"This has to be the cats" - Personal Data Legibility in Networked Sensing Systems

Peter Tolmie, Andy Crabtree, Tom Rodden and James Colley
School of Computer Science
University of Nottingham
United Kingdom
{First.Last name}@nottingham.ac.uk

Ewa Luger
Microsoft Research
Cambridge
United Kingdom
a-ewluge@microsoft.com

ABSTRACT
Notions like 'Big Data' and the 'Internet of Things' turn upon anticipated harvesting of personal data through ubiquitous computing and networked sensing systems. It is largely presumed that understandings of people's everyday interactions will be relatively easy to 'read off' of such data and that this, in turn, poses a privacy threat. An ethnographic study of how people account for sensed data to third parties uncovers serious challenges to such ideas. The study reveals that the legibility of sensor data turns upon various orders of situated reasoning involved in articulating the data and making it accountable. Articulation work is indispensable to personal data sharing and raises real requirements for networked sensing systems premised on the harvesting of personal data.

Author Keywords
Networked sensing systems, personal data, privacy, articulation work, accountability, ethnography.

ACM Classification Keywords
H.5.3 Group and Organisation Interfaces; Computer Supported Cooperative Work.

INTRODUCTION
Digital applications and services increasingly rely on personal data. Indeed some authors go so far as to suggest that we are in the middle of a "personal data gold rush" [8], with advertisers, Internet services, and governments alike all accumulating increasing amounts of personal data about us. The accumulation of personal data is not passing without notice. An ongoing 'drip' of stories invoking inappropriate uses of personal data accompanied by national security 'leaks' have attracted broad media interest and have in turn engendered widespread societal concern and debate about privacy.

CSCW '16, February 27-March 02, 2016, San Francisco, CA, USA

ACM 978-1-4503-3592-8/16/02…$15.00
DOI: http://dx.doi.org/10.1145/2818048.2819992

Our interests with regard to personal data and privacy are circumscribed by a concern with large scale personal data harvesting driven by ubiquitous computing and networked sensing systems; interests that are otherwise encapsulated by slogans such as 'Big Data' and the 'Internet of Things'. Big Data analysis has proved to be remarkably effective at revealing patterns or clusters that correlate to human behavior - credit card transactions patterns that predict the likelihood of divorce [2] or lifestyle measures that predict the likelihood of death within five years [6], for example, not mention the correlations that increasingly drive digital marketing [10]. The existence of large data sets has led to a strong business interest in the use of statistical techniques to predict future behavior across a broad set of domains [17].

The power of predictive analytics associated with Big Data has also fuelled a growing interest in the development of richer pictures of human behaviour through *fine-grained* data collection enabled by the Internet of Things. The promise here is the ability to build an extremely detailed picture of human behavior that moves beyond the predominant approach of clustering and correlation. As Pentland [13] puts it,

> "It's not about the things you post on Facebook, and it's not about your searches on Google, which is what most people think about … This sort of Big Data comes from … the little data breadcrumbs that you leave behind you as you move around in the world. What those breadcrumbs tell is the story of your life. It tells what you've chosen to do. That's very different than what you put on Facebook. What you put on Facebook is what you would like to tell people … We are moving from the reasoning of the enlightenment about classes and markets to fine grain understanding of individual interactions and systems built on fine grain data sharing."

Our interest in personal data and privacy is in the breadcrumbs that are often captured without us realising it and the concomitant assumption that fine grain understanding of interaction, and with it our lives, can be had from sharing these. We explore the assumption through an ethnographic study of a networked multi-sensor package deployed in three homes in the UK. The study reveals that personal data generated through this networked sensing system is *opaque* when considered in isolation. This means that fine grain understandings of interaction cannot be 'read off' the data alone. Rather, the real world, real time sense or legibility of the data trades on 'articulation work' [16], the

practical indispensability of which impacts understandings of privacy in a domestic context. Far from opening up personal data on an unprecedented scale, networked sensing systems appear to lack the necessary transparency. Mechanisms are required then to support articulation of the data, which may reflexively enable privacy management.

We elaborate key features of articulation work, particularly how the intelligibility of personal data turns upon an understanding of the domestic routine and the taken for granted orderliness of domestic life. While derived from a domestic context we would argue that these issues generalise to other contexts. After all, it is not only the home that is characterised by routines and an underlying social order that shapes what goes on, and what may be recorded, about a setting and its activities.

Articulation work is of course an often-cited concept in CSCW, at least in early work in the field, and key to Schmidt and Bannon's elaboration of what makes CSCW distinctive [16]. They attribute the concept to Anselm Strauss and his studies of work and organisation. Strauss describes articulation work as a "supra-type of work" which involves "the accountability of actions" [18]. To appreciate what he is driving at here, Strauss construes of the "totality of work tasks" in an organisation as making up the "arc of work". The arc is "not automatically articulated, actors must do that". Actors are thus "rendered accountable" for accomplishing the tasks they are responsible for and "reporting" the accomplishment to others. The "reporting" is an overhead (a supra-type of work) that is not the work itself (the job as it were) but key to its coordination within the overall arc.

Articulation work speaks then to the work of "reporting" what has been done - where in the work we are now, what has been done, what needs to be done next, etc. – i.e., to making the work accountable to others so that they can "see" where in they work they are and "mesh" their actions into the work in a coordinated manner. Accountability thus speaks to the manifold ways in which people make action *observable and reportable.* Schmidt and Bannon extend the concept of articulation work beyond "formal organisational structures" to any cooperative arrangement:

> "The conception of cooperative work suggested here does not assume or entail specific forms of interaction such as mode and frequency of communication, comradely feelings, equality of status, formation of a distinct group identity, etc. or even specific organisational settings. Indeed, we do not want to restrict the scope of CSCW to cooperative work relations that are defined and bounded in legal terms, i.e. in terms of formal organisational structures. Cooperative work is constituted by interdependence in work … Thus, the boundaries of cooperative work networks are defined by actual cooperative behavior and are not necessarily congruent with the boundaries of formal organisations." [16]

Our study suggests that insofar as some arrangement of cooperative work is *necessarily* implicated in the harvesting of personal data [5], whether it involves end-users and service providers or as in our case researchers, articulation work is required to derive fine-grain understandings of interaction and to build systems on this basis. Our study elaborates the various orders of reasoning implicated in articulating sensor data to third parties and the methodological ways in which the data is made accountable. This, in turn, raises requirements for the support of articulation work and the design of communication, visualisation, annotation, and collaborative data management tools.

