Exploring domestic energy consumption feedback through interactive annotation

**Melanie R. Herrmann · Enrico Costanza · Duncan P. Brumby ·
Tim Harries · Maria das Graças Brightwell · Sarvapali Ramchurn ·
Nicholas R. Jennings**

Abstract We report on a three-week field study in which participants from nine households were asked to annotate their domestic electricity consumption data using a prototype interactive visualisation. Through an analysis of the annotations and semi-structured interviews, our findings suggest that the intervention helped participants to develop a detailed and accurate understanding of their electricity consumption data. Our results suggest that energy data visualisations can be improved by having users actively manipulate and annotate their data, as doing so encourages reflection on how energy is being used, facilitating insights on how consumption can be reduced. One of the key findings from our thematic analysis was that participants went beyond the data in their reflections, talking about generational issues,

M. R. Herrmann · E. Costanza · D. P. Brumby (✉)
UCL Interaction Centre, University College London,
London, UK
e-mail: d.brumby@ucl.ac.uk

T. Harries
Business School, Kingston University, Kingston, UK

M. das Graças Brightwell
Universidade Federal de Santa Catarina, Florianópolis,
Brazil

S. Ramchurn
University of Southampton, Southampton, UK

N. R. Jennings
Loughborough University, Loughborough, UK

upbringing, financial matters, socio-economic comparisons, environmental concern, mistrust towards utilities, convenience, comfort and self-reported waste. Reading beyond the data illustrates the importance of social practices in the context of energy feedback, embedding eco-feedback research into the relevant context of sociology and psychology research.

Anthropogenic climate change is one of the biggest challenges of the twenty-first century according to the United Nations (www.un.org). Residential energy use makes up 25% of total energy consumption in the UK and this share is forecasted to grow (Mogles et al., 2017). At the same time, countries are expected to reduce total emissions by 80% below 1990 levels by 2050 in order to achieve international carbon targets (Cosar-Jorda et al., 2013). To tackle this challenge, several human–computer interaction (HCI) research projects, as well as commercial products, have been put forward. It is hoped that by providing rich feedback to householders, they will learn to use energy more wisely and cut their consumption.

Eco-feedback technology provides feedback about individual or collective behaviour and aims to persuade people to improve their environmental footprint by reducing their consumption of resources, such as energy, water or food (Froehlich et al., 2010). On

the user's side, the aims of using feedback may vary. Looking at residential energy feedback, Karlin (2011) found that two main ends of feedback from the user's perspective are tracking and learning. Tracking refers to the monitoring of ongoing energy use with or without the intention to change behaviour, and learning is to gain specific information about one's energy use such as information tied to specific appliances or energy usage behaviours. The intentions of the user and the fit of the feedback design to meet these intentions shape the outcome of the interaction. This explains why the success of persuasion-based interventions for sustainable behaviour change has been reported to be limited and to vary between studies. Karlin et al. (2015) summarise in their meta-review of 42 feedback studies that there has been much research investigating *whether* feedback works, and not enough research examining the question of *how* it works best. They found a number of factors that moderate the effectiveness of feedback, such as the granularity and frequency of feedback. These two factors relate to how householders use the information in everyday life.

In order to be persuasive in everyday life, feedback needs to address the lived reality of householders and their energy-related behaviour. Research in practice theory (Shove & Walker, 2014; Shove, 2003) and energy policy (Satre-Meloy et al., 2020) highlights the role of daily activities, and their associated comfort and convenience in driving energy consumption, and sometimes suggests that existing culture is an insurmountable obstacle. Against this background, we are interested in how energy consumption data may be interpreted and understood by householders in relation to their own daily activities, and how they link this data meaningfully to their everyday lives. Can frequent and granular data be a starting point for reflection on everyday practices in the household? Data feedback only has implications for behaviour if users can meaningfully reflect on the information and derive actionable insights. Prior HCI work emphasised the role of reflection around personal informatics (Mamykina et al., 2008; Ploderer et al., 2014; Purpura et al., 2011). Reflection is crucial because it determines whether people think for themselves, which enables them to decide if they want to improve their skills or behaviour.

In this paper, we report on an exploratory field study where nine households were provided for three weeks with an interactive visualisation tool named FigureEnergy (Costanza et al., 2012). This tool allowed them to annotate their high-resolution electricity consumption data with labels describing the practice or appliance (such as 'breakfast', 'washing machine'). The aim of the tool, used in our study as a probe, is to help users reflect on their own usage patterns, to learn how, when and to what end they use electricity. The timing of energy use is relevant to householders who are on a time of use tariff, which encourages customers to use energy at off-peak times by offering lower prices. This is an important matter to utilities, who have to balance the load on the grid at peak times which remains a challenge with renewable energies.

The annotations generated by participants through the tool, as well as the tool usage logs, were analysed and used as primary research data, and complemented by semi-structured exit interviews. Every annotation was assessed manually by two researchers to evaluate whether it was plausible or not (e.g. a short 2 kWatts spike labelled 'kettle' would be classed as 'plausible'). In contrast to the initial evaluation of FigureEnergy, which was reported together with the system design and implementation in an earlier paper (Costanza et al., 2012), the focus of the new study presented in this paper is to understand how users reflect on their consumption, using FigureEnergy as a probe.

Our findings indicate that the annotations generated by participants appear to be most of the time plausible, suggesting that they were generally able to understand and explain their electricity consumption data. Throughout the interviews, participants' reflections on their electricity consumption practices brought up a number of themes, including issues of waste and efficiency, how they learned from the consumption data, and considerations related to energy that go beyond what the data show. Importantly, the active manipulation of and reflection on energy data was effective for self-reported behaviour change, at least for some participants. This is a hopeful finding in stark contrast to previous findings that smart eco-feedback was not fulfilling its potential (Hargreaves, 2018; Knowles et al., 2014). However, the study also highlighted the limitations of manual annotation, suggesting the need for a more advanced semi-automatic approach: in line with other recent publications (Mogles et al., 2017), we conclude that 'smart'

feedback must become smarter and offer interactive elements to engage and educate users.

Related work

Our work builds on prior research on eco-feedback and on users' engagement with and reflection about energy data. Positive effects of eco-feedback on energy saving have been confirmed, although evidence is inconsistent and meta reviews found that the effect varies widely from savings of only 1% up to 20% (Abrahamse et al., 2005; Darby, 2006; Fischer, 2008; Karlin et al., 2015). There is little work that provides a reference as to how much the average household could save (Cosar-Jorda et al., 2013). Mogles et al. (2017) considered personal profligacy and sought to address it by smart feedback that built on value framing and provided action prompts. This work found that energy literacy of the participants in the sample increased and consumption decreased by over 20%, where high consumers showed stronger saving effects. This underpins that energy feedback needs to be considered in the social context. Socio-technical problems around energy feedback often relate to engagement, awareness and meaningfulness of the data (Buchanan et al., 2015). This is to say that the technical aspects of the system are interrelated with the users' interactions. An expertly designed piece of technology does not perform perfectly if it's not fit for its purpose. Both technical and social components need to be equally considered and optimised.

The most relevant understanding of 'energy literacy' for the household setting is that of Mogles et al. (2017). They differentiate knowledge about energy on the one hand, and on the other hand motivation to reduce consumption. The latter is driven not by knowledge alone, but also attitudes, values and social practices. Conceptually, energy literacy breaks down into these concepts. Practically, the goal of energy feedback is to improve theoretical understanding and practical competencies at the same time (Schwartz et al., 2013). It is necessary for householders to first make sense of and learn from the feedback, so that subsequent literate choices can be made. However, energy literacy is typically low in the general population and previous work found that householders find it challenging to relate to energy consumption data. For

example, Herrmann et al. (2018) conducted a contextual inquiry with householders using a commercial electricity tracking tool. They asked participants to spontaneously make sense of a time series line graph showing the household's consumption. They found that the sample varied widely in their energy literacy, and independent of energy-related competence, participants struggled to map the data to specific appliances or behaviours.

To better support the sense making process, user interfaces and their interaction design must be optimised for ease of use and to support and augment cognition (Darby, 2010; Hegarty, 2011). Hegarty outlines the different cognitive concepts and processes involved in understanding digital displays. These include attentional and sensory processes to decode the visual information, as well as subjective interpretations to infer conclusions, integrate information into existing knowledge and mental models (internal representation). Particularly with complex displays for data exploration, goals and tasks must be well defined. Current energy feedback systems often fall short in this regard, as they provide complex data to householders with no or little instruction on how to interpret the information.

It is necessary to build on the body of knowledge on how to design feedback that can be easily monitored and learned from by householders. As mentioned before, Karlin et al. (2015) identified moderators that help householders use energy information wisely. One factor is frequency—which used to be low in conventional energy feedback such as bills. They distinguish the frequency with which a system updates, and the frequency with which a user accesses the information. Another factor is granularity; while the authors did not find a significant effect in their review, they still believe that granularity is helpful in the learning process even if it does not necessarily translate into changes in consumption. They also discuss that research has largely focused on feedback itself without further analysing or even reporting the visual design of the presentation. Some reviewed studies did not describe the interfaces in detail, and those that did were not representative of the interactivity and complexity that we see in innovative energy feedback systems. Further, current systems provide high granularity and can be accessed anytime, which makes more user studies necessary to confirm

whether these improvements lead to more learning and higher energy literacy in the home.

