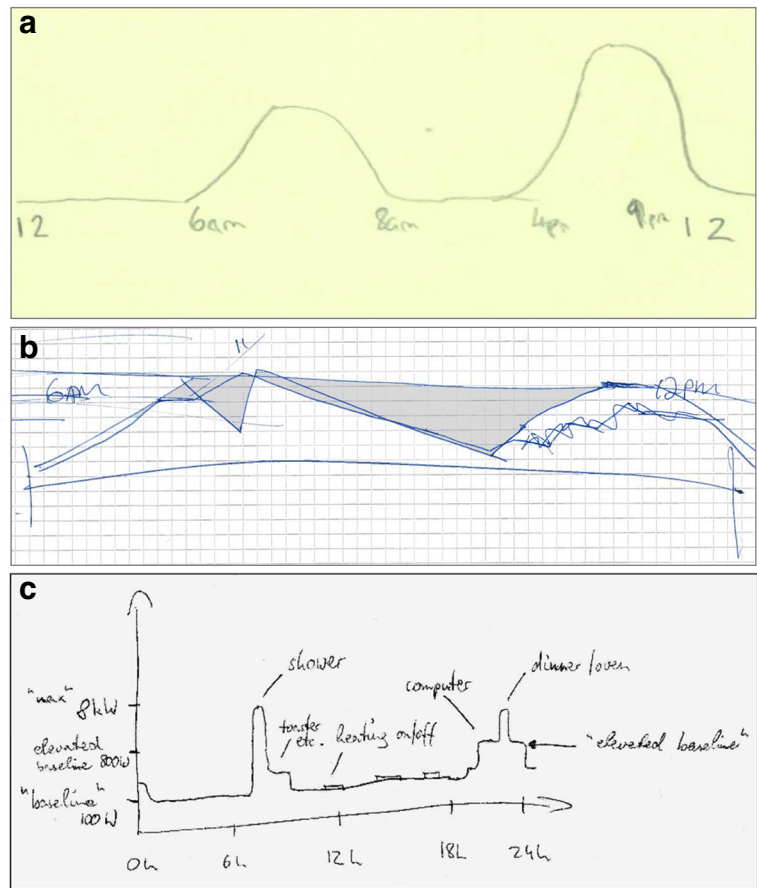
patterns). In the following, the results for the emerging themes are presented.

Routine-based and memory-based reasoning

Our first key finding is that participants had a harder time to account for peaks that were caused by less habitually performed actions—such as washing laundry or vacuum cleaning. Similar to the approach when drawing their sketches in the first interview, participants would focus on explanations that related to typical routines that they perform every day.

> P1: *it kind of goes up and down throughout the day depending on when we are up and about. It goes up a little bit in the morning when we are … eh … we are all kind of here and doing a few things and the electricity comes on and then it goes down again. And then … it dips when we all leave.*

Daily or weekly habits served as a basis for generating ideas about what might have caused the data pattern.

> P1: *Friday lowest period. Everyone out for a drink in the evening?*

Our second key finding is that participants relied on their memory. They would draw from their knowledge of what they had done recently to interpret the displayed pattern. Therefore, as their memory of events or activities faded with time, the more historical the data, the less confident the participants were about their interpretations. Two participants checked their calendars during the interview to help make sense of the data pattern. P8 could not explain a peak and with the backup of his digital calendar, he then described how he had people visiting that day; he was then able to identify the point of

their arrival and when he would be cooking roast chicken for the guests. Likewise, P7 checked her pocket calendar to find out whether she was working from home or at the office on a particular day.

Interpretation errors

Interpretation errors are the third key finding. As a consequence of the routine-based reasoning, people would commit errors such as assigning peaks around lunchtime to cooking: P8 inferred that his cooking caused a 2-kW peak around lunchtime, until he realized a bit later that he had a washing machine running at the same time. P6 tried to understand a big peak in 1 day when she was at work. First, she considered "*(…) maybe my husband was at home yesterday. I don't know (…) He would have his computer plugged in and be using it. And might also charge his phone cause he doesn't have a routine to charge it. He would be making himself tea and coffee and … stuff.*" However, the peak was higher and longer than computer and phone would account for, so the interviewer asked what other appliances could have been running in the period of question, which prompted P6 to realize "*Ooh, so that's what that could be as well. Oh, because Tuesday is the day that the cleaner comes. And she comes between 9 and 12. So that's who was at home and she is doing the washing.*"