THE STUDY

Building on prior deployments of network and sensor-based technology in domestic environments [1, 3, 4, 12, 19], this study reports on efforts to develop a prototype platform for capturing a diverse collection of data in the home to support an open and scalable 'multi-sided market platform' capable of serving both the needs of residents and future service providers. The platform aims to consolidate data from sensors, which can be embedded into household items, as well as from a range of Internet-enabled devices [9].

A key issue in developing this platform is to understand how people might live with fine grain data collection and orient to and understand sharing it. To that end, a range of sensors were deployed over a 2 month period in 3 households to permit initial investigation of the matter through detailed ethnographic work. The deployed sensor package contained AEON Z-Wave Multi-sensors to monitor temperature, humidity, light levels and motion; wireless solar power temperature sensors; and a current transformer sensor to monitor the rate of electricity consumption for the property. Additionally, a Raspberry Pi with a Z-Wave daughterboard was installed to receive and store data from the wireless sensors, and some simple visualisation techniques were used to display sensor data to participants on a basic time-rate graph (Fig.1).

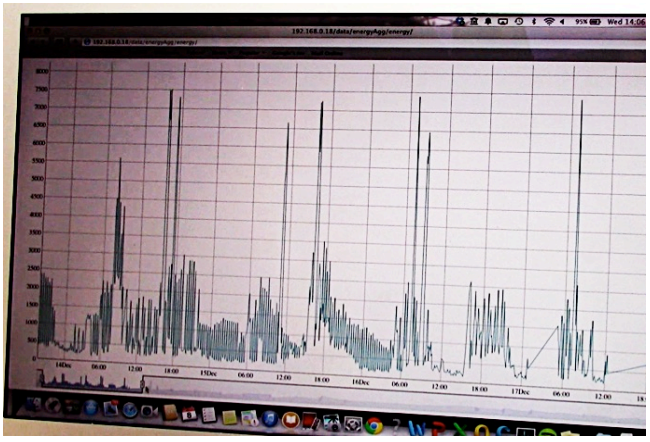

Figure 1. Displaying Sensor Data

The sensors were deployed in four principal areas within the home: bathrooms, kitchens, main living areas, and bedrooms. The energy consumption sensor was located wherever the electricity meter was situated. The Raspberry Pi was placed where it could best capture the output from

the various sensors wirelessly, something that itself led to re-positioning of sensors to find the best possible positioning.

Because of the anticipated sensitivity of the data it was decided to deploy the sensor packages in the homes of project members. These are real homes nonetheless, possessed of and exhibiting the ordinary day-to-day business and routine concerns of domestic life that can be found elsewhere irrespective of the particular project members involved. The households were predominantly professional in character. In one lived a couple where one was retired and the other semi-retired. In another, a family with one daughter still living at home and one of the adults retired. And the third, a family with two younger children, of 10 and 7 years old.

In each of these homes the participants were not computing experts – the project was interdisciplinary in character, not only involving computer scientists - though two of them had previously purchased off-the-shelf sensors out of personal interest. None of the participants therefore had anything beyond what might be termed a 'lay expertise' in using sensor-based technology. It should also be stressed that each of the households involved, despite their broad professional character, were quite distinct from one another in terms of both routine and disposable income. Additionally, it should be noted that the households involved were not those of the researchers who played a part in the installation of the senor packages or gathering of the ethnographic data. So none of them had any special relationship with the ethnographers that might have made them orient to the ethnographers as anything other than strangers with a relevant interest, the interest in this case being that of conducting research.

The studies of the deployments were phased across a number of steps. Before any observational work was undertaken members from each of the households were interviewed in order to get a general sense of the layout of their home, the people who lived within it and their preoccupations, and the kinds of routines that tended to shape domestic life. The interviews were recorded and followed up by deployment of the sensor package and data logger. At about the same time an ethnographer visited each household and met all of the inhabitants and conducted a brief tour of the house to get a sense of the kinds of locations sensors would be positioned in and the kinds of activities that tended to unfold there. The tours were videoed and the participants were also encouraged to keep online diaries in which they could record significant events within the household, such as people visiting or members being away, etc., so that shifts in patterns within the data might be inspected for and correlated with those events.

After the deployment had been in place for 3 to 4 weeks an ethnographer revisited the houses and used the visualization tool (Fig. 1) to talk through the data and capture on video the participants' accounts of it. In doing this, it was understood that the intended purpose of collecting sensor data was to share it with third parties. 'Talking through the data' thus provided an opportunity to investigate something of what is involved in sharing personal data with third parties (us) and, importantly, what sharing turns upon. A further visit was conducted after another 4 weeks to assess whether or not the prior visit had resulted in participants changing their domestic activities, and then the sensor package was removed. Analysis of these materials was undertaken using an ethnomethodological approach [7] to explicate the *accountable* features of personal data sharing.

Below we look at the ways in which the data gathered by the sensors could be seen as revealing certain *orderly* characteristics of the household and how both the participants in the study and the researchers working with them sought to arrive at accounts of the data in these terms. A critical point here is the apparent *gap* between what is captured by the sensors and what is necessary to render the data locally and socially meaningful. We then examine more closely the work participants engaged in to make the data accountable and how this is possessed of certain *methodological* features, before looking at what all of this might mean for design.

FINDINGS

In this section we work through the study findings to examine just how people *reason* about sensor data and how reasoned *accounts* of that data are reflexively produced. We begin with relatively basic reasoning about some gross spatio-temporal features of the home, about the people who inhabit the home, the way those people usually go about doing things, and about the kinds of events that ordinarily unfold within people's homes. This is followed by consideration of how these matters are taken to speak more broadly to how the home is constituted as a 'socially organised' affair, such as what is understood to be the home's routine, how that routine might be breached in various ways, and how the social order of the home is oriented to ultimately as a moral order, through and through. These various considerations are elaborated through a handful of perspicuous examples, but it should be understood that *all* of the great many accounts offered by the participants in bridging the gap between sensor data and its local and social meaning speak to one or more of these matters in some way.