Looking beyond monitoring and learning, Sanguinetti et al. (2018) propose a framework to improve the effectiveness of eco-feedback displays by investigating how design choices relate to behaviour. They propose that dimensions like granularity (information), frequency (timing) and the medium and its style (display) have an impact on salience, precision and meaningfulness which in turn influence attention, learning and motivation. The user's attention is directed through the salience of the display, the timing of the feedback and the meaningfulness of the information, e.g. about specific activities or appliances. It is assumed that appliance- and behaviour-specific information is the most actionable type of feedback for householders to inform behavioural changes.

Indeed, there is plenty of research to automate for example appliance-wise disaggregation of energy consumption, known as non-intrusive load monitoring (NILM), originally proposed by Hart (1992). Despite thirty years of research, challenges remain with NILM (Xia et al., 2019; Nalmpantis & Vraskas, 2019). NILM research utilises a range of metrics, datasets and methodologies leading to varying levels of accuracy in modelling load disaggregation. Disaggregating the energy use of any given household in the field without knowing the exact appliances is error-prone; i.e., the algorithm may misclassify appliances or their states. The feedback that could be given to householders on this basis may lead to confusion or misinformation if not accurate enough or containing too much missing data.

And despite the many potential benefits of automation, sustainable design ought to carefully consider the trade-off between automating smart systems and sustaining users' active engagement through interactivity (Alan et al., 2016). To design for both ease of use and engagement at the same time, a system should strike a wise balance between automation and interactivity. A lack of automation risks overwhelming the user. On the other hand, a fully automated system risks low user engagement. Prost et al. (2015) emphasise that in sustainable designs, interactivity and exploration of the data are crucial to empowering users (in contrast to trying to raise awareness through passive consumption). Interactivity allows users to explore the data and to investigate their usage patterns to learn specific information about appliances or

behaviours (Hegarty, 2011; Karlin, 2011; Rheingans, 2002). Recent experimental evidence by Salmon and Sanguinetti (2020) shows that interactive feedback is more effective than feedback without interactivity. The researchers tested a map-based visualisation of urban energy data with and without interactive features and found that interactive features increased interpretability of the energy data map. This explorative process is in line with constructionist learning theory (Papert, 1980). Actively engaging with data is effective for deriving insights about one's life and subsequently, for changing behaviour (Fogg, 2002; Ploderer et al., 2014).

It is therefore important that HCI systems enable engagement and reflection about everyday aspects of people's lives. The most important type of reflection, according to Fleck and Fitzpatrick (2010), is transformative reflection which fosters a change in understanding or practice based on the acquisition of new perspectives. Such new perspectives can be acquired through reflection-in-action and reflection-on-action (Schön, 1987). Reflection-in-action means to reflect at the time of doing; reflection-on-action means reflecting on previous activities. Ubiquitous technology has the potential to facilitate both types: the user could look at the feedback in the very moment when they are carrying out a certain behaviour, or at the end of the day or week to look at their consumption over time.

Neustaedter et al. (2013) conducted an interview study focusing on family's routines asking participants to relate them to their energy bills. They found that participants mostly found external explanations for their data (such as temperature) rather than looking at energy-related everyday activities inside the home. The advantage of smart technology over energy bills is that they collect and provide richer data and that they can analyse such connections and patterns as discussed by Schön (1987).

The purpose of this study is to investigate how users reflect on the energy intensity of everyday behaviours based on data feedback. We deployed the FigureEnergy system, which records electricity data and allows householders to annotate their consumption. Our analysis considers both the annotations generated by participants using the system and a thematic analysis of semi-structured contextual interviews carried out in participants' homes at the end of the study. The aim of this exploratory field study

is to gain insights into householders' ability to make sense of their consumption data, when given the opportunity to actively engage with it and reflect on it over the course of several days or weeks. This is in extension to existing research which shows limited ability in householders to explain their data spontaneously (Herrmann et al., 2018). Before describing this study in detail, we first give a brief description of FigureEnergy.

FigureEnergy

FigureEnergy allows users to browse and annotate their electricity consumption logs. It is implemented as a Web application accessible from standard Web browsers. Electricity consumption data are collected through off-the-shelf sensors. While we refer to the original paper introducing FigureEnergy (Costanza et al., 2012) for a detailed description, together with its design rationale and implementation details, in this section, we summarise the key features critical for the new study reported here, including changes from the original version. The system includes two main Web pages: the Consumption Graph and the Consumption Overview (Fig. 1).

As detailed below, the Consumption Overview was accessed only sporadically by our participants, and its usage did not emerge in the interviews. So, in what follows, we focus on the Consumption Graph. The Consumption Graph was known as the 'logger view' in the previous version of FigureEnergy, and it was renamed for clarity. It displays the recorded electricity data as a time series line graph, with variable resolution: users can zoom in or out, down to a resolution of 2 min, as well as pan left or right.

The Consumption Graph allows participants to annotate the line graph, thus allowing for reflection-in and reflection-on action (Schön, 1987). They can select a time period by click-and-drag, and then annotate this time period by adding an 'event label'. FigureEnergy comes with a set of event labels such as 'meal breakfast', or 'toaster' (full list in Fig. 4). Further, the system allows the user to optionally add text to describe in more detail what they were doing and which appliances they were using.

Fig. 1 FigureEnergy's interactive Consumption Graph (top) and Consumption Overview (bottom). The Consumption Graph was used by participants to annotate their electricity consumption during the course of this study. The Consumption Overview shows usage by label based on the annotations. In the interviews, participants in this study did not report using the Consumption Overview
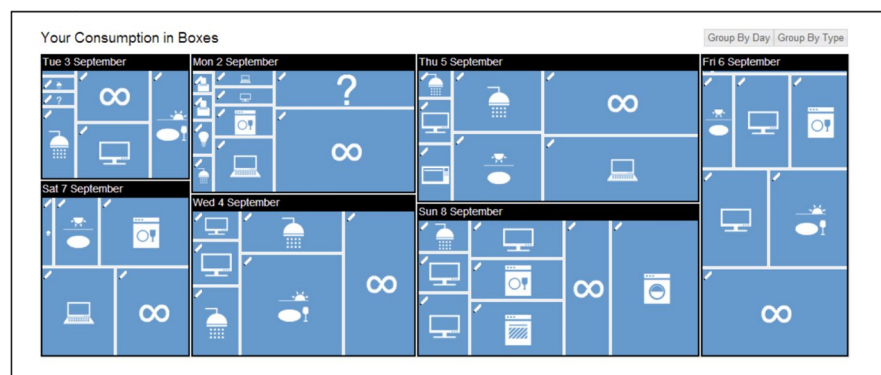
Subsequently, hovering over the icon would prompt that description to be displayed. While in the initial version of the system the labels were mostly appliance-specific (e.g. *toaster*, *kettle*), in the new version new practice-oriented (e.g. *meal breakfast*) and generic (*other A*, *other B*, *other C*) labels were also included. A short video demonstrating the annotation of events is available on https://vimeo.com/42328926.

Methods

Sample

Twelve participants from nine households from a suburban area of London, UK, took part in the interviews (eight females, four males). The households ranged from flats to terraced houses, to cover a range of lifestyles. Table 1 details the participants' information, using fictional names to protect privacy. Participants

Table 1 Household and participant details including occupation and property information. Not all household members took part in the interviews. Interviewed occupants in bold

household	occupants	occupation	years in property	owned	N bedrooms	average daily electricity consumption (kWh)		average daily always-on consumption (kWh)
1	**Helen** Alan dog	account manager	3	yes	2		9	3
2	**Faith (40s)** cat	recruitment consultant	n/a	no	1		5	3
3	**Lorraine (40s)** Derrek (40s)	HR consultant logistics management	7	no	3		8	4
4	**Boris (80s)**	retired (advertising)	13	yes	4		6	3
5	**Justine (50s)** husband daughter (10s) son (10s) dog fish	admin in a company working full time	10	yes	3		35	12
6	**Linda (40s)** husband son (11-18) son (4-10) son (4-10)	charity (not curr. work.) univ. lecturer (history)	20	yes	3		15	4
7	**Sarah** **Andy** Jenny (0-4)	tech. proj. manager horse racing	7	yes	3		9	3
8	**Elain (70s)** husband (70s) **Keith (son, 40s)**	retired retired gardening & delivery driver	47	yes	3		24	13
9	**Chloe** husband daughter (20s) daughter (20s)	clinical receptionist eng. consultant working full time working full time	15	yes	4		9	3

were recruited by a combination of convenience sampling and snowball sampling. In addition, because the sample in the earlier FigureEnergy (Costanza et al., 2012) study comprised mostly very numerate participants, an effort was made to avoid participants from professions that demanded high levels of numeracy. As an incentive to participate, householders were given £10 at the start of the trial and a further £40 on its successful completion.

Material

Participants were provided off-the-shelf electricity consumption sensors measuring total household electricity consumption and were given a username and password to use the FigureEnergy system through their own desktop or laptop computers. At the time of the study, FigureEnergy did not support mobile or tablet interfaces, so access from these devices resulted in an error message being displayed, suggesting users to access the service from a desktop or laptop computer. Appliance-level data unfortunately were not available, due to financial and installation constraints.