P3, P4, and P5 (who live together and were interviewed together), spend some time discussion a reoccurring high peak in the morning hours while going through the data recording of different days, until they finally figured out that the electric shower was causing the peak:

> *P4: What is that peak?!*
> *P5: That is the washing machine.*
> *P4: That's 8 o'clock this morning. That is my hair dryer.*
> *P5: Ah.*
> *P4: Oh my god, I am not gonna be allowed to dry my hair again (P4 and P5 laugh)(…)*
> *P3: That's 7, there, that peak.*
> *P5: What were you doing?*
> *P4: Drying my hair.*
> *P5: That was your hairdryer?*
> *P4: Wasn't for very long, though (laughing)(…)*
> *P5: Did you wash your hair? Where you late or something?*
> *P4: No, I do it every other day.*

> *P5: Your hairdryer doesn't show up.*
> *P4: No, I didn't use it, I don't do it every day.*
> *P5: Oh, I see.(…)*
> *P3: I got up and had a shower. Maybe my showers are quite hot.*
> *P5: Maybe your showers are quite long.*
> *P4: Your shower is higher than my hairdryer.*
> *P5: It's amazing it uses that much. Just for a shower!*
> *P4: But … I use the shower before I dry my hair … so it's the shower that's using an awful lot!*

Unaccountable patterns

Several times, participants tried to recall what they did on a given day when explaining the peaks but concluded "*I cannot remember,*" staying unclear about what had caused the data pattern. Generally, the longer the day of question dated back, the less confident participants were about their interpretations of the graph's ups and downs. Other than that, there were periods in which the interviewee had not been home and therefore could not have memories (as in the earlier example when P6 puzzled over what her husband could have been doing).

P3, P4, and P5 were surprised that their consumption was higher on Thursday and Friday than on Saturday and Sunday of the same week, although there was one person more in the house during the weekend. Similarly, P9 struggled to comprehend how the usage in the afternoon could be marginally lower than in the middle of the night. P6 could not explain to herself why there would be small increases in usage during the night and why the graph would start rising between 6 and 7 am, when none of the household members gets up before 7 am. P13 did not come to a conclusive explanation of what was going on during one of his Saturday evenings. First, he reasonable argued that his computer's graphics card must be a significant contribution to the evening consumption because he had been playing a computer game. Then, however, he remembered that they had cooked that night and, thinking more about it, decided that the usage from the evening in question was just the elevated baseline and the computer would not be that high after all.

A design factor of the visualization that concealed distinct events in the data was the low temporal resolution of the graph (for example the use of the electric shower and the hair dryer in the morning, which would

blend into one peak in the graph and prevent the users from detecting the two events). P9 articulated criticism, thinking the tool is "*mostly a gimmick*" and leaving them "*frustrated. Not frustrated, too strong a word. But I would like more granular control.*"

Disaggregation

Some participants were able to identify big events such as the electric shower or the washing machine and the tumble dryer. Overall, however, they performed rather poorly at mapping the data patterns to their everyday actions. They lacked insights about which actions were contributing towards high or low usage days.

> P7 joked: *It says what was your lowest day and then ... what did you do differently that day? I suppose I did go to work (laughing).*

P10 could not make sense of his high evening peak until the interviewer pointed to the fact that the graph displayed global usage and that one peak could consist of several appliances running at the same time. He then listed several devices that would most likely contribute to his big peak around bedtime (he switches on his dish washer and his washer dryer before going to sleep).

P13 concluded that the lack of mapping between data and everyday actions made it difficult to proceed to the stage of changing behavior: "*If I wanted to ehm, actively consume less energy- yes I think it would be helpful to know how much I am actually using. Ehm, it's a little bit like ... obviously I know if I turn off the light, I know I use less energy. But it's a little bit like putting things to scale, how much difference does one light bulb compared to the fridge make? That sort of thing would be interesting to know and I think without that sort of software it might be difficult.*" He added he would be able to work out this information over time.