Reasoning about place

It is no great surprise that the spatial configuration of people's homes is oriented to as a meaningful configuration that carries with it common-sense understandings and expectations regarding just how that space is organised. Thus it is equally no surprise to find that a certain amount of how sensor-based data gets reasoned about is attached to those understandings and expectations. Here we have a case in point:

> Connie: (Inspecting output from the motion sensors over the Christmas period). December 28th, 29th, lots of motion activity. Samantha and the family were all here.
> Connie notices motion in the upstairs bathroom. She suggests that all the other bathrooms were probably being used so she might have gone upstairs instead.
> Connie: I might have used it.
> Andrew: To run away from them.
> Connie: Yeah, probably (laughs) yeah; taking refuge.
> Connie: I took a shower around 11o'clock that day, as I had come down in the morning to cook. Then I got changed. So I was using the bathroom a lot that day.

What is visible to the sensor here, and made manifest in the graphical representation, is motion of a certain duration on particular days at particular times; additional data includes the location of the motion sensor through the way its input is labeled by the Raspberry Pi collating the data. By contrast, the husband and wife who are talking through the data with the researcher are actively making comparisons with other things they can see on the graph, such that they might appropriately describe what they are seeing as 'lots of motion activity' and even exceptional activity in the upstairs bathroom.

In searching for how to account for this they come up with the general gloss that this was a period over which family members were present in the house, rather than it just being the two of them. Seeing that a measure of the motion activity is visible in the upstairs bathroom Connie comes up with an account for this regarding other bathrooms being in use. This makes manifest an understanding that a) the upstairs bathroom is not usually used at that time of day and b) that this therefore needs to be accounted for. The account itself turns upon already having explained that there are exceptional numbers of people in the home and it is understood that unusual numbers of people using the bathroom at the same time might appropriately lead to using other bathrooms that are not usually used. There is an added half-joking interchange about escaping the other people present in the house, which rides upon that also being seen as a morally defensible course of action. Finally, Connie elaborates and underscores the account further by presenting rationales for why she was 'using the bathroom a lot that day'.

So some clustered peaks on a graph have provided for the following: a) seeing that they need to be accounted for; b) seeing what surrounding situational circumstances might have provided for them (understood to be people moving around); c) seeing how the peaks are associated with specific locations (in this case an upstairs bathroom); d) seeing that this requires a more specific account than the general account so far provided; e) coming up with such an account that trades upon what anyone knows about bathrooms and their use, and what might reasonably be done if one bathroom is in use but another one is available; f) seeing that there are additional, elaborative accounts available regarding what anyone knows about how bathrooms might be used to get some time away from people in a crowded house (because who could reasonably object to you going to the bathroom?); g) recognising that this is a somewhat fragile account that presents a certain attitude towards family members that not everyone might deem appropriate, and thus providing a further elaborative account regarding what anyone knows about having to fit in ablutions around a charged domestic schedule occasioned by 'having visitors'.

Thus the peaks on the graph are accounted for, but note just how much *social understanding* is wrapped in the accounting and how much it trades upon a body of common-sense knowledge about bathrooms and how they might be used. The gap between what is available to the sensor in terms of motion x at time y in location z and this *exercise* of social understanding is startling.

Reasoning about time

It is equally the case that an ordinary way of understanding what kinds of phenomena might be indicated by sensor data is to open up to inspection just *when* things occurred and to seek to account for them in terms of the *temporal order* of the home. This is, of course, a feature of the reasoning visible in the preceding example, where exceptional use of the upstairs bathroom turned upon seeing the time. Below we can see a similar order of understanding, but this time regarding how bathroom use is wholly unexceptional:

> Connie and Andrew are looking at the humidity data. They see regular 'spikes' on 23/24th January at about 9.00am. Connie says that that's right, as they often have a shower after breakfast.
> Andrew: I bet you can tell the difference in showers between you and me, because of the amount of shower. I shower for a lot longer than you do.
> Connie: And at a higher temperature.
> Andrew: I bet the humidity comes up there.
> Connie: But you weren't there on the 23rd of January.
> Andrew: Oh that's right. There's only one of those. There's you.
> Connie: I must have thought thank goodness, I'm going to have a really long nice shower (laughs).

Here the sensor is seeing different quantified measures of humidity at different times on different days and the sensor is recognised as one associated with a particular bathroom. The representation of this shows different levels of humidity over time and therefore makes it is possible to see 'spikes' (and troughs). The first step towards an account for the household members here lies in seeing an unexplicated connection between 'spikes' in humidity in a bathroom and the possibility of equating that to someone having a shower. Thus place is an integral feature of members' reasoning, but timing is also crucial because it allows them to determine that a shower at this time amounts to business-as-usual.

Time is crucial in interesting ways, for wrapped up in this account is seeing that the time – 9.00am – is *subsequent* to them having breakfast: the account here is not that they usually have showers at 9.00am; it's that they usually have

showers *after* breakfast. It is also interesting how calendar time comes to play a part in their reasoning. Andrew makes the proposition that they should be able to see which one of them is showering in the data by seeing different spikes in humidity, and Connie concurs with this. However, Connie then points out that the day they are looking at is a day when Andrew wasn't there. This then presents the need for another account because suddenly what was seen as normal comes to have the character of an exception: Connie has managed to provoke a similar peak all on her own. Thus she jokingly suggests she must have decided to have a longer shower as he wasn't there, an account that is both appropriate and inappropriate, and therein lies the joke.

So what we have here regarding matters of time is members using its availability in the data as a resource for discovering *the routine* and for discovering ways in which the routine might be seen to vary but nevertheless be explicable. This isn't something that just falls out of the measurements coming from the sensor. It takes local understanding of the 'patterned' character of their lives to see this in the data. More importantly still, it takes local competence to be able to see how apparent anomalies in the data still amount to wholly ordinary courses of action.

Reasoning about people

Something else that becomes visible in the way people account for the data relates to what they know, as members of the household, of one another's routine doings and *who* might therefore be responsible for some particular piece of data looking the way it does. Indeed, we have already seen certain aspects of this kind of reasoning in the preceding examples regarding just who might be responsible for unusual activity in a bathroom at a time when there are guests and whose data might look like what as a consequence of how long they routinely stay in the shower. Consider also the following:

> Becky is looking at the temperature and humidity data in the kitchen. There are clear places where the figures are rising: Yes, it goes up when he's boiling things … days when Ron does the stews or soups. If he was frying or baking it goes up. You can see when Ron comes home.