Procedure

The study started with an initial home visit. After getting formal consent, an experimenter installed the electricity consumption sensor, and demonstrated the use of FigureEnergy. Participants were asked to access the system daily to annotate the 'peaks' in their consumption data. The season at the time of the study was autumn. During the duration of the study, participants were emailed a weekly reminder to annotate their Consumption Graph. The data collected in this study includes house-level electricity consumption, the annotations created by participants and interaction logs (i.e. when and how long participants used FigureEnergy).

After approximately three weeks (depending on availability), a follow-up visit was arranged, to conduct a contextual semi-structured exit interview, and to collect the sensor kit. Questions asked in the interview were for example: How would you go about annotating the graph? What kind of things caught your attention when doing this (annotating the graph)? Was there any change in the way you understood electricity? If you had to reduce 10% of your overall consumption, what do you think you would cut out? The interviews were audio recorded and fully transcribed. They lasted from 50 to 108 min, with an average duration of 67 min ($SD = 19$). In what follows, households are identified as H1 to H9, and participants are referred to by pseudo names as in Table 1.

Data analysis

The interview data was analysed thematically by two researchers (Aronson, 1994; Braun & Clarke, 2006). The data collected by FigureEnergy was analysed as follows. The number of annotations and annotations per type (e.g. per appliance or activity) were counted. Additionally, two researchers assessed every single annotation with a view to how plausible it was. To the best of our knowledge, letting users annotate their electricity consumption data is novel, and we chose the following approach to assess the accuracy of annotations in lieu of automated appliance level feedback. Each annotation was coded as 'plausible' or 'not plausible'. This analysis was based on the duration, energy amount and the power pattern for each annotation. In particular, an annotation (e.g. 'kettle') was coded as 'plausible' if its duration, energy amount and maximum power were compatible with typical appliance ratings (e.g. up to few kWatts for a kettle) and usage patterns (e.g. up to few minutes for a kettle). In contrast, an annotation would be labelled 'not plausible' if the label did not match the duration, energy amount and power pattern (for example a spike in power up to 2.5 kWatts labelled 'lights'). Annotations where participants expressed uncertainty (e.g. 'there was a bit of a mystery') or were very general (e.g. 'housework') were counted as plausible, because they were meaningful and did not indicate misunderstanding of the data. If the attempts to explain the data were reasonable, annotations were considered 'plausible', since the aim of FigureEnergy was to provide an active engagement with the data to foster deeper reflection. After an initial round of independent coding by the two researchers, the instances where there was disagreement were discussed to reach consensus. The analysis of the annotations was not discussed with participants, and conducted by the researchers after the interviews. This was done so as to not delay the interviews, but conduct the interviews immediately when participants would still remember their annotations. Figure 2 shows two annotations,

Fig. 2 Examples of annotations that were classed as 'plausible' (left, 2 kW spike labelled 'kettle') and 'not plausible' (right, 2.5 kW spike labelled 'lighting')
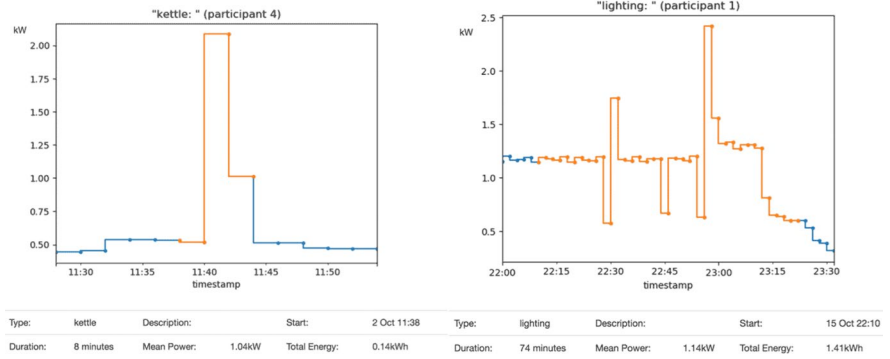

Results

We report findings from the semi-structured interviews through thematic analysis (Clarke & Braun, 2014). We also present information on system usage based on automatic interaction logs, and an analysis of the annotations created by participants using the FigureEnergy system.

Engagement with the FigureEnergy system

To characterise the participants' activity during the study, access logs from the server running FigureEnergy were processed to identify interaction sessions: periods where each user is not inactive for more than 5 min. Each interaction session was used as a unit of analysis, considering which pages were accessed. The information was then aggregated on a daily basis, counting in how many sessions per day each page was accessed, and the results are reported in Fig. 3. It can be noticed from the figure that throughout the study the Consumption Graph page was accessed considerably more than the Consumption Overview page. Moreover, the overall frequency of access went down over time.

Annotations

Annotation of one's own energy consumption data was the primary form of engagement for participants in our study. We analysed the annotations to
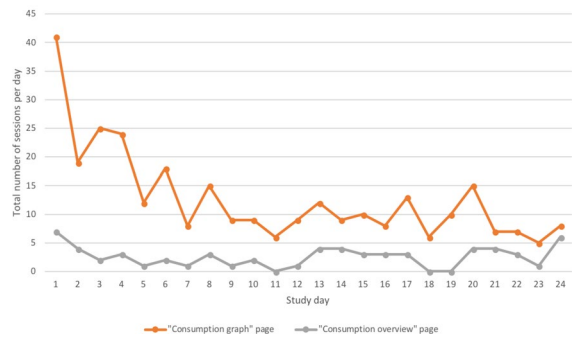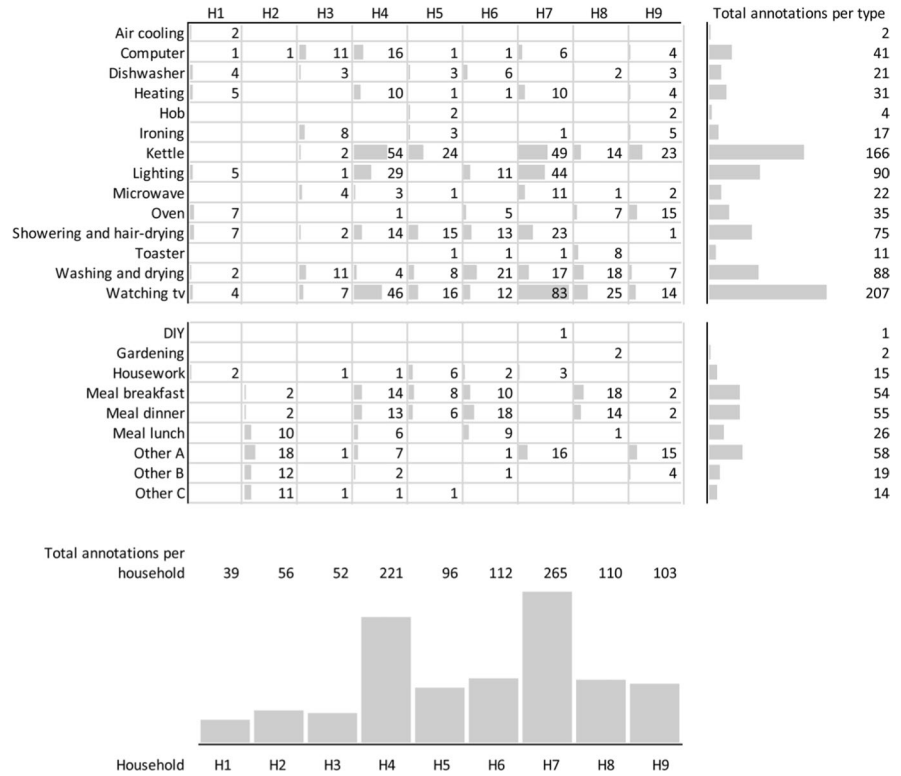

Fig. 3 Participants' page access over the course of the study both for the interactive Consumption Graph (used to make annotations) and the Consumption Overview (no explicit instruction to use in this study)

characterise them, and to complement the analysis of the interviews.