By contrast, other participants repeatedly asked for breakdowns of the information on the appliance level and graphs of activity: P2 tried to figure out which appliances were contributing to a big peak and was looking for a breakdown on the device level on the website, in vain. P3 stated that the software disappointed him, because he thinks it should do the work for the user to "*look at patterns and classify usage of different things.*" He reported having used other personal informatics systems before, so he would have expected

graphs of activity. Thereupon, P5 added to the discussion with P3 that she could have taken notes: "*You know, when I had the washing machine on and the cooker. Which would have been helpful. Cause these are ... big things. I would have made a note, had I known it might be more useful.*" Likewise, P12 said that it is "*a bit like a food diary. It would be a lot of work. Because you have to recall what you were doing.*"

The rest of the sample, when asked if and what they learned from the Loop and if they have ideas how they could change their behavior to save electricity, would not give a concrete answer. They proposed universal ideas that were not related to what they had seen in their graphs before. They said they would need to look further into it. One person suggested to look at the live-spend widget while another household member is having a shower to determine how much the shower consumes or to look at the graph and read the activity level at times of the day when a certain routine such as making cups of tea is performed.

Long-term usage

In the follow-up interviews after 3 months, there were three participants (P1, P2, and P7) who had never looked at the data again after the first interview. P1 found that the software did not give her anything tangible. P2 did not see any benefit and P7 said she had been interested but had completely forgotten about it. They would have wished for better explanations, for information that was easy to grasp and for a breakdown of the global usage.

Four participants (P3, P6, P9, P12) had looked at their data again in the first weeks after the initial interview but had quit by the time of the second interview, because they did not gain any further understanding over time.

> P3 said: *We were like "oh okay I am here for the weekend" (laughing). So. It wasn't gaining any insight into electricity usage as such.*

P6 had looked again when being prompted by her children who wanted to know how they were doing. She described her insights as a "*one-off thing.*" She explains how in the first interview she had first been puzzled about her data: "*I thought why on earth is Tuesday the biggest day?*" She then thought about it and realized her cleaner was in on Tuesday and she concludes that "*You get that understanding the first*

time you look at your data and think about it. And I don't think it becomes any more nuanced over time."

P9, who had only logged in again to "*kill time*" further stated that "*I find that the information it offers is too general to really give me any inspiration to log back in and continue using it.*" The software does not offer the services he is really looking for, namely information on "which particular appliances or what particular events might have used more energy."

P12 stated that by using the comparison function of the Loop (that compares the household to other households), she had learned that "*we actually have a lower resting energy consumption than I thought. We thought our fridge is really bad, but it's actually not that bad.*" She also noted that in a "*more complex household, where you had a lot of appliances, it would be difficult to gage what's doing what.*"

Finally, P8, P11, and P13 were still monitoring their electricity usage with the Loop, where P13 read the weekly digest emails and did not log in to the website anymore. P8 and P11 stated they had been looking every 2 to 3 days at times. The consensus was weekly, especially for reading the email digest, with a tendency to check more often when using the app. P13 had unsuccessfully tried to scrape the website for the full dataset.

With these participants, we inquired if they had gained understanding of their data patterns over time and if they had learned more about their household's electricity usage.

P8 had learned to identify when his son is home from school but was not quite sure what is happening at home: "*I need to look slightly in more detail but yeah if someone makes himself a cup of tea then you can see that. And he might get home anywhere between 4 o'clock and 6 o'clock. So … but as soon as he gets home he turns on the computer. But I am not sure, the computer isn't that visible.*"

When asked if the data had helped him learn about his household, P11 said "*A little bit. I think it could be more micro. I'd like to get down to the nuts and bolts. I'd like to know about (…) zonal areas where it actually shows you, kind of in a visualization of your house, showing you heat maps of where the usage is taking place so you can quickly zone in on these areas, as opposed to me having to turn things on and go back turn things off* (…) *I wouldn't say I've become better. I could preempt when (…) electric usage was gonna happen (…) I think with it being the peaks of the entire*

house … if it could be more detailed. I think that'd be better." One specific insight he had gained over time was related to the electricity usage of his teenage daughter: "*Yeah, the peaks and troughs (…) in the diagram. Looking at what has been used over the particular day and when (…) I can do as I say turn to the rooms and be "hang on a minute there must be something left on somewhere." And I can actually trace those things (…) her room is slightly … you know, the epicenter of all electricity usage.*"

