

An Empirical Investigation of Domestic Energy Data Visualizations

MELANIE R. HERRMANN¹, DUNCAN P. BRUMBY¹, LONGMIN CHENG¹, XAVIER M. P. GILBERT², AND TADJ ORESZCZYN³

¹ UNIVERSITY COLLEGE LONDON, INTERACTION CENTRE, DEPARTMENT OF COMPUTER SCIENCE, FACULTY OF ENGINEERING, 66-72 GOWER STREET, LONDON, WC1E 6BT, UK

² IMPERIAL COLLEGE LONDON, GRANTHAM INSTITUTE, SOUTH KENSINGTON CAMPUS, LONDON, SW7 2AZ, UK

³ University College London, The Bartlett School of Environment, Energy and Resources, Energy Institute, Faculty of the Built Environment, 14 Upper Woburn Place, London, WC1H 0NN, UK

Corresponding Author: Duncan Brumby, University College London, Interaction Centre, Department of Computer Science, Faculty of Engineering, 66-72 Gower Street, London, WC1E 6BT, UK. E-mail: d.brumby@ucl.ac.uk.

ABSTRACT

Which device in your home uses the most electricity? Many people have a poor understanding of their domestic energy consumption. In this paper, we evaluated three data visualizations used to deliver feedback. These were: (1) an aggregated line graph – showing changes in total electricity consumption over time, (2) a disaggregated line graph – showing changes in electricity consumed over time but separated out at the appliance-level, and (3) an area-based visualization – showing the cumulative energy consumed by different appliances over a given time period. In an experiment, 65 participants used one of these three visualizations to make sense of the same pattern of domestic electricity data. Participants who used the area-based visualization gained a more accurate understanding of how much electricity different domestic appliances were using compared to participants who were shown time series data. These results suggest that the choice of data visualization will impact people’s understanding from smart metering systems, and that appliance-wise disaggregation offers the most promising approach for visualizing domestic electricity consumption data.

KEYWORDS: SMART METERING; DISAGGREGATION; INFORMATION VISUALIZATION; GRAPHICAL LITERACY; TIME SERIES DATA.

HIGHLIGHTS

- Investigates how the choice of information visualization influences people in making sense of data for domestic electricity consumption.
- Results of experiment suggest that by using an area-based data visualization, people can gain a more accurate understanding of how much electricity different domestic appliances use.
- These results are important because the information display of current smart meters tend to use time-series data visualization, which we show to be less effective for enabling users to develop an understanding of domestic electricity consumption.

1. INTRODUCTION

Over the past couple of decades, there have been advances made in the development of technologies that make domestic energy consumption data easily accessible to users [21,22,58]. By deploying Smart Meters with In-Home Displays and other smart Home Energy Management Systems, it is hoped that householders will be able to gain a better and more accurate understanding of how they are consuming energy in the home. In turn, this improved knowledge can potentially help people reduce their consumption and save both energy and money. When integrated across all of society, such reductions in energy demand can lead to significant decrease in greenhouse gas emissions, contributing to the ongoing transition to a low-carbon future and the fight against anthropogenic climate change.

How exactly are Smart Meters and Home Energy Management Systems supposed to help householders gain a better and more accurate understanding of how they are consuming energy? Conventional energy bills have been likened to buying groceries in a store where products are sold without price tags and billed via monthly statements [42]. If customers do not know how much they spend on individual items, how can they possibly hope to economize? The purpose of this metaphor is to highlight why it is important to develop systems that can give people far more detailed feedback on how they are consuming energy. There have been many advances in current technologies for recording energy consumption at a high sampling rate. It is now possible for users to easily access data on how much energy is being used in their home every second of every day. Despite this fine level of temporal granularity, most systems still do not offer ‘price tags’ for separate items; instead, data on total consumption across all devices and appliances in the household is given, giving little improvement to householders over conventional bills.

There is a lot of work being done on *disaggregating* a household’s electricity data so that consumption can be tracked down to the appliance-level [6]. This has the potential to have a transformative impact on end-users’ behavior [3], allowing them to learn how much energy each

appliance in their home is using. For example, does a typical household refrigerator use more or less energy during the week than a typical household dishwasher? The hope is that by providing people with feedback on disaggregated energy data, they will be able to answer these kinds of questions about how they are consuming energy. Surprisingly though, a meta-review of twelve studies investigating how people make sense of disaggregated energy data [41], found very limited empirical evidence to support the idea that this actually helps people to understand how they are consuming energy in the home and to make actionable energy saving behavior changes. This finding is surprising, and we wonder why these attempts to give people disaggregated feedback on their domestic energy consumption does not appear to work. This brings up the question whether the way in which disaggregated energy data is presented to end-users has an impact on whether they can use the data to improve their understanding of how energy is being consumed in the home.

An important consideration is how best to *visualize* energy data to the householder. Domestic energy use is captured as time series data, typically visualized as a line graph depicting power over time. Reading such representations requires a level of both energy literacy and graphical literacy. Energy literacy is the understanding of physical concepts, such as comprehending that energy is the product of power consumed over time [10]. Graphical literacy [16] is the intellectual skill to communicate and understand visual aids, such as line graphs. Given that many people have both sparse energy literacy [10] as well as poor graphical literacy [25], the question of how best to visualize energy data has been much overlooked in the literature [53]. Understanding human perception is critical to visual design [34] and assessing people's ability to understand visualizations should be of paramount importance when communicating user-relevant data [8].

In this paper, we present the results of an experiment to evaluate three data visualizations of domestic electricity consumption. These data visualizations were: (1) an aggregated line graph – showing changes in total electricity consumption over time, (2) a disaggregated line graph – showing

changes in electricity consumed over time but separated out at the appliance-level, and (3) an area-based visualization – showing the cumulative energy consumed by different appliances over a given time period. In a laboratory experiment, participants used one of these three visualizations to make sense of patterns of domestic electricity consumption data. We were interested in finding out which visualization helped participants gain a more accurate understanding of how much electricity different domestic appliances were using.

The contribution of this work lies in:

- Addressing the topical question of whether disaggregation is crucial for eco-feedback technologies.
- Providing a systematic empirical assessment under controlled conditions and free from confounding variables.
- Testing whether people gain more knowledge about the electricity consumption of everyday actions depending on the choice of data visualization.
- Offering design recommendations for future Smart Meter and Home Energy Management Systems to summarize how much energy domestic appliances use.

2. BACKGROUND

Before describing the details of this study, we first review important related research that has addressed relevant aspects regarding disaggregation, information visualization, and users' literacy to understand energy-related visualizations.

2.1 Disaggregation

The comparison of energy bills and monthly statements for groceries shopping [42] has been cited many times over the past few decades and it has been suggested that disaggregation of household energy data could be the 'holy grail of energy efficiency' [3]. The common belief is that low-cost energy reductions will be achieved through behavior change once disaggregated feedback is provided to households. The

cost of rolling out smart meters would be justified by the long-term savings and further advancements such as demand-side management and load shifting. Research has therefore been focused on addressing the many technical challenges associated with disaggregating electrical signals down the powerline. This is usually achieved through a process of computationally analyzing changes in voltage and current to identify appliances, known as Non-Intrusive Load Monitoring [3,6,26,33].

Few systematic investigations have been carried out to assess the ramifications of disaggregated eco-feedback on end-users' behavior and knowledge of how energy is consumed. Indeed, studies that have directly compared aggregated and disaggregated feedback have found that aggregated feedback was as good as or better than disaggregated feedback [44,48,56]. Kelly and Knottenbelt [41] reviewed these, and nine further studies, and concluded that the validity of their findings was severely limited by methodological biases. For example, Sokoloski [56] compared different groups using entirely different feedback mechanisms and data visualizations. In the study, one group received aggregated feedback through an in-home display that participants could easily access throughout the day, while another group received disaggregated feedback through an online platform that they could only access after logging into the system. These confounding factors make it difficult to assess how effective giving disaggregated energy feedback is.

Controlled laboratory studies have found promising results regarding disaggregation. Yun et al. [63] investigated different levels of resolution of energy feedback, considering differences in spatial resolution (e.g., total, per room, or per appliance) and temporal resolution (e.g., current usage, history of usage, and prediction of future usage). Yun et al. found that appliance-wise disaggregation was beneficial, both for novices and experts.

In a previous experiment, we [36] compared different data visualizations for representing disaggregated energy feedback that varied the temporal resolution of the data being presented. We found that showing moment-to-moment changes in consumption over time across different appliances in the

form of a time series line graph resulted in participants having a less accurate understanding of how much energy was consumed compared to when a data visualization was used that showed only the total consumption for each appliance normalized across the duration of usage. These results suggest that giving high temporal resolution feedback can potentially hinder people's ability to make sense of and interpret domestic electricity consumption data. We further discuss this in the following paragraphs on data visualization.

2.2 Information visualization

People's ability to understand visualizations should be of paramount importance when communicating user-relevant data. Boy et al. [8] define visualization literacy as the ability to use data visualizations to handle information in an effective, efficient, and confident manner. Boy et al. also point out that few studies examine how well people can extract information from graphs. Previous research shows that data comprehension greatly depends on the manner of presentation and the design of the display interface [14,63]. As a result, we are taking a closer look at how to visualize energy data for the user, ranging from abstract and ambient information visualization to graphic representations.

There is a rich tradition of developing information visualizations that use abstract, ambient and artistic techniques to communicate data to users. For example, Holmes [37] proposes to combine art and technology to help householders reduce their energy consumption. Looking to visualize the use of individual appliances, Rodgers et al. [54] found that a viable way to provide feedback in a pleasingly aesthetic way was to use artistic symmetrical and colorful shapes that responded to and changed in line with energy use. Taking a similar approach, Gustafsson et al. [28] deployed a prototype which looked like a common electrical power strip that was modified so that it displayed the amount of energy passing through it using dynamic glowing patterns produced by electroluminescent wires molded into the transparent electrical cord. These examples provided tangible feedback at localized points in the home.

In contrast to these abstract and ambient displays, graphical data visualizations are more commonly developed for digital interfaces, such as in-home displays and web apps for domestic smart meters. Here we focus on two published studies, one by Costanza et al. [17], and one of our own studies [36], both of which have addressed the question of how to visualize disaggregated electricity data using graphical displays.