Overall, participants created 1054 annotations over the three weeks of the study, corresponding to an average of 117.1 per household ($SD = 77.0$), and 5.6 per household per day. For 359 annotations (34.1%), participants entered a textual description, for the remaining 695 they only selected a type with the associated icon. Figure 4 illustrates the number of annotations generated by each participant (columns) per type (rows). The types are presented in two groups: annotation types related to specific appliances (such as 'toaster', 'computer', 'kettle') at the top of the figure, while annotations that are more generic (such as 'housework', 'meals', or 'others') at the bottom. The histogram at the very bottom of the figure summarises the total number of annotations per participant. This plot reveals that two of the participating households, H4 and H7, were particularly prolific in their annotation, having generated 221 and

Fig. 4 Number of annotations per type, i.e. appliance or practice (rows) and per household (columns)

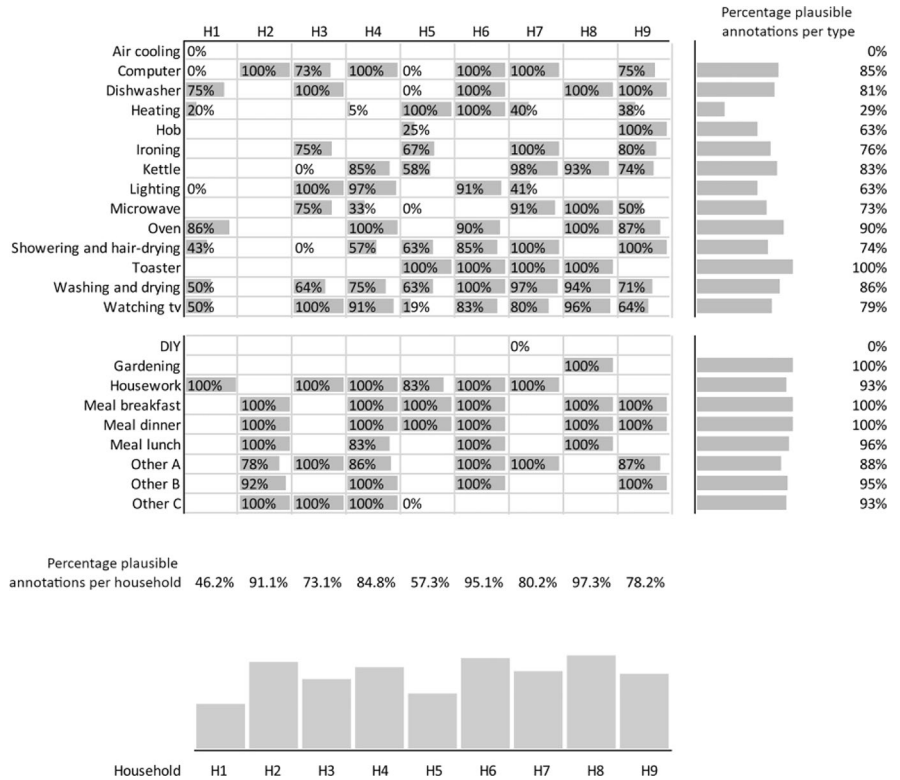|	H1	H2	H3	H4	H5	H6	H7	H8	H9	Total annotations per type
Air cooling	2									2
Computer	1	1	11	16	1	1	6		4	41
Dishwasher	4		3		3	6		2	3	21
Heating	5			10	1	1	10		4	31
Hob					2				2	4
Ironing			8		3		1			17
Kettle			2	54	24		49	14	23	166
Lighting	5		1	29		11	44			90
Microwave			4	3	1		11	1	2	22
Oven	7		1		5			7	15	35
Showering and hair-drying	7		2	14	15	13	23		1	75
Toaster				1	1	1	8			11
Washing and drying	2		11	4	8	21	17	18	7	88
Watching tv	4		7	46	16	12	83	25	14	207
DIY							1			1
Gardening								2		2
Housework	2		1	1	6	2	3			15
Meal breakfast		2		14	8	10		18	2	54
Meal dinner		2		13	6	18		14	2	55
Meal lunch		10		6		9		1		26
Other A		18	1	7		1	16		15	58
Other B		12		2		1			4	19
Other C		11	1	1	1					14
Total annotations per household	39	56	52	221	96	112	265	110	103	

265 entries each respectively. Another group of participants, H5, H6, H8 and H9 generated around 100 annotations each (ranging from 93 to 110), while the remaining 3 participants (H1, H2 and H3) generated about half this amount (from 39 to 56). The horizontal bar chart at the right of Fig. 4 presents the total number of annotations per type. This chart reveals that 5 annotation types were considerably more popular than the rest: *kettle* (207), *watching tv* (166), followed by *lighting* (90), *washing and drying* (88), and *showering and hair-drying* (75). The rest of the types were used at most 58 times.

We analysed the 'age' of the annotations, defined as the amount of time passed between when each event took place (the end of the event) and when the corresponding annotation was created. Values ranged from a little less than 6 min (5.8 min) to more than 6 days (9209.9 min), with a median value of 9.4 h, and the third quartile at 20.1 h (indicating that 75% of the annotations were created within 20.1 h of the end of the corresponding event).

Figure 5 illustrates the percentage of plausible annotations per participant (columns) and type (rows). Similar to Fig. 4, at the top are types that are appliance-specific, while at the bottom those that are practice-oriented or more generic. The histogram at the very bottom of the figure shows the average accuracy (i.e. percentage of plausible annotations) for each participating household. This histogram reveals that for seven households, plausibility scores were over 70%, with H8 reaching the highest plausibility score of 97%. Instead for H1 and H5, the overall annotation plausibility reached 46% and 58%, respectively. It is worth noting that the two households who were most prolific in creating annotations, H4 and H7, had scores in the 80 s range. The bar chart on the right of the image reports the overall plausibility per annotation type. Within the specific appliances, the event label 'heating' received the lowest number of 'plausible' scores (32%), followed by 'lighting' (63%). Microwave, showering, hob, ironing and watching TV range between 73 and 79%. Dishwasher, kettle, computer, washing and drying, and oven reach 81% to 89%. All events labelled 'toaster' have been assessed as plausible (100%). The more generic annotation types tend to show higher plausibility, with the exception of DIY (which was used only once, Fig. 5).

Fig. 5 Percentage plausible annotations per type, i.e. appliance or practice (rows) and per household (columns)

	H1	H2	H3	H4	H5	H6	H7	H8	H9	Percentage plausible annotations per type
Air cooling	0%									0%
Computer	0%	100%	73%	100%	0%	100%	100%		75%	85%
Dishwasher	75%		100%		0%	100%		100%	100%	81%
Heating	20%			5%	100%	100%	40%		38%	29%
Hob					25%				100%	63%
Ironing			75%		67%		100%		80%	76%
Kettle			0%	85%	58%		98%	93%	74%	83%
Lighting	0%		100%	97%		91%		41%		63%
Microwave			75%	33%	0%		91%	100%	50%	73%
Oven	86%			100%		90%		100%	87%	90%
Showering and hair-drying	43%		0%	57%	63%	85%	100%		100%	74%
Toaster					100%	100%	100%	100%		100%
Washing and drying	50%		64%	75%	63%	100%	97%	94%	71%	86%
Watching tv	50%		100%	91%	19%	83%	80%	96%	64%	79%
DIY							0%			0%
Gardening								100%		100%
Housework	100%		100%	100%	83%	100%	100%			93%
Meal breakfast		100%		100%	100%	100%		100%	100%	100%
Meal dinner		100%		100%	100%	100%		100%	100%	100%
Meal lunch		100%		83%		100%		100%		96%
Other A		78%	100%	86%		100%	100%		87%	88%
Other B		92%		100%		100%			100%	95%
Other C		100%	100%	100%	0%					93%

Percentage plausible annotations per household: 46.2% 91.1% 73.1% 84.8% 57.3% 95.1% 80.2% 97.3% 78.2%

Household: H1 H2 H3 H4 H5 H6 H7 H8 H9

Annotation styles

In the interviews, participants explained how they went about annotating their consumption data, and when instead instances of electricity consumption were *not* annotated. In most cases, participants adhered to the study instructions, which was to annotate 'peaks' in the consumption data. For example, Elaine and Keith (H8) say in the interview that what they labelled most were the kettle, the toaster, the washing machine, the tumble dryer and the TV. Indeed, these are their most frequently used event labels (Fig. 4). Justine (H5), too, says she focused on the peaks, e.g. in the morning when getting up—indeed among her most frequent annotation types we find 'kettle' and 'showering and hair-drying'. A notable exception are Chloe's (H9) annotations; her style was distinctively different from the others': she selected long time periods and wrote a list of all things she did in that time.

Sometimes during the interview, participants referred to events that they did *not* annotate because they did not translate into peaks. Helen (H1) for example speaks about the radio in the interview: she explains she leaves it on all the time for her dog, and her partner questions if that is necessary. When asked if she ever annotated the radio in FigureEnergy, she negates, explaining that '*It doesn't surge*'. She refers to it as constant usage and says she wouldn't know how much energy appliances are using in the background like the lights or her computer and she says it would be interesting to see a breakdown.

Food practices provide an interesting opportunity to contrast annotations that refer to specific appliances (such as 'microwave' or 'kettle') versus those that refer to more general events (such as 'meal breakfast', 'meal lunch', and 'meal dinner'). Keith and Elaine (H8) created annotations of both types. They would sometimes annotate specific appliances used to prepare food (i.e. oven 7 times, kettle 14 times, toaster 8 times, microwave 1 time; see Fig. 4). However, Keith (H8) explained that he used the more generic meal labels when he was '*pushed for time*' because it was easier than specifying which appliances he used. Boris (H4) also used annotations of both the appliance-specific and more general type. However, his strategy was different from Keith's—Boris explained that, for example, if he made a cup of

tea for breakfast first and porridge in the microwave later, he'd annotate the two appliances separately. If he did everything at once, he'd choose the breakfast symbol. Boris (H4) also mentions it would be nice to break everything down into the separate appliances that are used for making a meal, and the fridge which is always on in the background. He explains that '*making porridge*' is ambiguous because he could be using the microwave or the hob to make it. Justine (H5), like Boris (H4), says it was '*tricky*' to annotate when multiple things were on at the same time and she would have liked a button in the software to annotate multiple items. She used the breakfast label to summarise '*the toaster, the kettle and maybe the oven*' and similarly she used the 'washer dryer' label indiscriminately for washing and drying. Linda (H6) describes a period when she was cooking dinner and running the washing machine and the dish washer and her partner had a bath, so she '*just sort of put one thing [icon] and I listed the others in it [the textual description]*' and she '*couldn't really segregate what was happening in terms of usage with which, because it all seemed to be happening at the same time*'. Helen (H1) never used the 'meal' event labels and explains that they '*don't really eat like that*': she might only boil the kettle to make a cup of tea in the morning and when she cooks in the evening she would use the label 'oven' (used 7 times); she points out that '*you don't use any more electricity while you're having your meal*'. Her interpretation of the 'meal' event labels, then, is energy consumption during the meal, rather than to prepare the meal, as interpreted instead by other participants.