So what's visible to a sensor in this case is rising and falling temperatures and amounts of humidity in the air, together with the times such fluctuations occur, the extent to which they continue at certain levels for certain amounts of time, and (perhaps) the extent to which these things coincide with one another. Allowing for the coded-in indicator of the location of these particular sensors one can imagine that the sensor output might also provide for some crude allocations of possible activity, such as cooking, given that there are only so many things that are likely to be productive of heat and humidity in a kitchen. Note here, however, that the household member sees more than this: she sees *who* is doing the cooking and how cooking fits in with Ron's presence in the house and *his* routines.

None of this is visible in the output of the sensors; even a simple algorithm that might recognise 'cooking' as an activity does not see the 'who' of the matter. Thus we can see here how members are able to account for the data in terms of what they know of one another. That they are predisposed to see the *person* in the data, such that it is not just about cooking or showering or going to the loo, but about just who is doing these things as intelligible features of the organisation of their home. Furthermore, insofar as these are *locally* accountable matters, turning upon members' common-sense knowledge of 'what usually goes on round *here*', we can see that the challenge for sensing is not just one of getting a fix on social understandings. More than that, it is a matter of being able to arrive at what the members of *just this* particular setting know so that they can see not just cooking (etc.) but how the cooking (etc.) is indexical to *specific* people and their routines.

Reasoning about practices

Something noticeable in the preceding sections is the way in which sensor data can get tied up with reasoning about not just what members of the household do, but also about just *how* they do them. On occasion the data quite expressly brought to the fore just how certain people undertook certain activities, making them specifically accountable for the fashion in which they were done:

> Connie and Andrew are looking at the light sensor data for the bathroom. Connie expresses surprise that they leave the bathroom light on. She feels that they leave Andrew's study light on a lot and the kitchen lights.
> Andrew: I turn the light on when I'm having a shower; I noticed that the other day.
> Connie: Oh, even when it's light?
> Andrew: Even when it's light. I like having a really light shower, makes me feel wide awake."

The sensor here is detecting light levels and, once again, these are being mapped against time. The software is also able to identify exactly which light sensor is recording the data and this has been pre-labeled as coming from the upstairs bathroom. So, at its most basic level, the graph can be seen to show when it is day and when it is night and when the light in the bathroom is switched on or off. Though, in a sense, all the sensor is seeing is more light or less light plotted against time, it is easy to see how this data can be further refined to categorise the data in terms of basic lighting events. What the members reveal in their accounts here, however, is much more than just lighting events.

First of all the duration of the event is problematised by Connie because it does not meet her expectations. The very fact that it is problematised implicates the production of some kind of account for the anomaly. The problem here revolves around the fact that this is not so much a one-off anomaly as a routine anomaly: the bathroom light is being 'systematically' left on for longer than expected. But it turns out that Andrew has a ready-to-hand account: it is his routine practice to have the light on when he is showering,

'even when it is light'. This is a practice that Connie evidently knew nothing about and, having revealed a hitherto unknown practice, Connie renders it remarkable by posing the question, 'even when it's light?' Andrew is now under obligation to produce a further account for this particular practice. This he duly does by explaining how he likes a 'really light shower'.

The leap from light data to the adequacy of a local account, such as a particular individual liking to have 'light showers', is substantial and embedded in a body of social reasoning about how specific showering practices might be adequately accounted for that goes far beyond anything that a sensor might be capable of seeing. However, this particular example also brings into view another matter of substance. What has happened here is that the production of the sensor data and its inspection has made available to another party in the same household an individual practice that they previously *knew nothing about*. 'Having a light shower' is, of course, no grand crime, though it is possible that some further discussion about the matter may have subsequently ensued and clearly it could be argued that Andrew has been 'found out' in wanton consumption of electricity.

The deeper point here is that sensor data has the potential to *make visible* to others with whom you are quite specifically related matters that might otherwise be *invisible* to them. Thus sensor data has the potential to make household members' activities and practices accountable *to each other*, not to a third party who, importantly, might not recognise these as particularly accountable matters at all. It is of note here that it wasn't the researcher Andrew had to provide an account to, it was his wife. There are numerous everyday activities and practices we engage in that we don't 'broadcast' to those we live with and don't ordinarily expect ourselves to be accountable for. The presence of sensors breaches this visibility of action and presents householders with a whole new set of challenges regarding how they might understand themselves to be routinely monitored, not by governments or employers or even by marketing organisations, but by those with whom they have the most *intimate* of relationships.

The added twist here is that this cuts both ways: people are not just accountable for activities and practices that might now be opened up to being monitored, the people doing the monitoring are accountable for their actions as well. There are ordinary rights and expectations regarding just what you might appropriately watch and notice, and sensor data in the home has the potential to breach these. The following example is instructive:

> Susannah (talking about the children and the sensor data): So, I'm aware that there's evidence that Sally's gone for a wee. She's spent most of her life trying to define her space, and David is defining his space. So now there's evidence - now I can see into these spaces - so there's a sense of invasion. I can now look and find out who went for a wee when and where they went.
> Frank: You know when Sally's been there because she always puts the toilet roll on the floor.

There is, then, serious potential for sensor data breaching members' ordinary everyday practices of *disregard* and making an invasive step into people's personal relationships by making things accountable where previously there was *never* a need for an account arising at all. The potential fallout is readily imaginable.

Reasoning about events

Ultimately all of the above aspects of reasoning we have discussed get framed within a broader consideration of just what kind of activity or event might be indexed by the visible data. So the data is seen to point to *showers* happening in bathrooms at certain times of day or in certain kinds of ways, or *cooking* being done by certain people. Sometimes the events proposed within accounts of sensor data display a quite detailed understanding of just what might be going on to provoke the generation of certain kinds of data, and understanding this involves a profound concatenation of all of the other kinds of reasoning we have talked about so far:

> Connie is inspecting the output from the temperature sensors in her kitchen. She notices that there are regular temperature drops in the kitchen on certain days.
> Connie: Andrew leaves the garage door open when he's working out there, when I'm not there to keep shutting it. He forgets doors are on hinges. Born in a barn really. (The garage door opens onto a corridor that leads directly into the kitchen)

In the above example temperature sensors are capturing changes in temperature at certain times and on certain days. All that the sensor can see and all that is visible in the graphical representation is a lower temperature for a certain period of time in a location that is identified as being the kitchen. This relatively simple piece of data, however, produces a deceptively simple piece of accounting that turns upon understanding a whole range of matters about who does what in certain places at certain times in the home and in certain ways, and how that might impact upon measurable matters such as temperature in other locations.