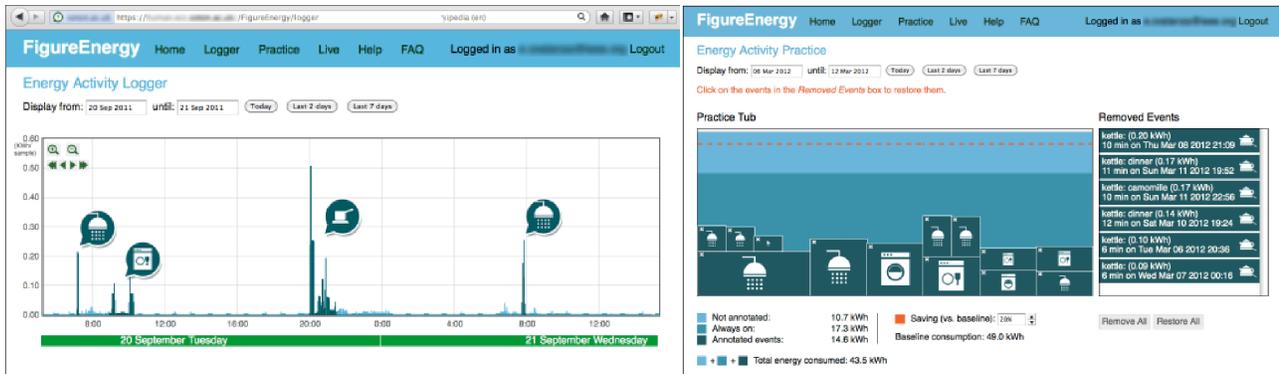


Figure 1. The FigureEnergy Data Visualization with Logger View (left) and Practice View (right). Taken from Costanza et al. [17].

Costanza et al. [17] built and evaluated a prototype data visualization, called FigureEnergy, which was designed to help people make sense of domestic energy data consumption. The FigureEnergy interface had two views. In the Logger View (left panel, Figure 1), users could see their electricity consumption as a time series line graph, showing power over the duration of a day. The interface was interactive meaning that users could annotate the graph and indicate which appliances they used throughout the day. To review how much electricity each appliance was using, users could switch to a second Practice View (right panel, Figure 1). The Practice View used an area-based visualization to show the cumulative energy consumed by each appliance: each appliance was represented by a separate two-dimensional box, the size of which was proportional to the amount of energy consumed by that appliance throughout the day. Pictographic icons were used to label each box to make it clear which appliance it referred to. This data visualization was designed to allow users to quickly and easily get a

sense of how much electricity each appliance was using: the appliance with the biggest box was using the most electricity. While we assume that the area-based summary provided by the Practice View helped users to learn which appliances consume the most electricity in the home, Costanza et al.’s study did not investigate this question in detail.

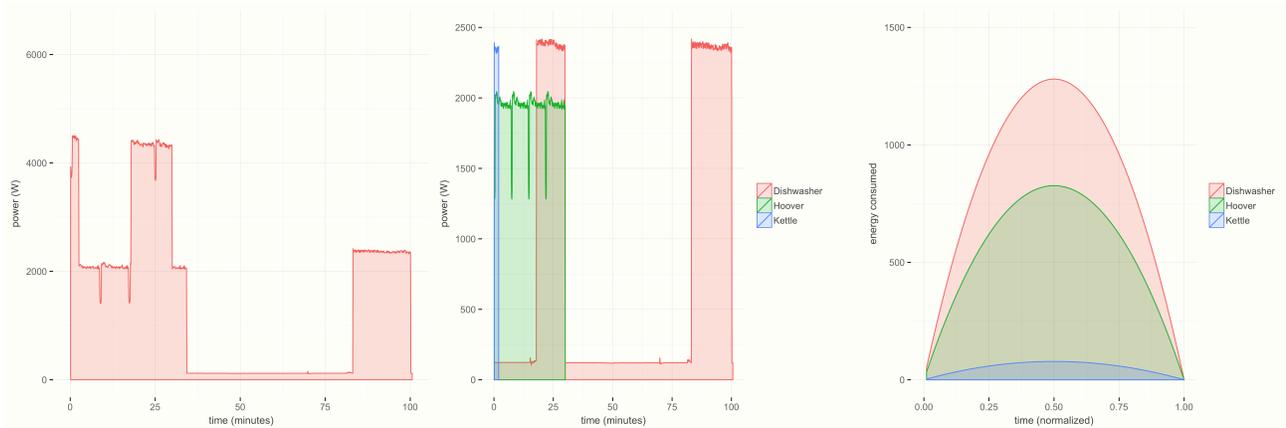


Figure 2. Three energy data visualizations: aggregated line graph (left), disaggregated line graph (middle), and area-based (right). Taken from Herrmann, Brumby, Oreszczyn & Gilbert [35].

We previously [36] conducted a lab study to assess the effectiveness of different data visualizations for helping people to learn how much electricity is consumed by domestic appliances per typical use. In the study, participants were shown one of three different visualizations of the same underlying domestic energy dataset: (1) an aggregated line graph, (2) a disaggregated and color-coded line graph for each appliance, and (3) a disaggregated *and statistically normalized* graph for each appliance. All three visualizations show the electricity usage patterns of a dishwasher, a vacuum cleaner, and a kettle running at the same time. In the aggregated condition (left panel, Figure 2), total consumption for all three appliances is represented using a single line graph. In the disaggregated condition (middle panel, Figure 2), consumption data from each appliance is separated and represented as different colored lines in the graph. In the area-based condition (right panel, Figure 2), the time-

dimension was eliminated from the data to normalize the consumption of the three appliances over their usage duration – creating a stacked visualization.

To evaluate each of these different data visualizations, we [36] asked participants questions about which of two appliances used the most energy. For example, a participant was asked: does running the dishwasher use more energy than using the vacuum cleaner? Participants who were exposed to the area-based visualization (right panel, Figure 2) were more likely to answer these energy literacy questions correctly than participants who saw either of the time series visualizations. Surprisingly, there was no difference in participants' accuracy on the energy literacy test between participants who used either the aggregated or disaggregated time series visualizations (left panel vs. middle panel, Figure 2).

While normalizing energy data may have yielded the best results in terms of test accuracy, it uses a non-standard and unfamiliar visualization format. Lee et al. [45] suggest that great care must be taken when giving novices an unfamiliar visualization because they can be difficult to understand, and users draw unintended conclusions from the visualization. For example, a user might assume that the x-axis of the normalized visualization represents time, and so therefore might not be able to make sense of the normalized units. Consistent with this concern, some participants in the normalized condition had to be excluded from the data analysis because they reportedly did not understand the visualization. This is not surprising as the normalized condition is very atypical and unfamiliar graphics are not recommended for use with an untrained audience [50]. A more appropriate area-based visualization are the pictographic representations developed by Costanza et al [17], which potentially reduce the cognitive effort needed to interpret the energy consumption data.

2.3 Household's literacy

Previous work in HCI has revealed that 'smart home' technologies are not as smart as users believe them to be and as a result, they often fail to fully understand them [13,35,62]. There are two factors that

played a role in users' failing to understand energy data feedback: (1) poor knowledge about energy and (2) poor understanding of how to read and understand data visualizations. We review each in turn.

First, most people have little knowledge about energy, its physical properties, or about how much electricity they consume when carrying out everyday behaviors at home, such as washing laundry or watching TV [9,15,19,23,43]. When being asked how to save energy, a classic response is 'switching off the lights when not needed'. Attari, DeKay, Davidson, and Bruine de Bruin [4] found that people often systematically overestimate the energy used by highly salient but low-energy activities, such as having the lights on. Conversely, people systematically underestimate the energy used by high-energy appliances that are used less frequently (e.g., the washing machine). These misconceptions affect how people go about saving energy and impair reasonable decision making in the process. There is a general consensus in the literature that feedback needs to address daily routines to help people learn about their energy consumption patterns and make reasonable changes [1,18,20,57,60]. In this paper, we focus on energy literacy as householders' understanding of how much energy they consume by carrying out everyday practices (such as watching TV or washing laundry).

Second, 'graphicacy' [5], or 'graphical literacy' [16,24,51], is poorly trained in most people. This skill refers to people's competence to extract relevant information encoded within a graph or data visualization. If householders are unable to learn the relevant information from energy feedback displays, they will be unable to make changes even if they are motivated to do so [49]. Kempton and Layne [42] explicitly argued that: "the conclusions consumers can draw ... are restricted by the form in which they receive price and consumption data and their limited analytic capabilities" (pp. 857). This suggests that energy data visualizations should be designed to make it easy for people to analyze and decode the information presented within the visualization [12,38]. Despite this, the question of how to present energy feedback information has been much overlooked in the literature [53].

For this study, it is relevant to see how previous work has operationalized and measured the effect of energy data representations on people's literacy (i.e., their learning about how much electricity everyday activities in the home consume). Power rating quizzes, pair-wise comparisons, and list rankings have been used for participants to decide which appliances consume the most energy and which ones consume the least after having been exposed to energy usage information [2,36,63]. As outlined above, we previously [36] evaluated how much people learnt from a data visualization by asking them a series of questions about which of two appliances used the most energy. The dependent variable in these tasks was response accuracy – the ability to correctly identify the appliance that uses the most electricity over a given usage period. Further outcome measures that have been used are response time and response confidence [13,32].

2.4 Purpose of current study

The aim of the current study is to investigate how the choice of information visualization influences people in making sense of data for domestic electricity consumption. This is an important and timely question as there are different ways in which disaggregated appliance-centric energy data can be presented to users. Line graphs are the most commonly used representation for energy data, but they seem poorly suited to convey to users how much energy everyday practices consume. Previous research has shown that there are problems with line graphs [31,35], and rather than knowing their consumption pattern over time, householders need to know which practices consume a lot of energy. Area-based appliance-centric (or event-centric) visualizations (like the one in FigureEnergy [17]) seem to be much more in line with householders' mental models. Following social practice theory [29,30], householders think about energy use in terms of everyday practices, such as making tea or washing laundry. Hence, visualizations that convey practice-centric feedback might be better suited to convey this information.