P13 commented on his monitoring over time saying he "*was interested a little bit in how it would develop. So with more data, because in the beginning fluctuations were high, and then after several months you can say, okay, so this is our weekly average.*" Further, he had gained understanding about the relative consumption of the washing machine and the electric shower over time: "*So about the first interview where I was surprised that the washing machine draws more electricity that the shower. Now, that is because of the integration period. So the actual amount when the shower is on is much higher, but then we don't shower for 90 minutes. While the washing machine runs for 90 minutes. Because of the binning of the graphs it looks like the washing machine draws … I mean the total amount of energy is higher. But the peak amount is less. So yeah. Given that some thought.*" He had also consulted the Loop data when he re-negotiated his direct debit with his energy provider. Although his bills and the Loop data were not completely congruent, he used the Loop to get a better idea what he was spending on electricity every month. He answered the question if he had learned from the data feedback by saying "*I put some thinking into it which otherwise I wouldn't have. And it gave you some insights (…) Insights like the washing machine, the shower.*"

User requirements

In the second part of the follow-up interview, we asked participants which characteristics and functions would be important to them if they could design the "perfect smart meter feedback." The functions that were requested most often were first, appliance-wise data disaggregation to make the information more actionable, and second, interaction with the software to explore the data.

In reference to the Loop's line graph, P3 pictured the following visualization:

… you could have like an aggregate of how much power you're using. But then underneath that, you can have other lines or bars or some sort of visualization showing 'ok so this is what was contributing to that much (…) you could maybe have a list of appliances. And you could roll over that appliance and then it would go from being like a greyed-out line to being a high-contrast line. So you can see what that particular appliance did over the week or over the month. Or over the day.

P9 brought up a similar idea:

… say having baseline usage, devices that run all the time, colored in one color, and when something new starts, that that takes on a new color. And so when another device starts that becomes another color again. So you end up with like a stacked graph.

Others had the idea of a screen that would show a schematic flat or house or the actual property with its rooms. The display would then show the consumption per room, and per room, they would be able to zoom in and obtain more details on the device level, such as their efficiency and how they could be improved.

I'd like to know about (…) zonal areas where it actually shows you, kind of in a visualization of your house, showing you heat maps of where (…) the usage is taking place so you can quickly zone in on these areas (P11).

One imagined a little display by every light switch and in the display every appliance would be represented with a little picture (of the appliance) that contains information about it.

One suggestion included a smart home system that would integrate information from the Internet and offer tailored advice; every plugged-in device would automatically communicate its specifications as well as its system status to the network, and thereupon, the user could be sent useful notifications. Similarly, another participant suggested that whenever there is a new appliance, there could be a training phase for the system to learn to recognize all appliances. As an example of specific tailored advice, P9 suggested the system could "*tell you that this particular light bulb is using more than the other ones. Or that your TV uses 25% more*

energy than most people's TVs; did you know that you could save energy by … upgrading this device or using it in a different way or something."

Discussion

Main findings

We asked participants which appliances or activities in their homes consume most electricity and how much electricity these consume. We observed a three-way split between participants with very accurate knowledge including quantitative specifications, participants with reasonable guesses, and participants whose guesses were inaccurate. The sketches granted an interesting insight into participants' energy literacy and mental models of how they thought electricity would be used over the day. As expected from the literature, there is a relationship between people's sketches and their understanding of data (Walny et al. 2015). The details of the graph and the ability to label the scales revealed that our sample ranged from very low to very high literacy. Three months later, we observed the same three-way split between those that had quit using the tool immediately after the first interview, a group that had tried but failed, and one group that was still using the tool. That some of our participants stopped engaging with the Loop soon after installing it is consistent with prior work that has also shown that some users abandon smart technologies within the first weeks or months (Harrison et al. 2015; Lazar et al. 2015). In terms of understanding why a person might choose to continue to use a device, Hekler et al. (2013) discuss moderating variables that influence how efficient interventions are for different people. Moderation is important in behavioral theory because research needs to address key differences and cater for different user needs or motivation. We found that only sufficiently energy literate users would continue using the Loop, while more illiterate participants could not be motivated to keep up the tracking. For the literate ones, the drivers seemed to be curious and fascinated with the data (Epstein et al. 2015; Rooksby et al. 2014).