So what Connie sees within the data are the following considerations: several days a week Andrew likes to do DIY activities in the garage; Andrew routinely leaves the garage door open when he's doing so; Connie has called him to account for doing this but he has continued to do so; it therefore falls to Connie to shut the door, but she's not always there to do so; the corridor the garage door opens onto is also connected to the kitchen; leaving the garage door open makes the corridor cold and this, in turn, makes the kitchen cold; the days and times when the temperature drop happens in the kitchen coincide with the days and times when Andrew does DIY in the garage; the cause of the drop must therefore be his practice of leaving the garage

door open. The account given does not by any means render all of these matters explicit but, despite this, it is proffered as an adequate account.

So in a single sentence account Connie brings together reasoning about places and times of events and the people who are enacting them and their practices. The account is adequate without her subsequent comments that make clear her own position on the matter: that it is oriented to as adequate is visible in the way it self-evidently glosses a whole constellation of matters pertaining to the household order, its inhabitants, the activities they engage in, and the discernible events these create. It speaks of a *social order* that it is presumed the recipient of the account, the researcher, can see too without any need for further elaboration. But social order is not something available to a sensor and it turns upon small details of local practice of which the sensor has no sense.

Reasoning about the routine

Of course, what underpins much of the reasoning visible in the preceding materials is the profound understanding members of a household exhibit regarding what constitutes the *routine* within their household: the things that happen in certain places and at certain times involving certain members of the household doing things in particular kinds of ways, in a regular and taken-for-granted fashion such that everyday life may proceed on an even keel. Sometimes discussion around the data made this deep appreciation of how the participants' lives were 'patterned' in certain ways especially visible:

> Susannah and Frank are looking through the energy consumption data for weekday evenings.
> Susannah: On a Monday - so Frank would pick David up - so this is Frank being in and then I would join him. So this is probably him getting sorted with some tea and then putting kids to bed and then …
> Frank: It's quite quiet I guess isn't it, putting the kids to bed?
> Susannah: Yeah I know but then what are they actually using. So that's lights on and off, and that might be - we've got a projector rather than a TV - so that might be this bit here … Now in the overnight the kids have a light on in the landing 'cause they don't like it to be completely dark.

So here all that is being looked at is a graph showing varying levels of energy consumption over time. It is, in a sense, nothing more than a set of numbers on axes and lines between varying points angling up and down. The sensor has produced a record or trace of changes in the amounts of energy being drawn upon by the household at the point where it enters the house and it has been rendered open to depiction in this way. What is impressive here is the extent to which such a simple resource can provide for a description of family life and its organisation, and in so doing illuminate the measurements in ways that are *not* visible within the measures themselves.

Members are able to point to specific points where the graph rises or falls and express them as points within the routine: this is probably him getting sorted with some tea; it's quite quiet … putting the kids to bed; that's lights on and off; we've got a projector … so that might be this bit here; overnight the kids have a light on in the landing. Here we have a snapshot of life in this household from the moment they enter it after work or school up until when the kids are in bed. But the snapshot is a property of the description, the account, of everyday life and its organisation, *not* the graph.

It trades on what *anyone* knows about the enactment of daily life in the home. There is no sense of bringing into question or needing to account for *why* Frank would be getting tea, why he would be putting the kids to bed, why the projector might have been used after the kids have gone to bed. Only leaving the light on in the landing is subject to further account, and this itself trades on what anyone knows about children not liking the dark. What anyone knows is not represented in the graph however, less still in what anyone knows round *here*. It resides in the conduct of domestic life and its routine organisation, and is manifest in *occasioned* talk when, for example, researchers ask members to account for the energy represented in graphs. Thus, in making the data accountable members understand that specific activities and practices will have particular energy consumption consequences. They understand too, without having to be told, that it is through spotting those consequences and imbuing them with an ordinary social significance that they inevitably make the routine visible.

Reasoning about exceptions

The flip side of members' reasoning about their routines and its use as a resource in accounting for sensor data is the recognition within the data of what might therefore constitute an *exception* to the routine. Reasoning about exceptions is often generative of quite a significant amount of work, as is visible in the following example:

> Frank and Susannah usually leave the house for work around 07.40. However, looking through the motion data they can see that the sensor detects movement in the house in their absence.
> Frank: OK, so I left the house and then there's some activity at 12.00.
> Susannah: So this has to be the cats because we're not in.
> Frank: Could be cat activity.
> Susannah: That has to be the cats. It can't be anything else can it?
> Frank: No, I don't think so.
> Susannah: Although that first one's longer isn't it? How long in duration is that? About eight minutes?
> Susannah: Generally, you see, my perception of what the cats do when we're not here – Fred – like this morning I took the car down to the garage at nine and he's asleep on David's bed and there he will remain, apart from getting up to have a look round for food. He'll get up if there's somebody that comes into the house but otherwise

> he'll stay asleep. But as I was coming down the stairs, Moomin shot up them, so I thought I bet she's going back to bed. So, they'll get up when we get up, but then as soon as we leave I do think they go back to bed then have periods when they might get up sometimes."

As with the discussion regarding bathroom usage earlier the data being examined here is motion data where a sensor has detected motion at a certain time of day and this is what is visible on the graph. In this case the data being examined relates to data from several motion sensors within the house rather than just one, but once again it is seen to require explication because the timing does not fit readily with local understandings of the day-to-day routine.

What is interesting here is the amount of work involved in making exceptional data reason-*able*: that is, the appeal is no longer to what just any member might take for granted about the routine, but rather to what might sound appropriate as a body of speculation, based on a set of highly specific and local rationales. This takes more effort than usual because it requires more detail than an account that trades on what anyone knows.

The specific problem the members are confronted with here is that there is motion in the house when there's nobody home. It takes inspection of their lives and the particular composition of their household for them to see that there are only so many candidate possibilities and it is no surprise that they arrive at the cats as causing the movement given their absence. However, convincing both themselves and the researcher of the *adequacy* of this proposition involves working through what they know of the cats' habits before they are able to arrive at the conclusion that this accounts for the exception and, in doing so, the example makes the curious nature of exceptions visible.

Exceptions are at first glance remarkable. To remark upon them is to make visible that they are not readily accountable, and if they are not readily accountable then there's a job of work to be done to understand them. Curiously, that work involves bringing the remarkable issue at hand into line with ordinary, run of the mill happenings in the home. So what at first appears remarkable turns out not to be particularly remarkable at all; just the cats doing what cats do, for example. This work of rendering the exceptional ordinary is something we return to again below.