It's important to note that we do *not* consider whether people are interested in learning about their actual domestic electricity consumption patterns or whether feedback actually leads to changes in

consumption behavior. Instead, the motivation for our study is to compare the standard time-series line graph to an area-based appliance-centric visualization for making sense of data for domestic electricity consumption data. Based on the reviewed literature, the purpose of this study is to evaluate the following two hypotheses:

- 1) Time series line graphs can provide some information about how much energy appliances consume. However, aggregated data hardly allows for comparisons between appliances. Disaggregated time series line graphs provide information by showing clearly visible start and end points, separating out the energy consumed per appliance. For instance, one can easily see in Figure 2 that the dishwasher has consumed more energy than the vacuum cleaner, which in turn has consumed more than the kettle. We assume that disaggregated line graphs allow users to learn more about the energy consumption of household practices than aggregated line graphs.
- 2) Area-based visualizations using pictographs, as utilized by Costanza et al. [17] and Haroz et al. [32], go a step further in processing the raw data, thus reducing the cognitive effort needed to interpret the information. The ups and downs in time series graphs make the visual comparison error-prone for appliances with similar profiles. Instead, the area under the curve could be transformed with one minimum alternation into a rectangle with duration on the x-axis and average power on the y-axis. However, using this method there is a risk that some rectangles would be very large while others may be very small because of the extreme differences in power and duration of usage between appliances. To overcome this issue, we're adding one further transformation, reshaping the representation into near-squares. The readability of these box shapes is further improved by using the icon of the appliance as the box itself (rather than relying on a key to match a line graphs' affiliation to an appliance). We assume that this visualization allows users to learn more about the energy consumption of household practices than line graphs.

To test these two hypotheses, we adapt the experimental setup of our previously published study [36] in which the effect of different data visualizations on users' data understanding was tested. That study found tentative evidence that area-based visualizations, which are free from the ups and downs in time series lines graphs, are superior in feeding back disaggregated data.

However, the area-based representation chosen in that study was an atypical graph that led to confusion amongst participants (see right panel, Figure 2). Also, the simulation used artificially isolated runs of appliances, not complex time series over a full day, as it would be the case in real-life eco-feedback displays. It could be due to this simplification that the study failed to find an advantage of disaggregated time series over aggregated time series line graphs. To create a more realistic visualization showing complex time series over a full day, we are adopting design ideas from FigureEnergy [17]. This field study presented both time series line graphs as well as an area-based visualization using two-dimensional boxes to represent the cumulative usage of appliances (see Figure 1).

We are interested in quantifying which of the two views yields better learning outcomes for users in terms of understanding how much energy they use for everyday household tasks: the time series line graph - which is still the most common form of visualization for electricity data - or the area-based view summarizing cumulative energy. This is a highly relevant question considering the hopes that are being placed in smart home technologies: if users don't learn which appliances are using a lot of energy, they cannot identify how to make savings; and if appliance-centric visualizations don't teach users where they are using most energy, then efforts would be wasted on disaggregation.

In the current study, we compare two different kinds of line graphs - aggregated and disaggregated - as well as an area-based visualization to evaluate their impact on users' understanding of how much energy typical household appliances consume. First, we predict that the disaggregated line

graph will result in higher test accuracy than the aggregated line graph. Second, we predict that the area-based visualization will yield better test results than both types of line graphs.

3. METHOD

3.1 Participants

We recruited 68 participants (41 female, 27 male) through the Psychology Subject Pool at UCL. Of the 68 participants, 57 were between 21 and 35 years old, three participants were younger (between 18 and 21) and five were 36 years or older. All had normal or corrected to normal vision and were accustomed to reading from left to right. Participants received course credit or a small payment for taking part in the study.

3.2 Design

The study used a single-factor between-subjects design with the electricity data visualization as the independent variable. There were three factor levels of data visualization: (1) an aggregated line graph – showing changes in total electricity consumption over time, (2) a disaggregated line graph – showing changes in electricity consumed over time but separated out at the appliance-level, and (3) an area-based visualization – showing the cumulative energy consumed by different appliances over a given time period.

We were interested in assessing the extent to which each of these different visualizations allowed participants to gain an accurate understanding of how much electricity typical household practices consume (e.g., making coffee or running the dishwasher). To assess this, we had participants complete an energy game. In this game, participants were asked to decide which of two practices used the most electricity (i.e., a two-alternative forced-choice). We recorded response accuracy, response time, and response confidence (on a 5-point Likert scale). These measures were used to compare the participants'

performance both before and after they used one of the data visualizations to make sense of a complex domestic electricity consumption dataset from a typical household (details on this are given below).

3.3 Materials

Throughout the experiment, we used ‘Jack’ as a persona to embed the experiment into a story about residential smart metering. Participants were told that Jack and his family live in a London end-of-terrace house and that they have a Smart Meter in their house. When participants used a data visualization to make sense of the electricity consumption data, we told them that this was the feedback that Jack also received from the Smart Meter. It was explained that they would see the data of one 24-hour period at a time, and that they would be shown a simulation of seven days in total, one day at a time. The simulation was meant to imitate a typical week in Jack’s house, where there was more activity during the weekend than on weekdays. Sometimes appliances were used in isolation, sometimes they would overlap. There were appliances that would consume a lot of electricity for a few minutes only (e.g., the kettle), and others that would consume less power over a longer period (e.g., the TV). The fridge generated a constant 24-hour baseload. The energy data that was presented was the same for all conditions, but the type of data visualization that participants saw in the simulation to make sense of this complex dataset depended on the condition they were assigned to.

The problem that developers face is how to best visualize a dataset to communicate the actionable information clearly and easily for users to gain knowledge. People are better at interpreting familiar diagrams [50] and according to Pinker’s [51] theory of graph comprehension, a graphic representation should preserve the physical properties of what they represent (power x time = energy). Hence, line graphs are predominantly used to visualize energy data because they are intuitive. In the first condition, participants saw the aggregated electricity of all appliances in a single line graph showing power in Watts on the y-axis over time of day on the x-axis. We refer to this condition as ‘aggregated’. Like the Logger View in Costanza et al. [17], the curve shows the total electricity consumption in the

household, merely labelling the onsets in time when the appliances were turned on. This was done to provide participants with an understanding of the pattern of domestic activities that were taking place in the home. Some information is visible enough in this condition: one can detect the baseline of the fridge, for example, and the spike of the kettle is quite distinct. On the other hand, it is nearly impossible to make out the exact pattern of the dishwasher.

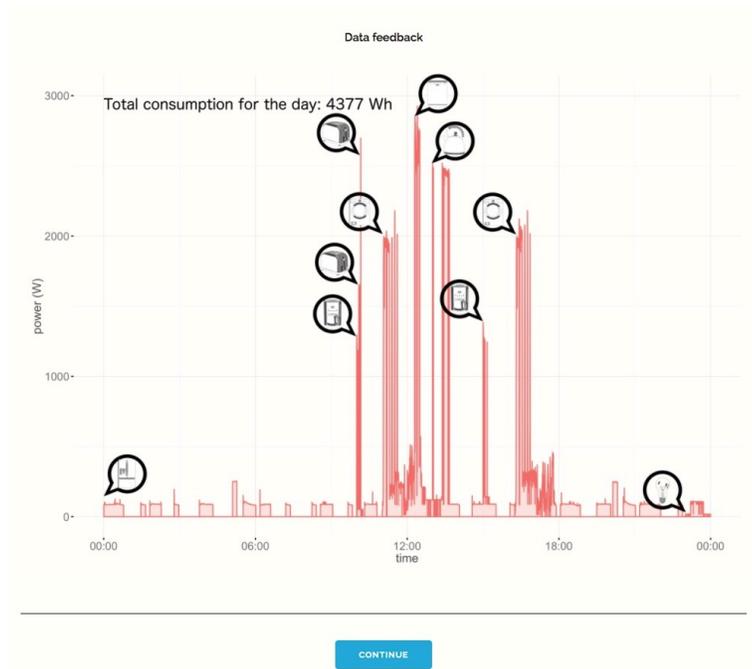


Figure 3. The aggregated time series line graph visualization.

Given the challenges of aggregation, we chose to separate out the patterns of appliances, using multi-color coded line graphs [39]. In the second condition, participants saw the electricity used over a day, but this data was disaggregated: each appliance was visualized as a separate line using a distinctive color (Figure 4). We refer to this condition as the ‘disaggregated’ condition. The idea is to facilitate the distinction of how the power consumption of each appliance varies over time throughout a period of usage and contributes to the total usage depicted. For example, the dishwasher has a distinct pattern with two peaks and a period of lower usage in the middle. This pattern becomes visible in the disaggregated condition but is hidden in the aggregated condition. Another example is the second usage of the toaster –

in the aggregated condition the toaster and coffee maker add up to a higher peak and the contribution of each becomes blurred. In this way, it seems intuitive that the disaggregated data visualization should aid participants as they make sense of how much electricity different devices are consuming. The information how much energy the appliances consume is depicted by the area under the distinct curves. It is well visible, for example, that the light is the smallest contributor to the total energy profile. The comparison between appliances is relatively easy for say the dishwasher versus the toaster or the dishwasher versus the kettle. It is more difficult and error prone to compare the dishwasher versus the washing machine. This condition likely allows for more learning than the aggregated condition, but still the visual system is not optimized to summarize the ups and downs of say the fridge and to compare them to the sum of the dishwasher's humps.



Figure 4. The disaggregated time series line graph visualization.

As explained above, line graphs are the natural choice for energy data because they show power \times time = energy; but they convey information on how much cumulative energy was used quite poorly. The practical question is how to show how much energy is consumed by different appliances in the

home. We therefore simplified the energy data into an easier-to-read visualisation so that the information would be more readily available and might facilitate encoding and retention [32].

The simplest way to transform the disaggregated line graphs into an area-based shape would be to maintain the time information and to eliminate the power fluctuations by using the average power as the y-value. However, this would result in a disorderly visual presentation with rectangles that vary greatly in height and width (e.g., the fridge would be 24 hours on the x-axis and around 100W on the y-axis, the kettle would be a couple of minutes on the x-axis and a couple thousand Watts on the y-axis, the light would be barely visible on the same scale). Therefore, we have opted to transform the area of *average use time x average power* into a near square-shaped rectangle. This approach also allowed us to impose pictographic appliance icons onto the shape.