The results of the contextual inquiry give insights into how people interact and reason about time series energy data graphs. We learn that all participants sketch time series graphs—hence, it is an intuitive visualization. However, we found that participants did not find this type of data visualization useful when exploring

their recorded usage data—they often struggled to understand and explain peaks and troughs in the graphs. Participants relied on top-down processes (i.e., they relied on their memory of what they were doing at specific times to help explain patterns of use, instead of processing the given information bottom-up). The implication of this is that people are often biased by "active events" and so possibly overlooked energy-intensive but less routinely performed activities—e.g., washing laundry—or background devices that consume electricity not specifically tied to an event—e.g., the fridge. These findings are in line with Costanza et al. (2012), where participants would mostly look at peaks in the graph, neglecting the baseline.

Concerning the waveform, the staircase-shaped square form (as in Fig. 2c) best represents the real-world matter (Pinker 1990), as it shows clean-cut on- and offsets (mind that P13, the most literate participant in our sample, chose the square wave format). The smooth sine waveform (as in Fig. 2a) reflected lower literacy in the preinterview, and in the contextual inquiry, the sine shape turned out to be a problem as it is impossible to see beginning and ending of an event, and worse, distinct events blend over into one another.

The most prominent theme in the user requirements interview was disaggregation. Everyone demanded a view that would break down usage per appliance, or at least render separate streams visible in the time series display. Appliance-wise disaggregation would allow users to decode the usage information more easily, thus eliminating the problem of using top-down processes in the interpretation. By showing aggregated energy data, current generation smart meters are preventing people from transitioning to the stage of meaningful reflection and action. Froehlich et al.'s (2014) description of personal informatics failures related, among others, to problems regarding the data collection and regarding the user interfaces. In our case, the approach for collecting and displaying aggregated data with low frequency does compromise comprehensibility of the graph. We assume that this very mapping is crucial because users reason in terms of everyday actions and educational approaches should take relevant routines and situations into account (Álvarez and Vega 2009). The action- or event-based nature of thinking about energy consumption is not mirrored in the data, so people fail to map data patterns to behavior and to gain relevant insights. It is assumed that disaggregated feedback would be more actionable for householders, but research has not yet delivered

strong evidence to support this assumption (Sokoloski 2015; Kelly 2016).

Contribution

In this section, we consider three contributions of the study: first, theoretical implications, second, a revision of the term energy literacy, and third, considerations for design recommendations. We discuss each in turn.

First, the data from our study suggests that cognitive information processing must be given more consideration in behavior change frameworks. Sense-making is significantly influenced by preexisting knowledge structures (i.e., mental models and in our case energy literacy) and depends on the users' analytical skills, which may vary considerably between individuals. In the contextual inquiry, we saw that participants struggle to understand the graph and they made guesses as to what caused the data pattern. If users do not understand the eco-feedback, they cannot derive insights for behavior change. The integration stage from Li et al. (2010) is automated as the technology collects and prepares the data. However, the integration has to serve the reflection phase, meaning that the collected data needs to be processed and visualized in a way that facilitates and catalyzes reflection and gain of knowledge.

Second, we plead for a concise and more practical definition of energy literacy. We need to consider in detail what and how much users know, and what specifically we want them to learn. Our data shows that householders lack an understanding of how much electricity they consume for everyday activities. We argue that feedback is "actionable" when it addresses use cases that are relevant to the householder in their everyday lives. That is for example, the awareness that certain appliances use much more power than others, but that some low-consumption appliances (such as the fridge) consume more over time because they are constantly on.

Third, we found that aggregated time series data with low temporal resolution is unfit to represent households' electricity consumption, because the user cannot decode the electricity consumption of everyday actions. If we would like them to use less electricity and more efficiently, a mapping between data and relevant everyday activities is crucial, though. High temporal resolution in line graphs or alternative views that break down electricity usage per room, appliance, or activity are recommended.

Limitations and future work

This study is an exploratory field work with the purpose of examining the problem space and generating more specific research questions regarding the cognitive sense-making processes in interpreting electricity data.