Reasoning about the moral order

Above all of the other matters we have so far discussed is the way in which accounts for data are grounded in what amounts to a *moral* organisation of the home. That is, members of any household have a sense of what *can* and *cannot* appropriately be said about their activities and what kinds of rights and obligations such accounts are demonstrably attentive to. All of the preceding materials are imbued with this moral understanding in one way or another, as indeed would be any account of the household and its doings.

Where accounts relate to the accomplishment of the ordinary everyday routine, members display no sense in which the pursuit of their activities requires *overt explication*. Rather, the statement of the routine is presumed to be an *adequate* account in its own right. The exception is exceptions. Here we see the taken for granted orderliness of domestic life being surfaced and made explicit. In the following example we return once again to the matter of appropriate bathroom practices, with one of the members of the household being called quite specifically to account as a consequence of data being made visible:

> Frank and Susannah are looking at the light and humidity data for their bathroom; a sudden peak in activity in the middle of the day on a weekday becomes remarkable:
> Susannah: That was Mary being off, and the 6th was you being at home. Oooh, what did you do?
> Frank: I didn't do anything.
> Susannah: You did. At 12 o'clock. Look at that.
> Frank: Where? Nothing.
> Susannah: No, here.
> Frank: I could have been up late, 'cause I've had this headache thing. So that's probably me getting up late isn't it? Having a late shower. It's high for a long time. I don't have that long a shower.
> Susannah: Yeah, but you could have had a shower and then you could have had a shave.

So here what is being sensed is light and humidity in a bathroom. As in all of the other cases this is being visualized on a simple time-rate graph, this time showing simultaneous increases in both light and humidity levels at a certain time and on a certain day of the week. This is all the sensor can see. However, what is visible to members is the way this indexes bathroom activity at an unusual time of day. More importantly, the activity is visibly connected to a specific member of the household, even though there is nothing in the data itself to identify any one particular person. It then becomes *incumbent* upon that specific person to come up with an account for why the data might look that way.

First of all Frank accounts for the matter by saying it might have been him getting up late because he had a headache. This, of course, is no kind of explanation of why there should be a light and humidity event in the bathroom, but already the very notion of getting up late pays testimony to his understanding of what the routine *should* be and provides acknowledgment of its status as an exception. He then ties the account more specifically to the visible data by proposing he could have been having a late shower. However, immediately after this he starts to question the adequacy of this account because the humidity and light levels stay high for a 'long time' and he doesn't have showers that last that long. This itself is bound up with a sense of what an appropriate amount of time to have a shower might be for this person in this household, underscoring the moral implicativeness of this event. Having a 'long shower' could be described as inappropriate *here*. Susannah, however, rescues his account by suggesting

he might have followed on from having the shower with a shave.

Underlying all of the more explicit reasoning here are a number of background understandings, including: a) what kind of things make the bathroom humid; b) what kinds of bathroom activities this person might engage in (e.g. shaving); and c) how that might provide a reasonable account of a phenomenon that at first sight seemed unusual and *demanded* accounting for. As observed recurrently over the analyses presented above, the work engaged in here is tightly concerned with making 'what the graphs say', and more specifically what the graphs render visible and even remarkable, an *ordinary* feature of everyday life.

Accomplishing this is utterly dependant upon membership of the household. This is because: 1) It takes membership to see what might or might not count as being 'ordinary' or 'exceptional' in 'this' particular household in the first place; in other households showers at 12 o'clock might count as business-as-usual, but not 'this' one, for example. 2) It takes membership to be able to see how the remarkable might be rendered intelligible in 'this' particular household. 3) It takes membership to be able to see what an *appropriate* account would need to look like for this intelligibility to be appealed to and invoked.

The gap between what is captured by the sensor and the kinds of moral reasoning being displayed here is substantial. However, there is a deeper challenge: the moral assumptions around which people organise the sociality of their homes form a *rarely explicated* backdrop to the things that are called to account. It takes the rather odd circumstance of a researcher (or some other third party) sitting down with a family to talk through graphs, which display data rendered from sensors that are putatively indexical to the things happening in their home, for such explications to be even vaguely forthcoming. In ordinary everyday life these matters remain largely *unsaid*. Explicitly uncovering accounts of what it is involved in being a member 'round here' is therefore going to be a challenging matter for anyone who has need of eliciting them; and designers have such a need if networked sensing systems are to deliver on their promise.

THE ACCOUNTABLE CHARACTER OF PERSONAL DATA SHARING

For members, the taken for granted orderliness of domestic life is largely unsaid; what need, after all, do members have to busy themselves in articulating for others what is ordinarily plain for them to see or to draw that which is usually invisible to them, and *should* be invisible to them, to explicit account? However, when the orderliness of domestic life is made an accountable matter, as the deployment of our sensor packages has done, then it becomes visible that its elaboration implicates various orders of situated reasoning. These are drawn upon to make the data generated by sensors accountable, i.e., observable and reportable. The making is a methodological matter that consists of *articulating* the data in terms of the social and moral organisation of domestic life and instructs us as to how personal data interaction might be appropriately handled. Here then, in summary, are the primary observations to attend to.

Accounts are occasioned

In the above materials we have assembled a number of perspicuous examples that together bring into view some of the methodological characteristics of accounting for personal data *to third parties*. In this case the third party is a researcher. This fact cannot be ignored [14]. It matters, then, *who* the third party is, and for *what reasons* and to *what ends* they want to pry into people's personal data. It is in this respect that we say that, as a methodological matter, the sharing of personal data is 'occasioned'. In our case the sharing is occasioned by the doing of research, whereas for others it will be occasioned by the delivery of some kind of service, which will have to be accounted for as a premise to personal data sharing just as our prying was for the families of project members.

The occasion was ostensibly accounted for by us in terms of the process of informed consent, an 'up front' process similar in kind to that which frames the use of many digital services in spelling out 'terms as conditions' as it were. We are aware of calls for new models of informed consent in systems design [11], however the point we would draw attention to here is that the 'up front' nature of the occasioning of accounts is misleading. Rather, it is the case that *each and every* account is occasioned, in our case by the researcher asking household members to inspect and talk through particular representations of the data. The occasioning of accounts is on *ongoing matter*, it cannot be settled 'up front', and this is of consequence for design.