Two-dimensional boxes have been found to be suitable for quick cognitive processing [7] and so have pictographic representations (simplified icons that are highly recognizable) where a picture of the represented item is used as a chart itself [32]. In the third condition, participants were presented with this area-based visualization, which showed the cumulative energy consumed by different appliances during a particular usage period. The individual appliances were represented by two-dimensional boxes (Figure 5). The size of these boxes was proportional to the energy consumed by the appliance. For example, if one appliance consumes twice as much energy as another, the box of one will be twice as big in area as the other. If Jack used the same appliance twice in a day, then the visualization will show two boxes for the appliance. We refer to this condition as the ‘summarized’ condition.

The idea behind the area-based summary visualization is to make it easier to distinguish how much electricity each appliance is consuming every time it is used. For example, the kettle and the toaster consume a lot of electricity for a few minutes only, using a lot of power per unit of time, whereas the TV consumes less power, but will be running for a much longer period. Line graph visualizations show energy consumption as a function of time, which presumably makes it difficult to estimate the

cumulative energy usage of a given appliance over time (i.e., calculating the total area under the line). The summarized visualization alleviates this by showing cumulative consumption over one usage of the appliance in an area-based representation which is easier to process for the human visual system. As discussed before, this representation is similar to the Practice View developed and deployed with the FigureEnergy system [17].

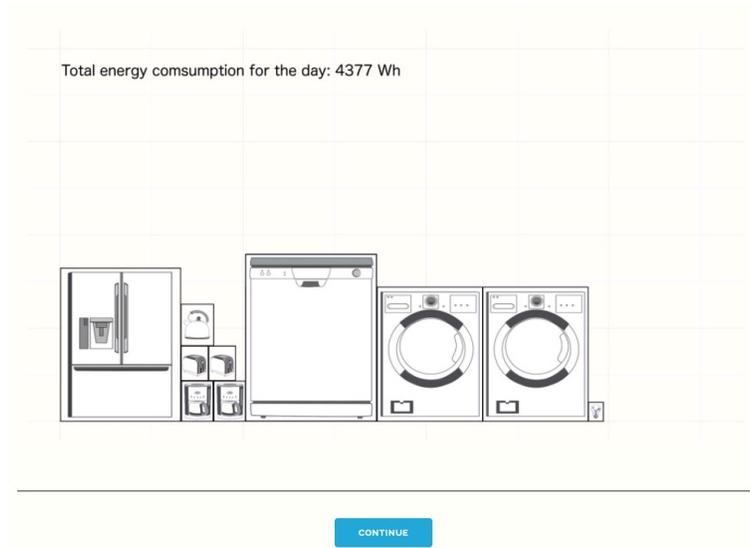


Figure 5. The summarized area-based visualization.

The energy data that was presented was taken from the UK-DALE dataset [40]. This open-access dataset is from a study that recorded domestic appliance-level electricity at a sample rate of 1/6 Hz from five UK houses, with the longest recording lasting 655 days in one house. We used the following eleven appliances from this house (house 1 in the dataset, a London end-of-terrace house, built c.1905): radio, lamp, microwave, toaster, kettle, coffeemaker, TV, vacuum cleaner, washing machine, dishwasher, and fridge. We modelled the appliances' energy consumption by identifying the typical duration of use and the power usage over time.

To assess the effectiveness of each of these visualizations, participants completed an energy game and answered open-ended questions. In the energy game, participants were asked to decide which of two appliances used the most electricity. This task was adapted from the ENLITEN energy game,

developed by the Universities Bath and Oxford as part of the ENLITEN project [47]. It has previously been used to measure changes in electricity consumption knowledge [36]. In the current study, participants made two-alternative forced-choice decisions, indicating which of two appliances consumed more electricity during a standard usage cycle (e.g., running the dishwasher or making coffee, see Figure 6). Participants were required to make 55 pairwise comparisons – this covers each combination of the eleven appliances taken from the UK-DALE dataset. For each decision, we recorded response accuracy and response time in seconds. Participants were further asked to indicate on a 5-point Likert scale how confident they were about the decision (1 being low confidence, 5 being high confidence).

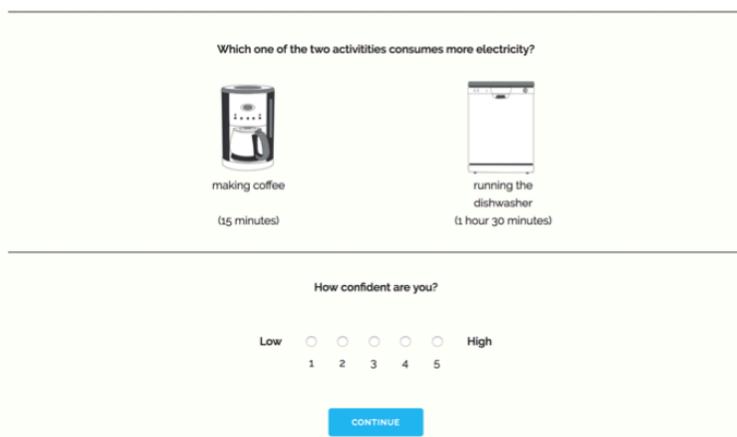


Figure 6. The energy game was played as pre- and post-test.

Some of the comparisons were relatively easy (e.g., dishwasher versus radio), whereas others were more difficult (e.g., dishwasher versus washing machine). The range of difficulty in the comparisons generated sufficient variance in participants' performance to assess differences between the three visualization conditions in post-test performance. The energy game targeted the understanding of how much energy is consumed by appliances per use. We further asked participants open-ended questions to better understand their cognitive processes during the experiment. The questions were as follows:

1. Before watching the simulation, you had to estimate which of two appliances consumed the most electrical energy. Were there any strategies that you used to make these decisions?

2. How did you make sense of the information visualization in the simulation?
3. What did you learn from the simulation?
4. How could Jack reduce his energy consumption?
5. Can you explain why the appliances use the amount of energy they consume?

The first question was to assess how much participants knew about the consumption of specific appliances beyond the measure of the energy game. The second, third and fourth question inquired how they made sense of the visualizations and what they learned from them. The fifth question was aimed especially at the line graph conditions. We previously [36] found that line graphs, with their ups and downs, triggered participants to reflect more deeply on what the appliances were doing throughout a usage cycle. The added cognitive effort may have benefits for encoding and recall [43]. We therefore assume that participants in the line graph conditions will give richer answers, because they require more cognitive effort than the area-based pictographs. The entire experiment was presented on a 27-inch iMac (2560 × 1440, graphics: ATI Radeon HD 4850 512 MB).

3.4 Procedure

The entire experiment lasted a little less than 30 minutes. It was conducted in a small quiet office, free from external interruptions and distractions, with a desktop computer placed on a table. Participants were randomly assigned to one of the three data visualization conditions. In each condition, they completed a short demographic questionnaire.

The experiment had three main parts. First, participants completed the energy game to generate their pre-test scores. Participants were not given any feedback on the accuracy of their responses. Second, participants were exposed to the simulation with the electricity data visualizations showing energy data over seven days. They were instructed to pay close attention and to try to learn how much

electricity the appliances consume. There was no time limit to how long each day was presented. Participants were asked to look at the data feedback and take as much time as they needed to learn how much the appliances consume. Once they felt they had completed one frame, they could click a ‘continue’ button and move on to the next frame, showing the next day. Participants could only look at the presented data, they could not actively interact with it or manipulate it other than moving forward to the next frame. Third, participants completed the energy game again. Completing it a second time allowed us to compare participants’ post-test performance with pre-test performance to determine whether there was any change in their understanding of electricity consumption after having been exposed to one of the data visualizations. After having completed the experiment, participants were presented with the five open-ended questions.

4. RESULTS

4.1 Quantitative data

Data from three participants was excluded from the quantitative data analysis because these participants said that they clicked through the simulation without trying to learn from it. We used a between-subjects analysis of variance (ANOVA) to compare responses on the energy game between the three different visualization conditions (aggregated, disaggregated, and summarized). We used a within-subjects analysis of variance (ANOVA) to compare responses between pre- and post-test. There were three dependent measures of interest: response accuracy (i.e., the proportion of correct decisions out of the 55 pair-wise comparisons in percentages), response time (in seconds), and response confidence (on a scale from 1 to 5). A significance level of 0.05 was used for judging the significance of effects.

We first checked to see whether there were any differences in participants’ baseline knowledge of domestic electricity consumption between the different conditions (as measured by pre-test responses to the energy game). All three groups had a similar level of response accuracy ($M = 75.15\%$, $SD = 11.68\%$, $M = 71.82\%$, $SD = 13.05\%$, and $M = 69.51\%$, $SD = 13.11\%$ for aggregated, disaggregated, and

summarized, respectively). Likewise, participants in each group reported similar pre-test response confidence levels ($M = 3.5$, $SD = .71$, $M = 3.78$, $SD = .59$, and $M = 3.55$, $SD = .59$, for aggregated, disaggregated, and summarized, respectively). Finally, each group had similar pre-test response times ($M = 7.97s$, $SD = 2.76s$, $M = 7.85s$, $SD = 2.3s$, $M = 8.32s$, $SD = 2.19s$, for aggregated, disaggregated, and summarized, respectively). Statistical analysis confirmed that there were no significant differences in baseline (pre-test) knowledge between participants that were placed in each of the three different data visualization conditions (all p 's > .05).

Given that all participants started out with a similar level of understanding of energy consumption, we next examine whether any one of the data visualizations helped improve this situation. To do this, we compared post-test responses to the energy game between participants assigned to each of the different data visualization conditions. As can be seen in Figure 7, participants had higher response accuracy in the summarized condition ($M = 87.36\%$, $SD = 12.71\%$) than participants in the aggregated condition ($M = 69.53\%$, $SD = 12.15\%$) or participants in the disaggregated condition ($M = 72.47\%$, $SD = 12.81\%$). Statistical analysis revealed a significant effect of visualization condition on response accuracy, $F(1,63) = 21.09$, $p < .001$. Participants gave slightly lower confidence ratings for their responses in the summarized condition ($M = 4.07$, $SD = .49$) than participants in the aggregated condition ($M = 4.33$, $SD = .54$) or participants in the disaggregated condition ($M = 4.4$, $SD = .41$), $F(1,63) = 5.08$, $p < .05$. We found no significant difference in response times between conditions for aggregated ($M = 6.06s$, $SD = 1.37s$), disaggregated ($M = 6.34s$, $SD = 2.27s$), and summarized ($M = 5.69s$, $SD = 1.24s$), $F(1, 63) = .53$, $p = .47$.