An appliance that stood out were the lights: annotations with this label were frequent, the third most annotated event. The breakdown in Fig. 4 reveals that such high frequency is due mostly to two households: H4 and H7. At the same time, lighting was the second least plausibly annotated type of annotation (Fig. 5)—the low score seems to be due mostly to the many lighting annotations by Sarah and Andy (H7), which were plausible only 41% of the times, and the few (5) by Helen (H1), which were all implausible. In the interviews, some of the participants referred to lights as being '*visible*' in the consumption data. For example, Linda (H6, who annotated lighting 11 times) says she can see '*quite clearly when people get up*' because there is a small increase in electricity consumption when the lights are switched on.

Similarly, Chloe (H9) noticed a '*slow rise*' when she switched lights on; she did list the lights twice in her long annotation lists. Helen (H1) did not talk about lighting in the interview. Sarah and Andy (H7) reported that they annotated the lights when they noticed a '*little blip every morning*'; however, their annotations for lighting often include high power (1 to 3 kW) peaks which are more likely to correspond to kettle usage. Consistent with her annotations, Justine (H5) mentions that she never annotated lights but, in the interview, she wonders how the lights and other appliances in standby (her multiple fridges, the TV, the alarm and chargers) may add up, contributing to her fairly high baseload. Elaine and Keith (H8) neither use the 'lighting' label for their annotations, nor do they speak about conventional lights in the interview—except that they do mention that the light in their aquarium has high wattage.

Learning, surprises and mysteries

The interviews brought to light instances where the use of annotations led participants to gain insights about their energy consumption data, as well as those where annotating was not enough to make sense of the data. Five of the participants reported '*surprises*' or '*discoveries*'. Keith and Elaine (H8) as well as Sarah (H7) describe participating in the study as an '*eye-opener*'. Keith and Elaine (H8) remember that the biggest peaks were caused by the dishwasher, the washing machine and the tumble dryer. Keith (H8) also reports being surprised how high the kettle and toaster spike. Helen (H1) and Sarah (H7) found out that the electric shower consumes more electricity than they thought. Sarah (H7) and Lorraine (H3) learned about the power needed for ironing. Sarah (H7) says that the iron causes a peak in electricity usage which had never '*crossed [her] mind*' before. Lorraine (H3) says she always believed that '*anything that heats really drains your electricity*' and found this assumption confirmed for the iron. Yet, Lorraine (H3) '*could not believe (…) the enormous spike*' her hairdryer caused. Justine (H5) says she was surprised by the height of her baseload (Table 1), which is caused by her having three fridges and freezers. Sarah (H7) states that '*when you think about electricity you don't think about the fridge freezer*' but the study made her think that she will consider energy

efficiency ratings when the time comes to replace the fridge freezer, realising that this is '*going to be a big percentage of your bill*'.

Occasionally, participants referred to 'mystery' events in their data, both in their annotations and in their interviews. For example, Derek and Lorraine (H3) talk about spikes that they could not explain and therefore had not annotated. In the interview, they referred to them as '*unknown*', '*in bed*' or '*out*'. Faith (H2), too, reports '*a couple of unexplained spikes*' that she labelled 'no one in' or 'don't know'. Boris (H4) noticed '*tiny little things*' during the night and wondered if it could be the fridge freezer. Helen (H1) mentions '*random patterns (…) at sort of three o'clock in the morning*' and she had no clue what that could be. Justine (H5) was '*surprised*' and '*puzzled*' that even when nobody was around there were still fluctuations and peaks.

One issue that came up in the interviews was that 6 participants explicitly said that they could not relate to kilowatts (Helen (H1), Linda (H6), Justine (H5), Chloe (H9), Faith (H2) and Elaine (H8)). Sarah (H7), instead, says she '*deal[s] with power at work*' so she understands kilowatts as a unit. Lorraine (H3) and Keith (H8) figure that the higher the kilowatt, the higher the cost of their energy consumption.

Personal and generational circumstances

The study sometimes triggered reflection that goes well beyond the energy consumption data that participants were presented with. Participants who appear to be reading 'beyond the data' refer to a variety of factors, such as their upbringing, generational issues, financial matters, socio-economic comparisons, trust or mistrust towards utilities, convenience, comfort and self-reported waste or avoidance of what they would consider wasteful.

Boris (H4) for example washes his clothes only after wearing them several times and takes three-minute showers. He is not concerned about cost, yet he is very mindful of getting the best deal: '*I suspect that when that contract expires (…) they would try and push me up (…) no, you know, you aren't going to bully me. I'll shop around*'. Reasoning about whether he'd change his behaviour to save energy, Boris (H4) didn't think the technology was going to change his lifestyle: '*If I want a hot drink I'll have one … I'm 80 years old*'. Boris (H4) said he has '*been fairly*

lucky in life' and explains '*just down the road in reality there are people who are probably not even as old as I am who are retired, and will be reluctant to put on heating because they're worried about the cost. I mean I think that is a factor in a lot of people's lives*'.

Chloe (H9) thinks that her daughters are from a generation that thinks '*everything is just automatic*' whereas she grew up in the 70 s with '*shortage*' and '*power strikes*'. Chloe (H9) says '*for my eldest daughter, she's not a silly girl by any means but I suppose she just thought jumping in the shower was using water – not electricity*'. Equally, this daughter '*doesn't like the house quiet. So the telly's on even if she's not in the room*'. Similarly, Sarah (H7) says that as a teenager she would put a pair of jeans that she wanted to wear in the washing machine and that she '*would never think of doing that now*'.

Linda (H6) states that she '*worr[ies] about the planet*', elaborating that '*it's all about being responsible to…being responsible for our planet and all the creatures that live on it, not just ourselves, and just being a good person, really*'. She feels guilty about the increased consumption compared to her parents' generation and finds that everyone needs to be more mindful of their energy use in a world with decreasing resources. Linda (H6) tries to get her kids to switch things off when not needed: '*I feel I am permanently saying "Why is the house lit up like a Christmas tree?"*'.

Reflecting on 'waste'

During the interviews, participants reflected on instances of consumption that they considered wasteful, yet often justifying them, and often reporting that despite the admission of *some* waste, they considered themselves overall quite efficient. Examples of self-reported waste included keeping appliances like the TV on standby out of '*laziness*' (Elaine and Keith, H8), or keeping the TV or the radio on to keep company for the pets (Faith, H2 and Helen, H1). Helen (H1) runs the heating for herself and her dog, whereas her partner Alan gets too warm. Helen critically reflected on their behaviour of sometimes running the heating and the fan at the same time. When the dog was unwell, Helen (H1) ran the heating during the night and Alan opened the window. Helen (H1) also uses the heating to dry her clothes and sometimes keeps it on all night so that Alan's work clothes are

dry in the morning. She comments on these anecdotes in a way that indicates that she considers them wasteful ('*That's bad use of electricity!*').

Faith (H2) has annotated in FigureEnergy that she leaves a hall light on during the night, and in the interview, she explains: '*I'll get up during the night and go to the loo and that way I can see where I'm going (…) and also (…) my cat is not allowed in my bedroom at night because she keeps me awake, so being the big softie I am I leave the hall light on for her, just completely ridiculous I know*'. She also leaves the radio on for the cat: '*I know that I'm using energy for that, but that's my choice and I'm paying for it*'. She explicitly considers this choice, and agency, justified by the payment of the energy.

Lorraine (H3) speaks about keeping the outside light in front of the house on because it helps her find her keys and because it makes it look like someone is home (she did not annotate this). This security is important to her despite her saying that it seems wasteful. In contrast, she refers to herself as the '*electricity police*' switching everything off when not needed, because '*waste generally irritates*' her and her heuristic is '*if you don't need it, you shouldn't really be using it*', explaining this is '*an environmental thing*' and '*general awareness*'.

Justine (H5) when thinking about wasteful behaviours explains that they have a TV in each of the bedrooms and she and her partner and her daughter sometimes all watch different programs simultaneously while her son plays the Xbox. She reasons they could vote on a program and all watch together, but other than that, she doesn't see how they could save as they are out during the day. This reflection only came up in the interview, the annotation data did not include any evidence of this.