Accounts are mutual accomplishments

As ongoing matters accounts are primarily oriented towards the business of articulating various patterns of activity and the reason-able character of exceptions. This turns upon seeing *what* needs to be accounted for, and there is a reflexive relationship between seeing this and representations of the data. This means that an account cannot just be 'read off' the data but has to be discovered *within* the interaction *between* occasioner, household member(s) and their situated use of the representation. Finding the 'just what' of the matter is therefore a mutual accomplishment of all parties to the articulation of sensor-based data.

Accounts are recipient designed

The mutual accomplishment of accounts turns upon the presumed appropriateness of accounts. It is not simply that not just any account will do - that there appropriate ways to account for what can be seen in the data based on what anyone knows of what goes round here - but that *who* the account is being told to is also consequential for what kind of account is offered about what has been seen. Articulating appropriate accounts of the data implicates various orders

of reasoning spanning the home's organisation, its routines, the activities that compose them, individual practices, and the moral probity of accounting for any of these things. *Who* wants to know about them is therefore of consequence. Thus, while there is very little to say to local members who already understand these matters, what members have to say to a researcher is a different matter, and what they might have to say to a service provider or a legal representative might well be different again; in each case accounts are and have to be 'worked up for the occasion' [15] or designed for particular recipients.

Accounts are fundamentally social in design

The recipient design of accounts is predicated on the supposition that they will need to be socially, rather than technically, intelligible. Thus data events are understood by members to be indexical to social events and accounts are rendered in these terms. Common-sense reasoning is brought to bear regarding the data made visible such that motion is seen and said to be indexical to the presence of people (or cats), for example, whilst temperature, light and humidity are all seen and said to be indexical to certain kinds of activity, and energy consumption is seen and said to be indexical to certain kinds of activity implicating certain kinds of device usage. Within these common-sense understandings are found situations and circumstances within the local order that can be and are drawn upon to appropriately account for what has been made visible.

Accounts are morally accountable

The making visible of the local order is itself an accountable matter. Simply put, the provider of an account is accountable for the account. Not only in the sense that the person offering an account may be required to make the *adequacy* of their account observable and reportable to other members as, for example, when the adequacy of Andrew's account of having the light on in the shower was called into question by Connie. In addition to this accounts are designed, as a methodological matter on the grounds of, and with respect to, moral probity. There are, then, features of the local order that are as a matter of moral compunction, to be glossed over and disregarded. There exist matters of domestic life that are *not* to be accounted for *by* members *to* members. To do otherwise would, as Susannah elaborates with respect to her children, be to make an 'invasive' intrusion into one another's lives, which in turn may disrupt the social order and potentially break it.

SO WHAT?

It might be thought that the 'rich' human analysis provided in this paper bears little relation to the 'thin' automated analysis that is typical of Big Data. It is therefore incumbent on us to explain how, exactly, the fine-grained articulation work exhibited here highlights the limitations of Big Data more generally. It is worth revisiting Pentland's comments by way of opening this up. Pentland assumes, like many working in the area of Big Data, that "little data breadcrumbs" – the kinds of data produced by sensors - can tell "the story of your life"; that "fine grain understandings of individual interactions" can be had from systems that produce these breadcrumbs, and that systems can therefore be "built on fine grain data sharing" [13]. Our study suggests that these assumptions are *problematic* in practice in a domestic context at least.

Our study makes it perspicuous that fine grain understandings of interaction cannot be simply be read off sensor data. It takes work to make the data legible, and thus 'speak about' real world interaction, and this work demonstrably implicates various orders of reasoning implicated in the local ordering of domestic life. The power of Big Data analysis becomes problematic then, for while the patterns and clusters that are detected through thin automated analysis work well for such things as energy optimisation (e.g., load balancing and demand side management), those same analytics start to break down at a local level where the patterns and clusters produced by manifold sensors are much more *nuanced*. Evaluation of the Nest thermostat, for example, documents its inability to recognise interactional 'context, situations and intent' [20], and this is just one single sensor, what of a host of them embedded in the home?

Of course it might be contended that even if each data point cannot be interpreted with a very high level of accuracy, thin automated analysis can still reveal a lot about the life of people in terms of gross patterns. Thus, the question is not if sensor-based accounts (rather than human accounts) of singular events are right or wrong, but if they allow us to understand patterns of activity that give away private information. Thin automated analysis may tell us when someone got up, the rooms they entered, the things they interacted with, the times when this occurred, etc., but it does not tell us what people are *doing* and without that it is difficult to see how systems built on fine grain data sharing are to respond *appropriately*. This, of course, is key to the harvesting of personal data in the Internet of Things. Data is not simply being harvested, it is being harvested to *do* something, to *actuate* something, to turn the heating up or down, for example.

If domestic sensor-based systems are to be built on fine grain understandings of interaction it is necessary to go beyond what sensors see - quantified measures of humidity, the rising and falling of temperatures, light levels, motion, varying levels of energy consumption, etc., mapped against location and time – to the *activities* that generate these measurements. The data generated by sensors is indexical to human action, but this is not seen by the sensors. Nevertheless, the data they generate, and any patterns that may be discerned in the data, obtain their sense from human action and the ways in which it is locally ordered.

We have a curious situation then in which the patterns derived from sensor data through thin automated analytics ignore the actual real world, real time *patterning* of domestic life as it is known, recognized and reasoned about

by household members in the interactional production of the local *order*. This makes a mockery of any claims that fine grain understandings of interaction and our lives can be had from little data breadcrumbs and it puts the building of systems predicated on thin automated analysis in jeopardy insofar as these are intended to respond to and support the localised delivery of services *in* the home (and not to it, as for example, with demand side energy management).

If we want to build effective networked sensing systems premised on fine grain data sharing then we need to *build in* mechanisms that support articulation work and enable people to make sensor data accountable to the local order. This is a non-trivial challenge, not only technically but socially, as no one is going to want to account for every moment of their day. Trying to do so would breach the taken for granted orderliness of everyday life, which leaves much unsaid and whatever is said morally accountable. Nor is there any need to; we are not suggesting that designers should seek to enable end-users to articulate their data as they have to us, only that the various orders of reasoning involved in articulating the data and the methodological ways in which accounts are assembled are matters to consider when tackling the problem.

The design challenge, then, is not one of enabling all sensor data to be articulated but of figuring out *just what* needs to be accounted for in building and using networked sensing systems and the services that will be delivered through them. Our studies provide some initial insight into what articulation mechanisms need to support to bring this about. Thus, in contrast to sensor networks 'seamlessly' harvesting and distributing personal data, mechanisms will be required that support methodological features of personal data sharing. This will involve developing a range of tools that support the ongoing occasioning of accounts, the mutual accomplishment of accounts, and the recipient design of accounts and in such ways that enable members to account for the social not the technical character of data, and allow them to exercise moral probity not only in sharing their data with others but in surfacing (or not) what is made accountable to the local cohort.