Finally, we compared the pre-test and post-test scores within each condition. The summarized group significantly increased their accuracy score from pre- to post-test, $F(1,42) = 21.02$, $p < .001$. There was no change between the pre- and post-test accuracy for either the aggregated group, $F(1,40) = 2.34$, $p < .13$, or the disaggregated group, $F(1,42) = 0.03$, $p < .87$. In terms of confidence, all three groups

became significantly more confident in their responses from pre- to post-test ($F(1,40) = 8.93, p < .01$; $F(1,42) = 10.54, p < .01$; and $F(1,42) = 30.77, p < .001$ for aggregated, disaggregated, and summarized respectively). All three groups also became significantly faster in their responses in the post-test compared to the pre-test ($F(1,40) = 8.06, p < .01$; $F(1,42) = 4.83, p < .05$; and $F(1,42) = 23.96, p < .001$ for aggregated, disaggregated, and summarized respectively).

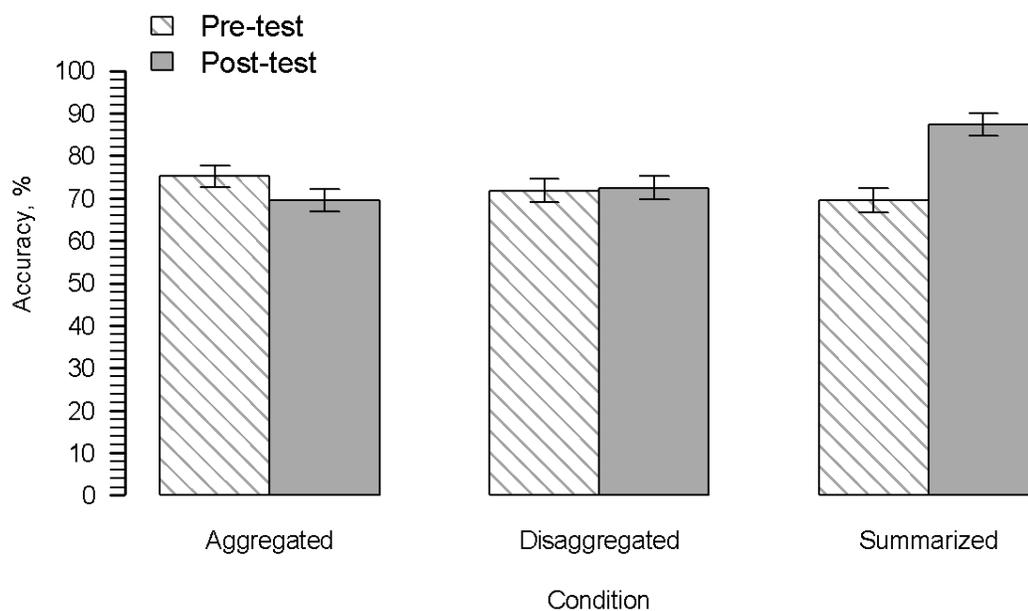


Figure 7. Pre- and post-test accuracy scores in the energy game for the three conditions aggregated, disaggregated and summarized. Error bars represent standard error of mean.

4.2 Qualitative data

4.2.1 *Before watching the simulation, you had to estimate which of two appliances consumed the most electrical energy. Were there any strategies that you used to make these decisions?*

The first open-ended question asked participants about the strategies that they had used to make their responses when completing the pre-test (baseline) energy game. Across the three groups, the same

themes emerged. Many participants considered the time the appliances were used for (as specified in the game) to inform their decisions. Also, many participants considered the power that they thought the appliance used. To estimate this, they thought about the size of the appliances and its functionality (e.g., whether appliances involved heating elements or generated kinetic energy). Prior knowledge and experience were named, too, such as information they had learned as a child or from energy ratings and information booklets. Few participants explicitly mentioned attempting to formally estimate energy as a function of power consumed over time.

4.2.2 *How did you make sense of the information visualization in the simulation?*

The second question asked participants about how they had used the information visualization to try and make sense of the illustrated domestic electricity dataset. Participants in the aggregated group reported looking at the spikes in the graph, understanding that *‘the higher up the red line, the more energy was used by an object’*. However, participants in this condition had to make guesses as to how long an appliance was on for. This might have been relatively easy to guess correctly for appliances with short duration and one high peak (e.g., the kettle), but is substantially harder for appliances that have longer durations and many peaks (e.g., the dishwasher). One participant said they looked at the annotations and *‘followed the spikes along the x-axis to deduce how long the appliance has been running for’*. Another participant stated: *‘I was not able to understand how long each item was used separately’*. Other than determining the duration of use per appliance, there was a second difficulty, which one participant described as follows: *‘I tried to compare the usage over time but found this difficult because some appliances used small amounts of energy but over a longer period of time’*.

Participants in the disaggregated group could clearly see the duration. They reported *‘to take into consideration both axes: time and energy. So, bearing in mind that even if an appliance consumes much more energy per hour, it may only be on for a short period of time’*, and they *‘compared appliances with each other and memorized relationships, kettle > coffee maker; washer > microwave, for example’*. The

idea that the area under the curve represented the energy consumed was understood by participants: *‘For the vertical axis, the higher the more it will consume per hour. Multiply it with hours. The result is the energy it has used. So, the more total space it has, the more energy it has consumed’*. Still the second challenge that the aggregated group encountered, occurred in the disaggregated group, too: *‘it was a lot harder to compare the refrigerator, for example, as it had relatively low consumption per hour, but it was turned on for many hours’* and another explained *‘It was very easy to compare high-consuming appliances with low-consuming appliances, and come up with a hierarchy of consumption for high versus low consuming appliances. However, it was harder to compare low-consuming appliances with other low-consuming appliances as the spaces on the graph were so small that it was difficult to visualize the difference between them’*.

Participants in the summarized group only looked *‘at the sizes of the boxes relative to each other’*. The duration and power information were eliminated from the data visualization, and the cognitive task was reduced to memorizing hierarchies and recalling relationships such as that the dishwasher, washing machine and fridge *‘took up most energy’* and that the *‘radio used least energy’*. One participant described grouping large appliances together and smaller appliances together, and then within one category compared appliances to each other to memorize their ranking. Several participants remarked that the visualization was clear and understandable, but one participant in the group misinterpreted the boxes thinking they showed how much energy the appliances consume per hour, whereas all others understood that we displayed the typical or average duration of use of an appliance which matched *‘the amount of time used in the questions’* from the energy game.

4.2.3 *What did you learn from the simulation?*

The third question asked participants what they had learnt from looking at the information visualization used to make sense of the illustrated domestic electricity dataset. Across conditions, several participants reported learning that certain appliances (i.e., dishwasher, washing machine, vacuum cleaner, kettle,

coffee maker, microwave, and toaster) consume more than they expected. Equally, they reported surprise as to how little radio, lights and TV consume. In the aggregated condition, participants spoke mostly about peaks (*'I was surprised how much of a power spike there is for kettles'*) and they used the terms energy and power interchangeably. Only one participant in this group considers that kettle and toaster are high in power, but only used for a very short time. None of the participants mentioned the fridge.

In the disaggregated condition, the fridge generated controversial statements: several participants in the disaggregated condition listed the fridge as one of the appliances that consumed surprisingly little energy. Only one participant in this condition noted that the fridge would consume *'a lot because it is always on'* and another inferred there would always be a base rate due to the fridge. It seemed that some participants focused more on the power dimension than on the duration: *'I never thought that the fridge would consume so little electricity per hour, and I never knew that the kettle would consume that high amount of energy'*. Another participant in contrast summarized *'we should consider the power of the appliance and the length of time it works together to get the power [energy] consumption'*. Learnings were that *'more time used'* does not necessarily equal *'more energy used'*, and that *'some appliances do not have a stable consumption, but it changes during the cycle of use'*.

In contrast to the disaggregated group, participants in the summarized condition consistently stated that the fridge uses a lot and contributes *'greatly towards an energy bill'*. The participant who realized the fridge's contribution towards the bill said: *'I have always been conscious as a bill payer about smaller things like charging phones, using lights etc.'* A couple of participants in the summarized condition further mentioned that they *'learned how much energy each appliance uses relative to other appliances'*, for example, that *'making a cup of tea was more energy consuming than listening to radio for an hour'* and this participant mentioned that they found that surprising. The summarized group also slightly differed from the other conditions in their assessment of the kettle and toaster – generally these

were listed as appliances that were shown to consume more than participants had expected; one participant in the summarized condition however must have had prior knowledge that they are power-intensive and he focused on the fact that *'in total'* they use less than expected.

4.2.4 *How could Jack reduce his energy consumption?*

The fourth question asked participants to help Jack, the persona in the scenario, with reducing his energy consumption. Most responses across the three conditions drew strongly from previous knowledge and included generic recommendations such as using appliances less and more efficiently. To use appliances less, one of the most common suggestions was to wash the dishes by hand. Others were to switch the lights off or to have fewer cups of tea and not to keep the radio or TV on in the background. To use appliances more efficiently, they recommended fully loading the dishwasher and the washing machine and boiling only the amount of water needed in the kettle.

In the aggregated group, two participants said they didn't know how Jack could save energy. One requested itemized information because it would be *'helpful for Jack to be made more aware of how much energy simple household items, e.g. microwave, use up'*. In the disaggregated condition, too, one participant said he did not know because there was not enough information provided in the simulation, for example, the *'kind of washing cycle'* for dishwashers and washing machines would impact energy consumption.

In terms of referring to power and duration, there was one person in the aggregated group who suggested reducing the *'using time of those with high electricity power, as these will have a significant increase on the whole electricity consumption'*. In the disaggregation group, too, some participants made the general recommendation to reduce the use of appliances with 'high power' or 'high energy'. A couple specifically named kettle, toaster, dishwasher and microwave amongst these high consumers.

A subtle difference emerged in the summarized group: two participants went beyond the generic suggestion to use energy intensive appliances less, and they specifically named the dishwasher (the dishwasher was indeed the biggest consumer of all appliances in the presented data).