Despite their reflections about waste, in general, all participants reported believing that they are '*reasonably energy efficient*' (Faith, H2) and '*only use what's necessary*' (Helen, H1). Linda (H6) and Lorraine (H3) say they already only wash clothes if they are dirty and that they only wash full loads of laundry. Relating to cooking practices, Linda (H6) '*multitasks*' when using the oven trying to use the heat to cook several things.

When asked about how they could save energy, Helen (H1) suggests, theoretically speaking, using less lighting, not listening to the radio all day, and wearing jumpers to have the heating on less. Linda

(H6) considers to maybe precook meals on the weekend and reheat them on weekdays, and jokingly said they '*could eat more salad*'. Justine (H5) reckons they '*probably do waste power (…) like any family*' but she would not change anything because she doesn't consider themselves '*that wasteful*'. Justine's baseline is high due to the three fridge freezers in the house. When asked about getting rid of one or two of these appliances, Justine (H5) explains '*you hear people, their freezers break down and so at least with having the three, you know, if there was a problem with one, we could then swap it into another*'. In contrast to such resistance to change, in the next subsection we report instances where participants mentioned behaviour change in reaction to taking part in the study.

Behaviour change prompted by FigureEnergy

Sarah (H7), who had described the feedback as '*quite an eye opener*', states that she has become concerned about her consumption. She and Andy (H7) report a range of insights and consequently behaviour change. For example, Sarah (H7) has '*been more frugal with the use of the dryer since doing this [taking part in the study]*' and she switches the lights off more often as opposed to having them on during the day (we have not found evidence for this in the annotations). At times, they had been running an electric heater in their daughter's room but upon discovering how much it consumes they reconsidered using it and concluded warmer pajamas would do. Equally, they used to put their daughter's towel in the dryer '*just quickly to warm it up when I got her out of the bath*', which '*made [them] think that it was a pure (…) luxury rather than a necessity*'. They point out that the information does not make them say '*that's got to stop*' but rather gets them to think. Andy (H7) reasons: '*Well, you got to have lights on when it's dark, you got to have your fridge on. You've got to make a cup of tea now and then, you know, you've got to have a shower. There's things that you just can't avoid, but there are a lot of things you can avoid*'.

Chloe (H9) found that the study is '*making you aware again because I think you do get complacent*'. She has occasionally turned the radio off completely instead of keeping it on standby. She's made further changes regarding the washing of dishes and clothes: '*I thought, no, actually today I'm not going to put [the dishwasher], I'm going to wash the breakfast things*

up, I'm going to wash the lunch and wash the evening meal things up (…) I think I could live without the dishwasher (…) I think I would reduce it down to probably just the weekends'. For washing her clothes, Chloe (H9) reported washing less and using a more efficient program: '*I have put the washing machine on quick washes as opposed to longer cycles*' and "*I said to my youngest daughter: you've only worn these jeans today…*" *I think really you could wear them sort of two or three times (…) you could air them (…) they're not dirty. And they smell fresh still. You can still smell the comfort on them for goodness sake*'. In contrast to Chloe's account of changing her washing practices, we did not see an apparent reduction across the three weeks of the study.

Derek (H3) admits that in the past he would sometimes '*be ironing, having the telly on, have the laptop on, stop ironing for a bit, answer a couple of emails or something like that but you've left the iron going at the same time*'. Derek (H3) has learned to make changes based on what Lorraine (H3) learned from the feedback: '*One of the things Lorraine said to me, and I'm conscious of now is, "don't turn the iron on and iron one shirt. If you're going to turn the iron on, you know, iron quite a lot of stuff because otherwise you're going to get a great big spike*"'.

Discussion

Making sense of energy consumption data

HCI research on behaviour change has acknowledged that there is more to monitoring data than feeding back information to consumers. Sometimes, users might wish to monitor only, without the desire to change, or they monitor simply out of curiosity, with no intention to change at all (Karlin, 2011; Epstein et al., 2015). Therefore, rather than measuring behaviour change (which is influenced by a plethora of factors), we focus on monitoring, understanding the data, and learning from it. The research objective was to inquire how people make sense of and reflect on energy data. In contrast to earlier work, our study found that participants did rather well in terms of explaining their data patterns: Overall, annotations were plausible which is in contrast with the field study interviews by Herrmann et al. (2018), who concluded that householders perform rather poorly in

making sense of their energy data when presented as a time series line graph. The core difference between the two studies is that thanks to FigureEnergy as an interactive prototype, in our study participants could annotate and hence actively reflect on their data patterns as often as they wanted, and they were encouraged to do so. Indeed, they created 75% of their annotation within 20 h of the end of the corresponding event, reflecting on their consumption repeatedly over the course of the study. In contrast, earlier work instead asked participants to reflect on their data as a one-off activity and without featuring interactive elements, looking back at their energy consumption for the past day(s) and week(s) (Herrmann et al. (2018). The annotation feature added value to the line graph visualisation. This adds to the existing evidence that interactive feedback is more effective than feedback without interactivity (Karlin, 2011; Hegarty, 2011; Rheingans, 2002; Prost et al., 2015; Salmon & Sanguinetti, 2020). Schön (1987) suggested that people can learn from reflection on previous actions and FigureEnergy helped participants in this study to do this. The good performance can then be explained in relation to memory: most participants annotated events on the same day when they had taken place, based on recent memories about what they did on the day. These results are also in contrast to those reported by Strengers (2011), who highlights the limitations of simple, and low resolution, energy consumption feedback displays. The high resolution of FigureEnergy and the possibility to zoom in and out provided rich data for participants to link energy use patterns to everyday practices. It is worth noting that FigureEnergy was available as a desktop application only. On a mobile device, it would be interesting to explore whether the time between an event and its annotation would decrease, and whether annotation plausibility would increase.

Our findings are typical in that participants cannot relate to kilowatt-hours as a measure of energy use, and in that lights stand out as an appliance that is annotated frequently, but often incorrectly (Strengers, 2011). This can be explained in terms of cognitive biases: small but salient appliances come to mind more easily, and householders tend to overestimate them (Attari et al., 2010). However, a contrasting and encouraging finding is that Linda (H6), Chloe (H9) and Sarah and Andy (H7) learned that lights cause only a very small increase in the graph, and because

of that Justine (H5) and Elaine and Keith (H8) did not annotate them at all, in line with the instruction to annotate the peaks in particular (as demonstrated also in Fig. 4).

There were data patterns that participants referred to as mystery events. Again, this is in line with previous findings. Herrmann et al. (2018) interviewed householders on a similar looking time series graph and there were data patterns, particularly small fluctuations during the night, but also spikes during the day, that participants were unable to explain. In this study, annotations were classified as plausible when participants said they did not know what caused the data pattern. Often, it was a case of minor fluctuations which might not be caused by a specific event.

Knowledge about the energy consumption of everyday behaviours has been improved by the study. This is evidenced by the examples of participants who gained insights and self-reportedly changed their habits. Our findings on waste resonate with those by Schwartz et al. (2013): reflecting on their energy consumption led our participants to reflect on waste, and relate the concept of waste to a complex of needs (keeping the lights on for safety) or responsibilities (heating and cooling the house at the same time for the wellbeing of a pet), similar to what was reported from that previous study. A positive finding in this study is that in some cases, participants (reportedly) discontinued certain practices, commenting that the study made them realise they were not justified by a need (such as Sarah and Andy (H7) preheating their daughter's pajamas). However, we did not see the self-reported behaviour changes reflected in the consumption data, maybe because our study was too short to reliably identify changes in consumption. Indeed, this study did not focus on behaviour change and changes in consumption, so it remains to be investigated whether learning and self-reported adjustments result in or correlate with measurable savings. It is possible that participants would respond to social factors in their interaction with the interviewer and exaggerate improvements. Also, it is possible that participants' behaviour did not actually change to the extent that they believed it did.

From annotations to conservation orientations

The design of the study allowed us to combine data from the interviews with that generated by participants through the annotations. Three households reported behaviour change, and six didn't. Two households explicitly justified continuing practices that they considered inefficient.

Overall, there is a split between most participants whose annotation plausibility was quite high, and two participants whose plausibility score was notably lower. Interestingly, Helen (H1) and Justine (H5), whose annotations have the lowest plausibility in the sample, also mention some of the most energy-inefficient behaviours in the interview, such as heating and cooling the house at the same time (Helen, H1) and having an exceptionally high baseline due to three fridge freezers (Justine, H5). We see from the energy data that Justine (H5) is a high consumer, alas she does not realise it, and she does not change her behaviour. Helen's (H1) interview data, too, reflects her annotation style: Helen (H1) reported generic actions such as switching appliances off from standby when she went away for a weekend which shows that her general awareness might have been increased by the study but she did not make any specific discoveries in the data which corresponds to her few plausible annotations. These findings seem to suggest that low annotation plausibility predicts poor learning about the energy consumption in the participant's home.