We aimed for a mixed sample to see how different types of users would read and reflect on the data. Yet, there are a number of limitations in our study that we address in the following:

The sample is relatively small. However, the contribution is an in-depth understanding of how people understand energy data. A qualitative research approach is appropriate to address this research question. As opposed to quantitative data where certain sample sizes are required for the validity of statistical tests, an increase in sample size is only useful as it reveals additional themes in the qualitative data. In their guide for qualitative research, Blandford et al. (2016) argue that a pragmatic approach should be taken to recruiting participants. Participants should be recruited until "theoretical saturation" is achieved (i.e., the point at which gathering and analyzing more data on the chosen theme does not yield further insight). We believe that this was achieved in our study. For example, only one out of 13 participants came up with an alternative idea for sketching her daily consumption (and in the end opted for the timeline as everyone else did). The analysis of the interview data revealed reoccurring topics between participants (e.g., memory-based reasoning and disaggregation as the most prominent ones). Given the considerable overlap between participants' data, we are confident that we have reached saturation in observing the cognitive processes involved in householders making sense of the Loop data feedback. Moreover, our sample size is consistent with that seen in previously published qualitative research (e.g., Blandford and Wong 2004; Yang et al. 2014).

We could not interview all participants in person. All are UK households, but in different cities, and some of the participants had very busy schedules and found it inconvenient to meet in person. To respect participants' wishes and due to limited resources for traveling far, we chose to interview some participants via Skype given the practical constraints. The data between Skype and personal interviews are comparable and we have no reason to assume that they had different experiences from one another.

As with the Skype interviews, we interviewed all three participants from household 2 together for convenience reasons and this interview revealed an interesting dynamic. Discussing the data together brings the advantage of combined knowledge and the discussion may stimulate the sense-making process. On the other hand, the disadvantage is that it is harder to determine the individual sense-making capacities. Although potentially confounding in our study, we would argue that interviewing participants alone and in groups within the same study increases ecological validity. The follow-up revealed that many households differed with regard to who looked at the eco-feedback and how much the matter was talked about between different household members.

Another limitation is that we only examined one tool, i.e., the Loop, which specifically uses a line graph with minimal resolution whereas other tools output the data in different formats. While shape and temporal sampling rate may vary, most systems feature aggregated feedback only. Finally, we have investigated how people reason about the usage data displayed by the Loop and possibly learn from it, albeit we did not record actual behavioral measures. It was beyond the purpose of the study to explore if people would reduce their consumption. Although we can confirm that users ask for appliance-wise feedback, it remains to be investigated if they'd perform better with disaggregated data.

There are two core questions for our further research, which concern energy data disaggregation and visualization. The first is to test if disaggregated data is superior to global household consumption data. Second, we plan to examine how disaggregated usage data is best processed and visualized for users to make sense of it and learn from it. The line graph with low temporal resolution did not provide sufficient detail to yield actionable insights. Visualizations that depict each device's usage might be more useful. Also, interactivity could improve users' performance by allowing them to actively explore the data.

In summary, we have shown that for eco-feedback to increase householders' practical knowledge and energy literacy, the information must be communicated in a way that makes sense to them. For the information to be relevant and easy to act upon, it has to map to householders' everyday life

with its routines. While unfit visualizations of the data make this mapping hard or impossible, we have identified promising design factors (such as high temporal resolution, wave form, and disaggregation) that take cognitive information processing into account and could significantly enhance users' learning experience with residential electricity data.

Compliance with ethical standards

Conflict of interest   The authors declare that there is no conflict of interest.

References

Álvarez, P., & Vega, P. (2009). Actitudes ambientales y conductas sostenibles. Implicaciones para la educación ambiental. *Revista de Psicodidáctica, 14*(2), 245–260.

Aronson, J. (1994). A pragmatic view of thematic analysis. *The Qualitative Report, 2*(1), 3. doi:10.4135/9781446214565. n17.

Bager, S. L. (2014). Information feedback, behaviour and 'smart meters': using behavioural economics to improve our knowledge about the potential effectiveness of Smart Meters to use electricity efficiently. IIIEE Master thesis.

Baur, D., Lee, B., & Carpendale, S. (2012). TouchWave: kinetic multi-touch manipulation for hierarchical stacked graphs. Proceedings of the 2012 ACM International Conference on Interactive Tabletops and Surfaces, 255–264. doi:10.1145 /2396636.2396675.

Blandford, A., Furniss, D., & Makri, S. (2016). Qualitative HCI research: going behind the scenes. *Synthesis Lectures on Human-centered Informatics, 9*(1), 1–115.

Blandford, A., & Wong, W. (2004). Situation awareness in emergency medical dispatch. *International Journal of Human-Computer Studies, 61*(4), 421–452. doi:10.1016/j. ijhcs.2003.12.012. 27, 35, 38, 68.