4.2.5 *Can you explain why the appliances use the amount of energy they consume?*

Most participants across the conditions referred to whether an appliance is generating heat or kinetic energy as reasons why an appliance uses a lot of energy. In less technical terms, they also mentioned that the dishwasher needs power to *'clean and dry'*, the Hoover needs a lot of energy *'to suck in air'*, a radio consumes little because it only outputs sound, and lights do not have *'such an intensive task'*.

There was one theme that only came up in the disaggregated condition: A few participants referred to whether appliances need a *'boost'* or not, for example the kettle, the Hoover and the dishwasher do, whereas the fridge does not have a boost but it's *'on throughout the day'*. Another participant explained in a similar way that the heating element in a kettle *'causes a huge energy consumption in a short time'*, whereas the fridge is *'keeping the temperature inside instead of changing it'*. Similarly, yet another participant said: *'a light bulb just needs a little power to keep it on'*.

Again, the fridge came up as a contentious item: one participant in the aggregated group said it consumed *'very little energy per hour'* but *'across the day [it] adds up to more than most appliances'*. For one participant in the disaggregated group, the framing was exactly the other way around, they said the fridge *'consumed less energy although it is used 24/7'* and another said a fridge was *'designed (...) to be kept on all the time'* and *'a key feature of it was that it needed to minimize electrical consumption'*. A second participant in the disaggregated group used the same logic saying: *'Fridge need[s] to be opened [on] all day, so the consumption must be relatively lower'*. In contrast, a participant in the summarized group reiterated that *'The fridge has to be turned on for 24 hours, so it consumes a lot of electricity'*.

5. DISCUSSION

5.1 Quantitative findings

There are numerous barriers that prevent householders from the effective understanding and interpretation of their domestic energy data. They may not understand, or be interested in, the relationship of power and duration. They may not be good at reading data visualizations. They may not wish to investigate their consumption patterns, or change their habits, whatsoever. All these factors limit the willingness and effort to learn from domestic energy feedback. These concerns are not controlled for in this study, which only looks at how visualization choices can address the unresolved ‘groceries bill problem’. The results show that indeed participants’ performance in the energy game was affected by the kind of visualization they were exposed to and hence confirm that what householders are likely to learn, depends greatly on how the data is visualized [14,36,63].

Most importantly, given the ongoing research on smart feedback and energy data disaggregation, it shows that disaggregation is necessary, but not sufficient. Participants in the summarized condition gained a more accurate understanding of how much electricity different domestic appliances were using (in terms of post-test responses), compared to participants who were shown time series data. Similar to previous research [36], we also found that participants who were shown disaggregated energy data in a line graph developed no better of an understanding of how much electricity different domestic appliances were using than participants who were shown a simpler aggregated line graph. This shows that disaggregation alone does not help householders to identify where they are using the most energy, but the choice of information visualization is critical: it is crucial to properly couple the level and type of data aggregation with the information needed to succeed at any given task.

The summarized condition facilitated learning about how much electricity appliances consume per typical use relative to each other. In the disaggregated time series condition, this comparison was still too difficult, as participants must work out how much energy is consumed by an appliance over

time. Estimating the cumulative energy consumed, that is, visually gaging the area under the curve is a challenging cognitive task. In general, people are very good at detecting deviations from the horizontal (i.e., to process the obvious peaks), but not at integrating power over time (i.e., the area under each color-coded line graph). As a result of this, people tend to overly focus on peaks in line graphs, neglect the baseline and trends [31,35,61].

The quantitative findings regarding the accuracy measure can be explained by existing design knowledge from the data visualization field. It may seem obvious in hindsight that the area-charts are easier to learn from than time-series graphs for the specific experimental task that was tested. The comparison between line charts and area-based graphs is biased in so far, that the area-based charts have undergone further transformations: they literally eliminate the need to eyeball the energy depicted under an oscillating curve. The summarized condition is therefore specifically designed and predicted to achieve the highest increase in the accuracy scores. The line graphs, in contrast, are by design less helpful to answer the comparative questions. One might even argue that a ranked list, without any visualization, would have been a fairer comparison. The findings are therefore not surprising from a visualization research point of view.

However, they have bearing on the energy feedback industry, which has not sufficiently picked up on the existing data visualization knowledge: Industrial products use brute line graphs, that are merely more than a depiction of the raw data. In fact, they require willingness and literacy from an untrained user, both prerequisites that householders won't necessarily meet. HCI research often takes place at the other end of the spectrum, venturing towards artistic design, investigating ambient and creative visualizations that do not require any previous expertise, or maximum concentration.

The groceries bill problem is still considered to be meaningful but has remained unresolved. FigureEnergy is a rare example of a feedback tool that provides a pragmatic visualization, that has attempted to address the groceries bill problem, and to show the householder an easy to grasp

breakdown of their total consumption. This specific experiment addresses only the question whether an area-based visualization, like the one in FigureEnergy, is effective in teaching users about their practice-specific energy consumption.

We expected that with higher accuracy, participants would also be more confident and give quicker responses if they felt they knew the answer, as opposed to hesitating if they were uncertain. However, all three groups increased their confidence score and decreased their response time from pre- to post-test. Given that participants in the aggregated and disaggregated condition did not improve on accuracy, it seems that becoming faster and more confident could merely be an effect of repeating the energy test; we discuss other possible reasons for this in section 5.2.

5.2 Qualitative findings

5.2.1 A priori heuristics

Part of the findings from the qualitative data are in line with previous studies [31,35] and help to explain the quantitative performance data in the first energy game: people go by heuristics and prior knowledge in their assessment which activity consumes more energy. Few consider energy consumed as the result of power demand over time.

5.2.2 Making sense of the visualizations

In the simulation, they learn that some of their decisions are biased by flawed heuristics [4]. As expected, challenges occur in the disaggregated condition in working out power over time, and in the aggregated condition even more so as participants focus on the amount of power but can only make guesses about duration. The comparisons are particularly difficult for appliances that do not differ greatly in the energy they consume, whereas participants find it feasible to work out the hierarchy between appliances with sufficient differences.

5.2.3 Subjective learning experience

It is interesting that several statements in the qualitative data show that participants *think* they learn from the line graphs. It seems they overestimate their ability to work out the duration in the aggregated condition and the respective area under both types of line graphs. This might explain the increase in confidence and decrease in response time: participants subjectively report learning from the visualization, but this does not result in any improvements in accuracy. In contrast, participants in the summarized condition have an improvement in both subjective confidence as well as the accuracy of their responses.

5.2.4 Actionable tips for Jack

We asked this question to see if participants would pick up on idiosyncrasies of Jack's household, and if they would give recommendations that demonstrate clearly that an energy saving tip is based on Jack's personal data. Most of the responses to the question of how Jack could save energy are generic, and similar between the conditions. Only in the summarized condition did one participant recommend to use the dishwasher less, explicitly referring to the data showing that the dishwasher is Jack's top consumer.

5.2.5 Inferences about appliances

Particularly striking are the responses relating to the fridge. Participants in the summarized condition correctly identify the fridge as an appliance that consumers a lot, explaining this by saying it is on constantly. This explanation is likely based on prior knowledge (because the time information was not provided in the visualization), but it seems that the size of the fridge stuck with participants. One participant says he used to be very conscious of the consumption of small appliances such as lights or charging his phone, and he now realizes he misplaced his efforts, identifying that it's the big consumers like the fridge that contribute most to the energy bill. This links back to the grocery bill metaphor [42] and the need for disaggregation, that is, the itemization of the energy bill for householders to learn where they are spending most of their energy and hence, where to tackle the expense and economize.

For the quantitative findings, we were expecting to find the best results for the summarized condition, because the area-based visualization was designed to simply convey information on how much energy an appliance used. But based on our previous experiment [36], we expected to find shortcomings of the summarized condition and an advantage of the line graphs in the *qualitative* data – the previous study found that the shape of the line graphs triggered people to think about what the appliances are doing. We cannot necessarily confirm that this is the case for our participants (we had only one participant observing that appliances consumption is not stable throughout but subject to fluctuations), but we certainly acknowledge that the summarized condition takes information about energy patterns and any time-related information away which is relevant for other learnings but knowing which appliance consumes most energy.

5.3 Discussion across qualitative and quantitative findings

We cannot say for certain why participants increased their confidence levels and decreased the response speed, independent of their variation in response accuracy. Participants erroneously believe they learn from the line graph conditions. Both the quantitative increase in confidence, as well as some of the qualitative comments indicate that they are unaware of how poorly they interpret the graphs. The example of the fridge shows how participants fall victim to their visual and cognitive biases, i.e. in the line graph conditions, participants neglect the impact of a low, but continuous baseline. The mismatch between increased confidence and unchanged accuracy, is reason for concern when it comes to showing householders energy time series data in the form of line graphs. Assuming householders need to know the cost breakdown of their total bill, the raw data requires further transformation into a visualization that is specifically designed to communicate this information.

5.4 Limitations

There are several limitations to the study. First, our sample is not representative of the general population with regards to gender, age, and education and these demographics might have an impact on

people's abilities to understand infographics [46,59]. This compromises the generalizability of this study. Ideally, we would like to replicate the study with a more heterogeneous sample showing a wider variation in age and education and a balanced gender distribution. However, our sample consisted of many university students, which means they were highly educated, relatively young, and probably highly computer literate. If they were unable to extract the relevant information from the line graphs, there is little reason to assume that the general population would perform differently. The fact that our results are consistent with one previous experiment [36] also supports the validity of our findings.

Second, laboratory experiments are naturally somewhat artificial and so there is always a risk when generalizing findings to more naturalistic contexts of use [55]. One limitation to ecological validity is that our study relied on a recall task, whereas householders often do not have to memorize and recall the information they see to answer questions, but they can look and reflect on it at the same time. We chose to replicate the method from [36], but for further replications it would be interesting to change the study's design and allow participants to answer test questions while engaging with the feedback data. That being said, it is unlikely that householders in their home would approach smart feedback in such a systematic way, trying to answer questions without being prompted as part of an intervention. One might argue that a recall task is relevant because what matters is what householders do when they walk away from the energy display; if they were able to easily memorize the biggest contributors in their homes from looking at pictographic information, they might be able to subsequently recall this and change relevant behavior.