In contrast, participants whose annotation plausibility is above 70% seem more likely to be either more economic or to identify and change wasteful habits. Boris (H4) for example is not motivated to reduce his consumption saying he can afford to be '*reckless*' at his age, and yet he is '*naturally conservative*' in his behaviour, i.e. only washes clothes after wearing them multiple times. Linda (H6) reports feeling very strongly about not lavishing energy. Sarah and Andy (H7) used to have '*wasteful*' habits but upon discovering this in FigureEnergy they changed them. Looking at Boris (H4) and Linda (H6), one might think that participants whose annotations, as estimated by the authors, are mostly plausible, were environmentally aware and economical to start with. The cases of Sarah and Andy (H7), Chloe (H9), and Lorraine and Derek (H3) invalidate this assumption. Schwartz et al. (2013) have introduced the idea that it is key what people do with technology, as opposed to what technology does to people. Sarah and Andy (H7), whose annotation accuracy was classified as good by the authors, used to engage in highly energy-intensive behaviours before taking part in the study,

but upon reading in the data and recognising their profligacy, they reported they stopped using the tumble dryer when not necessary.

The results provide novel evidence of how people read and reflect on energy data. We found two patterns of 'reading energy data': we refer to them as reading in the data and reading beyond the data. By reading in the data, we mean householders who are analytically reflecting on the energy data. A couple of participants referred to the study as an eye-opener, and several identified the biggest consumers in the home. This is valuable because knowing where the energy goes is the first step in reassessing one's energy use.

Going beyond the data

Some of our participants were reading (also) beyond the data: they talked about generational issues, upbringing, financial matters, socio-economic comparisons, environmental concern, mistrust towards utilities, convenience, comfort and self-reported waste. This theme is key to our findings and implications. It is important to note that reading beyond the data, as opposed to reading in the data, is not indicative of either good or bad energy use or of good or poor annotations. As it can be noticed from Fig. 5, participants who went beyond the data (e.g. H4, H6, H9) tend to have plausible annotations, and they also produced a comparatively high number of annotations. So, these comments should not be considered a 'replacement' for engaging with the data, rather an extension. For example, in the interview, Boris (H4) indicates that he feels strongly about not using energy without purpose and the lion's share of his interview reflects him talking beyond the data. At the same time, his annotation plausibility is high. Based on the data collected in this study, we can only observe that Boris' actual consumption is aligned with his attitude towards energy use, and that his annotations are plausible, without demonstrating a relationship between his pro-conservation behaviour and his energy literacy.

The idea of excessive use, or waste, is an interesting example of going beyond the data because it always goes beyond what we can see from data—the data only reflect how much energy is being used. It comes down to personal assessment if this use is necessary or profligate. How economically energy

is used is typically shaped by childhood education, comfort preferences and material circumstances. We acknowledge that cultural context has an influence on lifestyle, perception of waste, generational differences, and that such social influences could be further explored in future research.

Shove (2003) found that convenience and comfort come first for many householders and prevail over attempts to save energy; i.e., learning about one's consumption does not automatically lead to change. This is because it is everyday routines, habits, and household tasks that are on people's minds, rather than sustainability and energy efficiency. Even if feedback makes consumption visible, energy as a resource hardly becomes the main consideration, but the need to cook food and wear clean clothes remain the social drivers. Strengers (2011) depicts the householder not as a micro resource manager, rationally decreasing cost as much as possible, but as a social being facing the realities of everyday life mediated by cultural factors. Her suggestion therefore is to rethink energy feedback as an opportunity to script sustainable interactions, such as washing machines running cold washes by default (p. 2141). We can learn from studies like this to understand better how people live their lives and use energy, and to transform the socio-technical domain by changing norms and homes to create greener lifestyles.

Limitations

This study comes with a number of limitations. First, as this is exploratory work with a small sample, this approach is necessarily limited in how much we can learn about the wider population, particularly considering that there may be an opt-in bias of participants wishing to take part in a study to learn more about their consumption. We took care to recruit participants with average energy literacy, but they are not necessarily representative of the general population and it can be assumed that some householders would not be motivated to actively engage and interact with energy feedback.

Second, the short duration of this study is a limitation that poses an opportunity for further work. This includes the need for more research on how to keep users engaged with smart energy feedback long-term, whether it is manual, automated or a mix of AI and interactivity. Some benefits of feedback are valuable

one-off insights, such as discovering that the tumble dryer consumes a lot of energy. However, nuanced learning, detecting fluctuations over time or responding to time-of-use incentives would require ongoing engagement with feedback. Long-term studies are valuable because it is likely that novelty effects of using the system would wear off. For example, Sarah (H7) and Lorraine (H3) said they were unlikely to use a system like FigureEnergy permanently.

Finally, a specific limitation of this study is that the Consumption Overview was barely used and its potential remains to be explored. This is because participants were primarily asked to annotate the line chart in the Consumption Graph. The instruction was given in the first place to ensure that participants would actively reflect on their data, but it limits our findings and we do not know if the Consumption Overview could have been a valuable learning interface. Further, the instruction to specifically annotate peaks resulted in participants disregarding the base load. It seems that the instruction actively discouraged householders from being more inquisitive with regard to energy use happening in the background.

Implications

Our findings suggest that in line with constructionist learning theory (Papert, 1980) and Schön's (1987) reflection-on-action, the active manipulation of the data has facilitated reflection, learning and in some cases self-reported behaviour change. A previous study (Herrmann et al., 2018) found that people struggled to make sense of their data when it was presented in a similar time series line graph. It seems that the interactive annotation feature was powerful in triggering reflection in our sample (for example, Sarah (H7) and Andy (H7) reportedly quit using the tumble dryer for expendable tasks). Based on our observations, we believe annotation plausibility might reflect the level of engagement and the level of how much users reflect on the data and its meaning for everyday practices. Admittedly, this is only an exploratory study and the hypothesis would have to be examined experimentally to confirm whether there is a robust effect of interactivity.

The annotation task increased knowledge, particularly for actions or appliances that cause peaks. However, participants mentioned that they are not clear how baseline appliances are contributing and would

like an automated breakdown. They also struggled to annotate events when they were tending to multiple household tasks simultaneously. Computational appliance-level disaggregation would meet these user requirements and provide an automated solution to identifying events. However, disaggregation (non-intrusive load monitoring) is a complex technical challenge, and despite almost three decades of active research, it is not completely solved to the end of providing fully automated and highly accurate disaggregation for any given household without training the model specifically for the given scenario (Nalmpantis & Vrakas, 2019; Parson et al., 2015; Xia et al., 2019). The findings of this study indicate that device-specific information would be helpful to enable meaningful reflection on smart energy feedback. If accurate appliance-wise disaggregation became available in the near future, an automated breakdown would likely replace manual annotations. It remains subject of ongoing research whether such an automated breakdown would change energy behaviours.

A combined solution could then be a form of AI-enabled smart assistant: a system that is still centered on user-generated annotations, but with automatic assistance. Smart meters are not yet smart enough (Mogles et al., 2017) and our participants' annotations could be improved: for example, if the power for a selected event stays below a certain threshold or goes over it, the software could provide feedback to the user if the selected event label is unlikely to cause the annotated peak (i.e. implausible). This might prompt users to correct biases and identify the big consumers in their home. Similarly, the software could suggest likely labels for the 'mystery' events. On the other hand, manual annotation can help mitigate poor results from machine learning (e.g. due to one-off events) and weed out outliers.

The purpose of FigureEnergy was to provide richer context information to users and make the energy feedback as activity-centric as possible. Two participants (H1 and H4) spoke about whether it makes sense to label an event as a 'meal'—it is convenient and seems to reflect everyday practices, yet it limits the usefulness of the data feedback because it doesn't tell the user which appliances were used. This is in contrast to our assumptions that activity-centric feedback matters more than appliance-centric information based on the previous publication by Costanza et al. (2012). The work of Grünewald and Diakonova

(2019) supports the assumption that activity records matter in relation to householders' electricity consumption. They match electricity consumption to activity logs recorded via an app to determine the typical energy consumption of a household activity. Both in Grünewald and Diakonova (2019) and in this study, meals stand out: particularly cooked meals using high-power appliances produce a salient peak in demand. Based on this study, we suggest that the ambiguity of the 'meal' label highlights an opportunity for improved annotation tools. Our results may indicate that appliance-specific feedback on energy use may be helpful for users after all because it provides more information about consumption. More research is needed to investigate the difference between appliance-centric and activity-centric feedback, and their usefulness for householders. The difference between the two has emerged as a theme in this study, but we have not yet systematically examined the two alternatives.

Conclusion

This paper reported on a field study with 9 households who interacted daily with an interactive energy visualisation prototype called FigureEnergy. Participants used the software as a digital diary to record and annotate energy behaviours in the home. After three weeks, they took part in an interview in which we explored what they had learned. Both participants' digital annotations and semi-structured interview data were analysed. The findings of this study suggest that energy feedback can be made more meaningful through interactive systems that trigger users to reflect on their energy data and relate the information to their everyday life. FigureEnergy's annotation feature allowed participants to engage with their energy data more interactively in comparison to feedback that is just displayed for 'passive' consumption. Some of our participants discovered insights about everyday practices in the home, including wasteful behaviours that they wanted to change. This study did not provide enough evidence to conclude whether meaningful reflection and accurate insights significantly impact environmentally responsible energy use. As far as FigureEnergy is concerned, we emphasise how the system design needs to find a balance between users' input and AI to help users improve their annotations.

An opportunity for future work is to test autonomous assistance for annotations and to observe behaviour change and monitor energy savings in longer term studies.