Boy, J., Rensink, R. A., Bertini, E., & Fekete, J. D. (2014). A principled way of assessing visualization literacy. *Visualization and Computer Graphics, IEEE Transactions on, 20*(12), 1963–1972.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2015), 77–101. doi:10.1191/1478088706qp063oa.

Brewer, r. S. (2013). Fostering sustained energy behavior change and increasing energy literacy in a student housing energy challenge. Doctoral dissertation, University of Hawai'i at Manoa.

Buxton, B. (2007). Sketching user experiences: getting the design right and the right design (interactive technologies).

Cheng, P. C. H. (2011). Probably good diagrams for learning: representational epistemic recodification of probability theory. *Topics in Cognitive Science, 3*(3), 475–498.

Cheng, P. C. H. (2014). Graphical notations for syllogisms: how alternative representations impact the accessibility of concepts. *Journal of Visual Languages & Computing, 25*(3), 170–185.

Chisik, Y. (2011). An image of electricity: towards an understanding of how people perceive electricity. Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics). *LNCS, 6949*(PART 4), 100–117. doi:10.1007/978-3-642-23768-3_9.

Cleveland, W. S., & McGill, R. (1984). Graphical perception: theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association, 79*(387), 531–554.

Costanza, E., Ramchurn, S. D., & Jennings, N. R. (2012). Understanding domestic energy consumption through interactive visualisation: a field study, 1–10. doi:10.1145 /2370216.2370251.

Darby, S. (2006). The effectiveness of feedback on energy consumption. A Review for DEFRA of the Literature on Metering, Billing and Direct Displays, 486, 2006.

Department of Energy and Climate Change DECC (2012). Smart Metering Equipment Technical Specifications.

DeWaters, J. E., & Powers, S. E. (2011). Energy literacy of secondary students in New York State (USA): a measure of knowledge, affect, and behavior. *Energy Policy, 39*(3), 1699–1710. doi:10.1016/j.enpol.2010.12.049.

DeWaters, J., Powers, S., & Graham, M. (2007). AC 2007-1069: developing an energy literacy scale. Age, 12, 1.

Ehrhardt-Martinez, K. (2010). People-centered initiatives for increasing energy savings, 202–507.

Ellegård, K., & Palm, J. (2011). Visualizing energy consumption activities as a tool for making everyday life more sustainable. *Applied Energy, 88*(5), 1920–1926.

Epstein, D. A., Ping, A., Fogarty, J., & Munson, S. A. (2015, September). A lived informatics model of personal informatics. In Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing(pp. 731–742). ACM.

Fischer, C. (2008). Feedback on household electricity consumption: a tool for saving energy ?, 79–104. doi:10.1007/s12053-008-9009-7.

Fischer, J. E., Crabtree, A., Rodden, T., Colley, J. A., Costanza, E., Jewell, M. O., & Ramchurn, S. D. (2016). "Just whack it on until it gets hot": working with IoT data in the home.

Froehlich, J. (2011). Sensing and feedback of everyday activities to promote environmental behaviors. University of Washington Doctoral Dissertation.

Froehlich, J. E., Kay, M., Larsen, J. E., & Thomaz, E. (2014, September). Disasters in personal informatics: the unpublished stories of failure and lessons learned. In Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication(pp. 673–678). ACM.

Fry, E. (1981). Graphical literacy. *Journal of Reading, 24*(5), 383–389.

Gabe-Thomas, E., Walker, I., Verplanken, B., & Shaddick, G. (2016). Householders' mental models of domestic energy consumption: using a sort-and-cluster method to identify shared concepts of appliance similarity. *PloS One, 11*(7), e0158949.

Galesic, M., & Garcia-Retamero, R. (2011). Graph literacy: a cross-cultural comparison. *Medical Decision Making: an International Journal of the Society for Medical Decision Making, 31*(3), 444–457. doi:10.1177/0272989X10373805.

Greenberg, S., Carpendale, S., Marquardt, N., & Baxton, B. (2012). The narrative storyboard: telling a story about use and context over time. *Interactions, 19*(1), 64–69. doi:10.1145/2065327.2065340.

Harrison, D., Marshall, P., Bianchi-Berthouze, N., & Bird, J. (2015). Activity tracking barriers, workarounds, and customisation. *Ubicomp, 2015*, 617. doi:10.1145/2750858.2805832.