Related to the validity of the recall task, we measured the response time for the decision making in the forced-choice task and we did not find this measure to be meaningful. An improvement would be to measure the time participants spend looking at the data visualization instead. The other limitation to ecological validity is that users saw someone else's data which is certainly less meaningful than reflecting on one's own data. Other authors have found that in longer-term studies of eco-feedback using

time-based representations participants could eventually identify individual appliances in their home [11,52]. These are noteworthy limitations, but what experiments lack in ecological validity, they make up for with strong internal validity and provide a useful upper-bound on people's ability. We think experimental simulations like this are a rigorous approach to test cognitive processes free from confounding variables [27].

The third limitation is that our operationalization was one very specific task (i.e., making decisions in the energy game). We focused on one aspect of energy literacy as participants' knowledge about how much energy the appliances consume. We do not claim that learning the breakdown of a household's energy bill is the only valuable insight from smart energy usage data. In a naturalistic home setting, people might be interested in learning various other things about their domestic energy consumption (e.g., which appliances are inefficient and eligible for retrofitting, or at what time of day they are using most energy if they are on a time-of-use tariff). Our method was too limited in that neither the energy game nor our open-ended questions created sufficient opportunity to demonstrate time-centric learnings that were favored by the line graph conditions. It needs to be further investigated how time-sensitive information (e.g., time-of-use tariffs) are best communicated to the user. Feedback showing trends over the day (like line graphs do) will be indispensable for that kind of scenario. Possible measures for data comprehension could be tasks that require participants to determine the best time of day to run certain household appliances. Ideally, they would combine the knowledge of which appliances consume most energy (appliance-centric) and when it would be sensible to run them (time-centric).

Finally, the experiment only considers the understanding of energy usage information. The results of the study do not prove that increased knowledge results in behavior change and energy savings. More work is needed to determine how users can be encouraged to change practices based on their improved knowledge.

6. CONCLUSION

HCI technologies are aiming to nudge users into optimizing their behavior, such as reducing and shifting their energy consumption. People need to understand which everyday practices consume the most energy to make significant savings. The purpose of this study was to provide empirical evidence for an applied solution to residential energy feedback that is as simple and easy to understand as possible for an untrained audience (i.e., householders). For this purpose – knowing how much energy everyday practices consume – we’ve shown that simplified area-based graphs are preferable to line graphs, which are commonly used to display energy data. It also provided evidence that appliance-wise disaggregation is essential for home energy feedback. This finding is an important implication which should be considered in the design of future Home Energy Management Systems.

ACKNOWLEDGMENTS

This work is supported by an EPSRC Industrial CASE Studentship (EP/K504610/1) and EDF Energy R&D UK.

REFERENCES

1. Pedro Álvarez and Pedro Vega. 2009. Actitudes ambientales y conductas sostenibles. Implicaciones para la educación ambiental. *Revista de Psicodidáctica* 14, 2. Retrieved September 14, 2017 from <http://www.redalyc.org/html/175/17512724006/>
2. Will Anderson and Vicki White. 2009. Exploring consumer preferences for home energy display functionality. *Centre for Sustainable Energy, Tech. Rep.* Retrieved September 15, 2017 from http://www.cse.org.uk/app/webroot/downloads/file/consumer_preferences_for_home_energy_display.pdf

3. K. Carrie Armel, Abhay Gupta, Gireesh Shrimali, and Adrian Albert. 2013. Is disaggregation the holy grail of energy efficiency? The case of electricity. *Energy Policy* 52, Supplement C: 213–234. <https://doi.org/10.1016/j.enpol.2012.08.062>
4. S. Z. Attari, M. L. DeKay, C. I. Davidson, and W. Bruine de Bruin. 2010. Public perceptions of energy consumption and savings. *Proceedings of the National Academy of Sciences* 107, 37: 16054–16059. <https://doi.org/10.1073/pnas.1001509107>
5. W. G. V. Balchin and Alice M. Coleman. 1966. Graphicacy should be the fourth ace in the pack. *Cartographica: The International Journal for Geographic Information and Geovisualization* 3, 1: 23–28. <https://doi.org/10.3138/C7Q0-MM01-6161-7315>
6. Nipun Batra, Jack Kelly, Oliver Parson, Haimonti Dutta, William Knottenbelt, Alex Rogers, Amarjeet Singh, and Mani Srivastava. 2014. NILMTK: an open source toolkit for non-intrusive load monitoring. In *Proceedings of the 5th international conference on Future energy systems*, 265–276. Retrieved September 14, 2017 from <http://dl.acm.org/citation.cfm?id=2602051>
7. Judith Borghouts, Frank Soboczenski, Paul Cairns, and Duncan Brumby. 2015. Visualizing Magnitude: Graphical Number Representations Help Users Detect Large Number Entry Errors. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 59, 1: 591–595. <https://doi.org/10.1177/1541931215591130>
8. J. Boy, R. A. Rensink, E. Bertini, and J. D. Fekete. 2014. A Principled Way of Assessing Visualization Literacy. *IEEE Transactions on Visualization and Computer Graphics* 20, 12: 1963–1972. <https://doi.org/10.1109/TVCG.2014.2346984>
9. Robert S. Brewer, Yongwen Xu, George E. Lee, Michelle Katchuck, Carleton A. Moore, and Philip M. Johnson. 2013. Energy feedback for smart grid consumers: Lessons learned from the Kukui Cup. *Proceedings of Energy*: 120–126.

10. Robert Stephen Brewer. 2013. Fostering sustained energy behavior change and increasing energy literacy in a student housing energy challenge. University of Hawai'i at Manoa. Retrieved September 16, 2017 from <http://search.proquest.com/openview/720e98bff0da2a163f0f9d135f2ed24c/1?pq-origsite=gscholar&cbl=18750&diss=y>
11. Loove Broms, Cecilia Katzeff, Magnus Bång, Åsa Nyblom, Sara Ilstedt Hjelm, and Karin Ehrnberger. 2010. Coffee maker patterns and the design of energy feedback artefacts. 93. <https://doi.org/10.1145/1858171.1858191>
12. Peter C. -H Cheng. 1999. Unlocking conceptual learning in mathematics and science with effective representational systems. *Computers & Education* 33, 2: 109–130. [https://doi.org/10.1016/S0360-1315\(99\)00028-7](https://doi.org/10.1016/S0360-1315(99)00028-7)
13. Marshini Chetty, David Tran, and Rebecca E. Grinter. 2008. Getting to green: understanding resource consumption in the home. In *Proceedings of the 10th international conference on Ubiquitous computing*, 242–251. Retrieved September 14, 2017 from <http://dl.acm.org/citation.cfm?id=1409668>
14. Teresa Chiang, Sukumar Natarajan, and Ian Walker. 2012. A laboratory test of the efficacy of energy display interface design. *Energy and Buildings* 55, Supplement C: 471–480. <https://doi.org/10.1016/j.enbuild.2012.07.026>
15. Yoram Chisik. 2011. An image of electricity: towards an understanding of how people perceive electricity. *Human-Computer Interaction–INTERACT 2011*: 100–117.
16. William S. Cleveland and Robert McGill. 1984. Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. *Journal of the American Statistical Association* 79, 387: 531–554. <https://doi.org/10.1080/01621459.1984.10478080>

17. Enrico Costanza, Sarvapali D. Ramchurn, and Nicholas R. Jennings. 2012. Understanding Domestic Energy Consumption Through Interactive Visualisation: A Field Study. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing (UbiComp '12)*, 216–225. <https://doi.org/10.1145/2370216.2370251>
18. Sarah Darby. 2001. Making it obvious: designing feedback into energy consumption. In *Energy efficiency in household appliances and lighting*. Springer, 685–696. Retrieved September 14, 2017 from https://link.springer.com/chapter/10.1007/978-3-642-56531-1_73
19. Jan E. DeWaters and Susan E. Powers. 2011. Energy literacy of secondary students in New York State (USA): A measure of knowledge, affect, and behavior. *Energy Policy* 39, 3: 1699–1710. <https://doi.org/10.1016/j.enpol.2010.12.049>
20. Corinna Fischer. 2008. Feedback on household electricity consumption: a tool for saving energy? *Energy Efficiency* 1, 1: 79–104. <https://doi.org/10.1007/s12053-008-9009-7>
21. Derek Foster, Shaun Lawson, Mark Blythe, and Paul Cairns. 2010. Wattsup?: motivating reductions in domestic energy consumption using social networks. 178. <https://doi.org/10.1145/1868914.1868938>
22. Jon Froehlich, Leah Findlater, and James Landay. 2010. The design of eco-feedback technology. In *Proceedings of the SIGCHI conference on human factors in computing systems*, 1999–2008. Retrieved September 14, 2017 from <http://dl.acm.org/citation.cfm?id=1753629>
23. Jon Froehlich, E. Larson, S. Gupta, G. Cohn, M. Reynolds, and S. Patel. 2011. Disaggregated End-Use Energy Sensing for the Smart Grid. *IEEE Pervasive Computing* 10, 1: 28–39. <https://doi.org/10.1109/MPRV.2010.74>

24. Edward Fry. 1981. Graphical Literacy. *Journal of Reading* 24, 5: 383–389. <https://doi.org/10.2307/40032373>
25. Mirta Galesic and Rocio Garcia-Retamero. 2011. Graph Literacy: A Cross-Cultural Comparison. *Medical Decision Making* 31, 3: 444–457. <https://doi.org/10.1177/0272989X10373805>
26. Hugo Goncalves, Adrian Ocneanu, Mario Berges, and R. H. Fan. 2011. Unsupervised disaggregation of appliances using aggregated consumption data. In *The 1st KDD Workshop on Data Mining Applications in Sustainability (SustKDD)*. Retrieved September 14, 2017 from <https://pdfs.semanticscholar.org/5531/e05c2c76b150473c218a749bd83372d691e5.pdf>
27. Cleotilde Gonzalez, Rickey P. Thomas, and Polina Vanyukov. 2005. The relationships between cognitive ability and dynamic decision making. *Intelligence* 33, 2: 169–186. <https://doi.org/10.1016/j.intell.2004.10.002>
28. Anton Gustafsson and Magnus Gyllenswärd. 2005. The power-aware cord: energy awareness through ambient information display. 1423. <https://doi.org/10.1145/1056808.1056932>
29. Tom Hargreaves. 2011. Practice-ing behaviour change: Applying social practice theory to pro-environmental behaviour change. Retrieved July 18, 2018 from <http://journals.sagepub.com/doi/abs/10.1177/1469540510390500>
30. Tom Hargreaves. 2018. Beyond energy feedback. *Building Research & Information* 46, 3: 332–342. <https://doi.org/10.1080/09613218.2017.1356140>
31. Jordan Harold, Kenny R Coventry, Irene Lorenzoni, and Thomas F Shipley. Making Sense of Time-Series Data: How Language Can Help Identify Long-Term Trends. 6.
32. Steve Haroz, Robert Kosara, and Steven L. Franconeri. 2015. ISOTYPE Visualization: Working Memory, Performance, and Engagement with Pictographs. In *Proceedings of the 33rd Annual ACM*