Declarations

References

Abrahamse, W., Steg, L., Vlek, C., & Rothengatter, T. (2005). A review of intervention studies aimed at household energy conservation. *Journal of Environmental Psychology, 25*(3), 273–291. https://doi.org/10.1016/j.jenvp.2005.08.002.

Alan, A. T., Costanza, E., Ramchurn, S. D., Fischer, J., Rodden, T., & Jennings, N. R. (2016). Tariff agent: Interacting with a future smart energy system at home. *ACM Transactions on Computer-Human Interaction, 23*(4), Article 25, 28 pages. https://doi.org/10.1145/2943770.

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

Attari, S. Z., DeKay, M. L., Davidson, C. I., & Bruine de Bruin, W. (2010). Public perceptions of energy consumption and savings. *Proceedings of the National Academy of Science, 107*(37), 16054–16059. https://doi.org/10.1073/pnas.1001509107.

Buchanan, K., Russo, R., & Anderson, B. (2015). The question of energy reduction: The problem(s) with feedback. *Energy Policy, 77*, 89–96. https://doi.org/10.1016/j.enpol.2014.12.008.

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

Clarke V., & Braun V. (2014). Thematic analysis. In A.C. Michalos (Ed.), *Encyclopedia of quality of life and*

well-being research. Springer. https://doi.org/10.1007/978-94-007-0753-5_3470.

Cosar-Jorda, P., Buswell, R. A., & Mitchell, V. (2013). Estimating the potential reductions in energy demand through efficiency, control and lifestyle change in a real home. In *Proceedings of FutureBuild*, p. 135–144.

Costanza, E., Ramchurn, S. D., & Jennings, N. R. (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.

Darby, S. (2006). *The effectiveness of feedback on energy consumption: A review for DEFRA of the literature on metering, billing and direct displays*. University of Oxford.

Darby, S. (2010). Smart metering: What potential for householder engagement? *Building Research & Information, 38*(5), 442–457. https://doi.org/10.1080/09613218.2010.492660.

Epstein, D. A., Ping, A., Fogarty, J., & Munson, S. A. (2015). A lived informatics model of personal informatics. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*, 731–742. https://doi.org/10.1145/2750858.2804250.

Fischer, C. (2008). Feedback on household electricity consumption: A tool for saving energy? *Energy Efficiency, 1*, 79–104. https://doi.org/10.1007/s12053-008-9009-7.

Fleck, R. & Fitzpatrick, G. (2010). Reflecting on reflection: Framing a design landscape. In *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction (OZCHI '10)*, USA, 216–223. https://doi.org/10.1145/1952222.1952269.

Fogg, B. J. (2002). Persuasive technology: Using computers to change what we think and do. *Ubiquity*, Article 5, 32 pages. https://doi.org/10.1145/764008.763957.

Froehlich, J., Findlater, L., & Landay, J. (2010). The design of eco-feedback technology. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*, 1999–2008. https://doi.org/10.1145/1753326.1753629.

Grünewald, P., & Diakonova, M. (2019). The specific contributions of activities to household electricity demand. *Energy and Buildings, 204*(1), 109498. https://doi.org/10.1016/j.enbuild.2019.109498.

Hargreaves, T. (2018). Beyond energy feedback. *Building Research & Information, 46*(3), 332–342. https://doi.org/10.1080/09613218.2017.1356140.

Hart, G. W. (1992). Nonintrusive appliance load monitoring. *Proceedings of the IEEE, 80*(12), 1870–1891. https://doi.org/10.1109/5.192069.

Hegarty, M. (2011). The cognitive science of visual-spatial displays: Implications for design. *Topics in Cognitive Science, 3*(3), 446–474. https://doi.org/10.1111/j.1756-8765.2011.01150.x.

Herrmann, M. R., Brumby, D. P., & Oreszczyn, T. (2018). Watts your usage? A field study of householders' literacy for residential electricity data. *Energy Efficiency, 11*, 1703–1719. https://doi.org/10.1007/s12053-017-9555-y.

Karlin, B. (2011). Tracking and learning: Exploring dual functions of residential energy feedback. In *Proceedings of the*

6th International Conference on Persuasive Technology: Persuasive Technology and Design: Enhancing Sustainability and Health (PERSUASIVE '11). Association for Computing Machinery, New York, NY, USA, Article 10, 1–9. https://doi.org/10.1145/2467803.2467813.

Karlin, B., Zinger, J. F., & Ford, R. (2015). The effects of feedback on energy conservation: A meta-analysis. *Psychological Bulletin, 141*(6), 1205–1227. https://doi.org/10.1037/a0039650.

Knowles, B., Blair, L., Coulton, P., & Lochrie, M. (2014). Rethinking plan A for sustainable HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*, 3593–3596. https://doi.org/10.1145/2556288.2557311.

Mamykina, L., Mynatt, E., Davidson, P., & Greenblatt, D. (2008). MAHI: Investigation of social scaffolding for reflective thinking in diabetes management. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*, 477–486. https://doi.org/10.1145/1357054.1357131.

Mogles, N., Walker, I., Ramallo-González, A. P., Lee, J., Natarajan, S., Padget, J., Gabe-Thomas, E., Lovett, T., Ren, G., Hyniewska, S., O'Neill, E., Hourizi, R., & Coley, D. (2017). How smart do smart meters need to be? *Building and Environment* 125. *Supplement C, 439–450,*. https://doi.org/10.1016/j.buildenv.2017.09.008.

Nalmpantis, C., & Vrakas, D. (2019). Machine learning approaches for non-intrusive load monitoring: From qualitative to quantitative comparison. *Artificial Intelligence Review, 52*(1), 217–243. https://doi.org/10.1007/s10462-018-9613-7.

Neustaedter, C., Bartram, L., & Mah, A. (2013). Everyday activities and energy consumption: How families understand the relationship. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*, 1183–1192. https://doi.org/10.1145/2470654.2466153.

Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas*. Basic Books.

Parson, O., Fisher, G., Hersey, A., Batra, N., Kelly, J., Singh, A., Knottenbelt, W., & Rogers, A. (2015). Dataport and NILMTK: A building data set designed for non-intrusive load monitoring. In *Proceedings of the 2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 210–214. https://doi.org/10.1109/GlobalSIP.2015.7418187.

Ploderer, B., Reitberger, W., Oinas-Kukkonen, H., & van Gemert-Pijnen, J. (2014). Social interaction and reflection for behaviour change. *Personal and Ubiquitous Computing, 18*, 1667–1676. https://doi.org/10.1007/s00779-014-0779-y.

Prost, S., Mattheiss, E., & Tscheligi, M. (2015). From awareness to empowerment: Using design fiction to explore paths towards a sustainable energy future. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '15)*, 1649–1658. https://doi.org/10.1145/2675133.2675281.

Purpura, S., Schwanda, V., Williams, K., Stubler, W., & Sengers, P. (2011). Fit4life: The design of a persuasive technology promoting healthy behavior and ideal weight. In *Proceedings of the SIGCHI Conference on Human*

Factors in Computing Systems (CHI '11), 423–432. https://doi.org/10.1145/1978942.1979003.

Rheingans, P. (2002). Are we there yet? Exploring with dynamic visualization. *IEEE Computer Graphics and Applications, 22*(1), 6–10. https://doi.org/10.1109/38.974511.

Salmon, Klernan, & Sanguinetti, A. (2020). Use cases and best practices for map-based energy data visualizations. UC Davis. Retrieved from https://escholarship.org/uc/item/007295p0.

Sanguinetti, A., Dombrovski, K., & Sikand, S. (2018). Information, timing, and display: A design-behavior framework for improving the effectiveness of eco-feedback. *Energy Research & Social Science, 39*, 55–68. https://doi.org/10.1016/j.erss.2017.10.001.

Satre-Meloy, A., Diakonova, M., & Grünewald, P. (2020). Daily life and demand: An analysis of intra-day variations in residential electricity consumption with time-use data. *Energy Efficiency, 13*, 433–458. https://doi.org/10.1007/s12053-019-09791-1.

Schön, D. A. (1987). *Educating the reflective practitioner: Toward a new design for teaching and learning in the professions*. Jossey-Bass.

Schwartz, T., Stevens, G., Ramirez, L., & Wulf, V. (2013). Uncovering practices of making energy consumption accountable: A phenomenological inquiry. *ACM Transaction on Computer-Human Interaction, 20*(2), Article 12, 30 pages. https://doi.org/10.1145/2463579.2463583.

Shove, E. (2003). *Comfort, cleanliness and convenience: The social organization of normality*. Bloomsbury Academic.

Shove, E., & Walker, G. (2014). What is energy for? Social practice and energy demand. *Theory, Culture & Society, 31*(5), 41–58. https://doi.org/10.1177/0263276414536746.

Strengers, Y. A. A. (2011). Designing eco-feedback systems for everyday life. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*, 2135–2144. https://doi.org/10.1145/1978942.1979252.

Xia, M., Wang, K., Zhang, X., & Xu, Y. (2019). Non-intrusive load disaggregation based on deep dilated residual network. *Electric Power Systems Research, 170*, 277–285. https://doi.org/10.1016/j.epsr.2019.01.034.