Hekler, E. B., Klasnja, P., Froehlich, J. E., & Buman, M. P. (2013, April). Mind the theoretical gap: interpreting, using, and developing behavioral theory in HCI research. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 3307–3316). ACM.

Hofman, H. (1980). Energy crisis—schools to the rescue again. *School Science and Mathematics, 80*(6), 467–478.

Holtzblatt, K., & Jones, S. (1993). Contextual inquiry: A participatory technique for system design. *Participatory design: Principles and practices*, 177–210.

Kelly, J. (2016). Is disaggregated electricity feedback more effective than aggregate feedback at reducing energy consumption? In preparation.

Kirsh, D. (2010). Thinking with external representations. *AI and Society, 25*(4), 441–454. doi:10.1007/s00146-010-0272-8.

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science, 11*(1), 65–100.

Lazar, A., Koehler, C., Tanenbaum, J., & Nguyen, D. H. (2015). Why we use and abandon smart devices. Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, 635–646. doi:10.1145/2750858.2804288.

Lewis, C., & Rieman, J. (1993). Task-centered user interface design. A practical introduction.

Li, I., Dey, A., & Forlizzi, J. (2010). A stage-based model of personal informatics systems. Proceedings of the 28th International Conference on Human Factors in Computing Systems—CHI '10, 557. doi:10.1145/1753326.1753409.

Munzner, T. (2014). Visualization analysis and design. CRC Press.

Murugesan, L. K., Hoda, R., & Salcic, Z. (2014). Visualization of electricity consumption: software prototype through literature, 3(3), 2725–2729.

Pereira, L., Quintal, F., Barreto, M., & Nunes, N. J. (2013). *Understanding the limitations of eco-feedback: a one-year long-term study, In human-computer interaction and knowledge discovery in complex, unstructured, big data (pp. 237–255)*. Berlin Heidelberg: Springer.

Pinker, S. (1990). A theory of graph comprehension. *Artificial Intelligence and the Future of Testing*, 73–126.

Prochaska, J. O., Redding, C. a., Harlow, L. L., Rossi, J. S., & Velicer, W. F. (1994). The transtheoretical model of change and HIV prevention: a review. *Health Education & Behavior, 21*(4), 471–486. doi:10.1177/109019819402100410.

Rego Teixeira, A. (2014). Le smart-metering en Suisse Romande: conditions sociales d'acceptation et d'appropriation. University of Geneva.

Roberts, S., & Baker, W. (2003). Improving consumer feedback on energy consumption. A report to Ofgem.

Rooksby, J., Rost, M., Morrison, A., and Chalmers, M. (2014) Personal tracking as lived informatics. CHI, 1163–1172.

Sokoloski, R. (2015). Disaggregated electricity consumption: using appliance-specific feedback to promote energy conservation. M.A. California State University San Marcos.

Tong, X., Gromala, D., Bartram, L., & Carpendale, S. (2015). Evaluating the effectiveness of three physical activity visualizations—how people perform vs. perceive. In personal visualization: exploring data in everyday life, 1.

Tufte, E. R. (1985). The visual display of quantitative information. *Journal for Healthcare Quality, 7*(3), 15.

Tversky, B., Corter, J. E., Nickerson, J. V., Zahner, D., & Rho, Y. J. (2008). Transforming descriptions and diagrams to sketches in information system design. Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics), 5223 LNAI, 242–256. doi:10.1007/978-3-540-87730-1_23.

Wallenborn, G., Orsini, M., & Vanhaverbeke, J. (2011). Household appropriation of electricity monitors. *International Journal of Consumer Studies, 35*(2), 146–152.

Walny, J., Huron, S., & Carpendale, S. (2015, June). An exploratory study of data sketching for visual representation. In *Computer Graphics Forum* (Vol. 34, No. 3, pp. 231–240).

Yang, R., Newman, M. W., & Forlizzi, J. (2014). Making sustainability sustainable: challenges in the design of eco-interaction technologies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 823–832). ACM.

Yun, T. J., Jeong, H. Y., Kinnaird, P., Choi, S., Kang, N., & Abowd, G. D. (2010). Domestic energy displays: an empirical investigation. ACEEE Summer Study on Energy Efficiency in Buildings, 7–348–360.

Zhang, J., & Norman, D. A. (1994). Representations in distributed cognitive tasks. *Cognitive Science, 18*(1), 87–122.