- Conference on Human Factors in Computing Systems (CHI '15)*, 1191–1200.
<https://doi.org/10.1145/2702123.2702275>
33. George W. (Natick Hart, Jr Kern, and Fred C. (Carlisle Schweppe. 1989. *Non-Intrusive Appliance Monitor Apparatus*. MASSACHUSETTS INST OF TECH. Retrieved September 14, 2017 from <https://www.osti.gov/scitech/biblio/867077>
34. Jeffrey Heer and Michael Bostock. 2010. Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*, 203–212. <https://doi.org/10.1145/1753326.1753357>
35. Melanie R. Herrmann, Duncan P. Brumby, and Tadj Oreszczyn. 2017. Watts your usage? A field study of householders' literacy for residential electricity data. *Energy Efficiency*: 1–17. <https://doi.org/10.1007/s12053-017-9555-y>
36. Melanie R. Herrmann, Duncan P. Brumby, Tadj Oreszczyn, and Xavier M. P. Gilbert. 2017. Does data visualization affect users' understanding of electricity consumption? *Building Research & Information* 0, 0: 1–13. <https://doi.org/10.1080/09613218.2017.1356164>
37. Tiffany Holmes. Combining art and technology to reduce energy consumption. 188.
38. D. Huang, M. Tory, B. Adriel Aseniero, L. Bartram, S. Bateman, S. Carpendale, A. Tang, and R. Woodbury. 2015. Personal Visualization and Personal Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics* 21, 3: 420–433. <https://doi.org/10.1109/TVCG.2014.2359887>
39. W. Javed, B. McDonnel, and N. Elmqvist. 2010. Graphical Perception of Multiple Time Series. *IEEE Transactions on Visualization and Computer Graphics* 16, 6: 927–934. <https://doi.org/10.1109/TVCG.2010.162>

40. Jack Kelly and William Knottenbelt. 2015. The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes. *Scientific Data* 2: 150007. <https://doi.org/10.1038/sdata.2015.7>
41. Jack Kelly and William Knottenbelt. 2016. Does disaggregated electricity feedback reduce domestic electricity consumption? A systematic review of the literature. *arXiv:1605.00962 [cs]*. Retrieved September 14, 2017 from <http://arxiv.org/abs/1605.00962>
42. Willett Kempton and Linda L. Layne. 1994. The consumer's energy analysis environment. *Energy Policy* 22, 10: 857–866. [https://doi.org/10.1016/0301-4215\(94\)90145-7](https://doi.org/10.1016/0301-4215(94)90145-7)
43. Willett Kempton and Laura Montgomery. 1982. Folk quantification of energy. *Energy* 7, 10: 817–827. [https://doi.org/10.1016/0360-5442\(82\)90030-5](https://doi.org/10.1016/0360-5442(82)90030-5)
44. Tamar Krishnamurti, Alexander L. Davis, Gabrielle Wong-Parodi, Jack Wang, and Casey Canfield. 2013. Creating an in-home display: Experimental evidence and guidelines for design. *Applied Energy* 108: 448–458. <https://doi.org/10.1016/j.apenergy.2013.03.048>
45. S. Lee, S. H. Kim, Y. H. Hung, H. Lam, Y. A. Kang, and J. S. Yi. 2016. How do People Make Sense of Unfamiliar Visualizations?: A Grounded Model of Novice #x0027;s Information Visualization Sensemaking. *IEEE Transactions on Visualization and Computer Graphics* 22, 1: 499–508. <https://doi.org/10.1109/TVCG.2015.2467195>
46. Angela Locoro, Federico Cabitza, Rossana Actis-Grosso, and Carlo Batini. 2017. Static and interactive infographics in daily tasks: A value-in-use and quality of interaction user study. *Computers in Human Behavior* 71: 240–257. <https://doi.org/10.1016/j.chb.2017.01.032>
47. Tom Lovett, Elizabeth Gabe-Thomas, Sukumar Natarajan, Eamonn O'Neill, and Julian Padget. 2013. “Just Enough” Sensing to ENLITEN: A Preliminary Demonstration of Sensing Strategy for the

“Energy Literacy Through an Intelligent Home Energy Advisor” (ENLITEN) Project. In *Proceedings of the Fourth International Conference on Future Energy Systems (e-Energy '13)*, 279–280. <https://doi.org/10.1145/2487166.2487206>

48. L. T. McCalley and Cees J. H. Midden. 2002. Energy conservation through product-integrated feedback: The roles of goal-setting and social orientation. *Journal of Economic Psychology* 23, 5: 589–603. [https://doi.org/10.1016/S0167-4870\(02\)00119-8](https://doi.org/10.1016/S0167-4870(02)00119-8)

49. Barbara Mettler-Meibom and Bertold Wichmann. 1982. Informationsstand und Einstellung als Verhaltensregulative. In *Einfluß des Verbraucherverhaltens auf den Energiebedarf privater Haushalte*. Springer, Berlin, Heidelberg, 73–88. https://doi.org/10.1007/978-3-642-95404-7_6

50. David Peebles, Devina Ramduny-Ellis, Geoffrey Ellis, and John V.H. Bonner. 2013. The influence of graph schemas on the interpretation of unfamiliar diagrams.

51. Steven Pinker. 1990. A theory of graph comprehension. *Artificial intelligence and the future of testing*: 73–126.

52. Filipe Quintal, Lucas Pereira, Nuno J. Nunes, and Valentina Nisi. 2015. What-a-Watt: exploring electricity production literacy through a long term eco-feedback study. 1–6. <https://doi.org/10.1109/SustainIT.2015.7101365>

53. Simon Roberts and William Baker. 2003. Towards effective energy information. *Improving Consumer Feedback on Energy Consumption, Centre for Sustainable Energy, Bristol, UK*. Retrieved September 15, 2017 from https://www.ofgem.gov.uk/sites/default/files/docs/2003/07/4138-towards_effective_energy_information_cse_report_ofgem_july03.pdf

54. Johnny Rodgers and Lyn Bartram. 2011. Exploring Ambient and Artistic Visualization for Residential Energy Use Feedback. *IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS* 17, 12: 9.
55. Yvonne Rogers, N. Yuill, and Paul Marshall. 2013. Contrasting lab- based and in-the-wild studies for evaluating multi-user technologies. In *The Sage Handbook of Digital Technology Research*. London, UK. Retrieved September 15, 2017 from [https://books.google.co.uk/books?hl=en&lr=&id=vZjCAQAAQBAJ&oi=fnd&pg=PA359&dq=Rogers,+Y.,+Yuill,+N.,+%26+Marshall,+P.+\(2013\).+Contrasting+lab-+based+and+in-the-wild+studies+for+evaluating+multi-user+technologies.+The+Sage+Handbook+of+Digital+Technology+Research.+London,+UK:+Sage.&ots=PoltevITXr&sig=gNB82y5LG-Pt01Avfbqgt43grI4#v=onepage&q&f=false](https://books.google.co.uk/books?hl=en&lr=&id=vZjCAQAAQBAJ&oi=fnd&pg=PA359&dq=Rogers,+Y.,+Yuill,+N.,+%26+Marshall,+P.+(2013).+Contrasting+lab-+based+and+in-the-wild+studies+for+evaluating+multi-user+technologies.+The+Sage+Handbook+of+Digital+Technology+Research.+London,+UK:+Sage.&ots=PoltevITXr&sig=gNB82y5LG-Pt01Avfbqgt43grI4#v=onepage&q&f=false)
56. Rebecca Sokoloski. 2015. Disaggregated Electricity Consumption: Using Appliance-Specific Feedback to Promote Energy Conservation. Retrieved September 14, 2017 from <http://scholarworks.calstate.edu/handle/10211.3/141932>
57. L. Stankovic, V. Stankovic, J. Liao, and C. Wilson. 2016. Measuring the energy intensity of domestic activities from smart meter data. *Applied Energy* 183: 1565–1580. <https://doi.org/10.1016/j.apenergy.2016.09.087>
58. Yolande Strengers. 2011. Designing eco-feedback systems for everyday life. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2135–2144. Retrieved September 14, 2017 from <http://dl.acm.org/citation.cfm?id=1979252>
59. Christian Sturm, Alice Oh, Sebastian Linxen, Jose Abdelnour Nocera, Susan Dray, and Katharina Reinecke. 2015. How WEIRD is HCI?: Extending HCI Principles to Other Countries and Cultures. In

Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '15), 2425–2428. <https://doi.org/10.1145/2702613.2702656>

60. Jillian C. Sweeney, Johannes Kresling, Dave Webb, Geoffrey N. Soutar, and Tim Mazarol. 2013. Energy saving behaviours: Development of a practice-based model. *Energy Policy* 61: 371–381. <https://doi.org/10.1016/j.enpol.2013.06.121>

61. E. R. Tufte. 1983. *The visual display of quantitative information*. Graphics Press, Connecticut, CT.

62. Rayoung Yang, Mark W. Newman, and Jodi Forlizzi. 2014. Making sustainability sustainable: challenges in the design of eco-interaction technologies. 823–832. <https://doi.org/10.1145/2556288.2557380>

63. Tae-Jung Yun, Hee Young Jeong, Peter Kinnaird, SeungEok Choi, Namwook Kang, and Gregory D. Abowd. 2010. *Domestic Energy Displays: An Empirical Investigation*.

Credit Author Statement

Melanie Herrmann designed the study, analysed the data, and wrote the first draft of the paper. Xavier Gilbert programmed the materials for the experiment. Longmin Cheng ran participants through the experiment and collected the data. Duncan Brumby and Tadj Oreszczyn contributed to revising the paper and provided advice and oversight. All authors participated in the development of the ideas presented in this paper.