Insofar as data sharing is envisioned to be distributed, not face-to-face, then these methodological matters frame the design of communication, visualisation, annotation, and collaborative data management tools, to enable end-users and third parties to engage in (machine-based) dialogue regarding the harvesting of personal data, to visualise what is being harvested, to make it accountable to the local order, and to enable end-users to exercise due diligence in its harvesting and use.[1]

What of the claim that thin automated analysis reveals patterns of activity that give away private information? There is of course some truth in this, but the data is nowhere near as *transparent* as presumed at least where networked sensors are concerned: human action is missing from the picture. The risk then in this context is less to privacy as it is to the *misrepresentation* of peoples' lives, which may of course have serious ramifications. Putting the end-user into the loop in the ways we have touched upon above and enabling them to articulate their data may also be a way of managing this problem.

If personal data derived from sensor networks is to provide the foundation for systems predicated on fine grain data sharing in the digital home of the future, it will be necessary to actively involve people in making the data accountable to third parties to deliver services that work in the local context, 'here' in 'this' home. The work of accounting – of articulating the data – lies at the heart of *privacy management*. Members' accounts are never generic, but always situated within particular circumstances and tailored for specific recipients. Thus, supporting articulation work will enable members' to manage the flow of private information and the representations of their lives that are implicated in fine grain data sharing.

ACKNOWLEDGMENTS

The work reported here was supported by EPSRC research grants EP/K039911/1 & EP/M001636/1.

REFERENCES

1. Pat Brundell, Andy Crabtree, Richard Mortier, Paul Tennent and Peter Tolmie (2011) "The network from above and below", *Proc. of SIGCOMM WMUST*, pp. 1-6, Toronto, ACM.
2. Nicholas Ciarelli (2010) "How Visa predicts divorce", *The Daily Beast*, www.thedailybeast.com/articles/2010/04/06/how-mastercard-predicts-divorce.html [Accessed 27-07-2015]
3. Andy Crabtree, Richard Mortier, Tom Rodden and Peter Tolmie (2012) "Unremarkable networking: the home network as part of everyday life", *Proc. of DIS*, pp. 554-563, Newcastle, ACM.
4. Andy Crabtree, Tom Rodden, Peter Tolmie, Richard Mortier, Tom Lodge, Pat Brundell and Nadia Pantidi (2014) "House rules: the collaborative nature of policy in domestic networks", *Personal and Ubiquitous Computing*, vol. 19 (1), pp.203-215.

[1] We also anticipate the need to make the underlying algorithmic *operations* and various *judgments* made about data accountable to people. This will require 'opening up' the principles, approaches and analytic techniques used to develop these in such a manner that they can be articulated to ordinary people. This will also be socially and technically challenging given the need to convey information about what are often quite sophisticated mathematical and statistical techniques to people in a manner that they can understand and integrate into their practical everyday reasoning.

5. Andy Crabtree and Richard Mortier (2015) “Human data interaction: historical lessons from social studies and CSCW”, *Proc. of ECSCW*, pp. 1-20, Oslo, Springer.
6. Andrea Ganna and Erik Ingelsson (2015) “5 year mortality predictors in 498 103 UK Biobank participants: a prospective population-based study”, *The Lancet*, DOI: http://dx.doi.org/10.1016/S0140-6736(15)60175-1
7. Harold Garfinkel (1967) *Studies in Ethnomethodology*, Prentice-Hall.
8. Hamed Haddadi, Heidi Howard, Amir Chaudhry, Jon Crowcroft, Anil Madhavapeddy and Richard Mortier (2015) “Personal data: thinking inside the box”, *Computing Research Repository*, http://arxiv.org/abs/1501.04737
9. HAT (2014) *Platform for a Multi-Sided Market powered by the Internet of Things*, www.hubofallthings.com
10. Eric Hazan and Francesco Banfi (2013) Leveraging Big Data to Optimise Digital Marketing, www.mckinsey.com/client_service/marketing_and_sales/latest_thinking/leveraging_big_data_to_optimize_digital_marketing [Accessed 27-07-2015]
11. Ewa Luger and Tom Rodden (2013) “An informed view on consent for ubicomp”, *Proc. of UbiComp 2013*, pp. 529-538, ACM.
12. Richard Mortier, Tom Rodden, Peter Tolmie, Tom Lodge, Robert Spencer, Andy Crabtree, Joe Sventek and Alexandros Koliousis (2012) “Homework: putting interaction into the infrastructure”, *Proc. of UIST*, pp. 197-206. Cambridge (MA), ACM.
13. Alex Pentland (2012) “Reinventing society in the wake of big data”, *Edge*, 8.30.12, https://edge.org/conversation/reinventing-society-in-the-wake-of-big-data [Accessed 22-04-15].
14. Melvin Pollner (1991) “Left of ethnomethodology: the rise and decline of radical reflexivity”, *American Sociological Review*, vol. 56 (3), pp. 370-380.
15. Harvey Sacks (1992) “Verb selection; interactionally generated invitations; adequacy of local environments, etc.” *Lectures on Conversation* (ed, Jefferson, G.), Vol. 1, Pt. VII, pp. 787-801, Blackwell.
16. Kjeld Schmidt and Liam Bannon (1992) “Taking CSCW seriously”, *Journal of CSCW*, vol. 1 (1), pp. 7-40.
17. Eric Siegel (2013) *Predictive Analytics: The Power to Predict Who will Click, Buy, Lie or Die*, Wiley.
18. Anselm Strauss (1985) “Work and the division of labour”, *The Sociological Quarterly*, vol. 26 (1), pp. 1-19.
19. Peter Tolmie, Andy Crabtree, Stefan Egglestone, Jan Humble, Chris Greenhalgh and Tom Rodden (2010) “Digital plumbing: the mundane work of deploying ubicomp in the home”, *Personal and Ubiquitous Computing*, vol. 14 (3), pp. 181-196.
20. Rayoung Yang and Mark W. Newman (2013) “Learning from a learning thermostat: lessons for intelligent systems for the home”, *Proc. of UbiComp*, pp. 93-102, Zurich, ACM.