Smart New World: An Enquiry into Social Acceptability of Future and Emerging Autonomous Systems for the Home

Tommy Nilsson

Thesis submitted to the University of Nottingham in partial fulfilment of the requirements for the award of the degree of Doctor of Philosophy

November 2019

Tommy Nilsson. *Smart New World: An Enquiry into Social Acceptability of Future and Emerging Autonomous Systems for the Home*. Thesis submitted to the University of Nottingham in partial fulfilment of the requirements for the award of the degree of Doctor of Philosophy, © November 2019

Supervisors:
Andy Crabtree
Joel E. Fischer
Boriana Koleva

To my grandpa

Abstract

The convergence of multiple technologies, including sensors, embedded computing, real-time analytics and machine learning, is in the process of enabling system solutions capable of perception, decision-making and actuation without depending on direct human control or oversight. A widespread adoption of such autonomous systems has been suggested to make our lives more efficient, convenient and healthy, with the domestic context being frequently described as one of the main potential growth areas in this domain.

However, it is important to treat such predictions with a degree of caution as they are based on mere extrapolations from the current trends in technical progress that fail to take into account (known and unknown) non-technical pitfalls that might yet impact mainstream adoption. Indeed, research indicates the adoption of domestic autonomous systems is already being hampered by a variety of critical issues including an impaired sense of trust, privacy and security.

This thesis elaborates such *human* challenges further by exploring the manner in which prospective users relate to the notion of having their homes augmented through autonomous systems, along with the key points of friction this might give rise to. It makes two contributions. The first one is of a methodological nature and focuses on the use of envisioning approaches centred around provocative and deliberately contrasting scenarios of future domestic autonomous systems. These are used to drive an innovative workshop-based approach to "breaching experiments"; a sociological research method employed to surface tacit and unspoken background expectancies implicated in the organisation of everyday life. We argue that understanding and respecting such background expectancies is instrumental for the design of socially acceptable future autonomous technologies.

The other contribution is substantive, produced through participants' efforts to repair the incongruity or "reality disjuncture" created by the contrasting scenarios. Our analysis of workshop participants' responses to the scenarios revealed

background expectancies centring on 4 key themes in the form of *coordination* and the need to build the human into the behaviour of autonomous systems, *control* and the need to protect human agency, *computational accountability* and the legibility of autonomous system behaviour, as well as *social accountability* and the compliance of autonomous behaviours with social norms. We conclude by arguing that each of these key topics is formed by and brings with it a range of ancillary requirements that impact the adoption of autonomous systems in the home. This in turn opens up design possibilities enabling developers to gear autonomous domestic systems in with the non-technical expectations that govern their uptake in everyday life.

Acknowledgments

I would like to thank my supervisors Andy Crabtree, Joel Fischer and Boriana Koleva from the Mixed Reality Laboratory for their extraordinary patience, guidance and support during my time as a PhD student. Had it not been for them, none of this would have been possible.

I'm also very grateful to my examiners Paul Coulton and Paul Tennent whose thoughtful feedback opened my mind to new ways of thinking about my work and ultimately helped make this thesis better.

My time at the MRL has been blessed with many supportive lab colleagues. In particular I would like to thank my office mates and friends James Colley, Huseyin Avsar, Martin Porcheron, Richard Ramchurn and Horia Maior. Thank you guys for not once - to my knowledge - having tried to kill me! Well, except for that one time when James deliberately short-circuited an Arduino board and then told me to touch it because he wanted to "test something".

I'm likewise grateful to Michel Valstar for shredding my initial prototyping work to pieces during my first ever lab talk, proclaiming that he "can't imagine anybody programming an autonomous system in a house that is this stupid". No matter how much it stings, in retrospect I've come to realize that it is such constructive criticism that helps us grow and improve.

I would also like to extend my thanks to John Carroll, whose work was a major inspiration for this thesis, and who kindly provided me with his feedback during a lunch event at CSCW 2018.

Last but certainly not least, I want to thank my parents, Jana and Glenn, and my two brothers, Ronny and Jakub, for their unwavering support and faith in me throughout this strange, muddled, lonely, tiresome, but ultimately exhilarating, inspiring, rewarding and worthwhile journey.

Table of Contents

Chapter 1 Introduction 13

1.1 The Problem 14
1.2 Aims 15
1.3 Objectives 16
1.4 Chapters overview 17
1.5 Key findings 20
1.6 Original contribution 20

Chapter 2 Literature Review 22

2.1 The Sociotechnical Context 24
2.2 Defining an Autonomous System 28
2.3 Domestic Autonomous Systems and Key Challenges 34
2.3.1 Coordination 37
2.3.2 Control 40
2.3.3 Accountability 44
2.4 Responding to the Challenges 48
2.4.1 Coordination 48
2.4.2 Control 51
2.4.3 Accountability 55
2.5 Synthesis 62

Chapter 3 Methodology 65

3.1 Scenario-based design (SBD) 72
3.2 Foresight and Plausible Futures 77
3.3 Contravision 80
3.4 Synthesis 84

Chapter 4 Storyboard-based Pilot Study 86

4.1 Implementation 86
4.2 Engagement with Users 93
4.2.1 Participants 93
4.2.2 Focus group sessions 94
4.3 Coordination 95
4.3.1 Reasoning about voice messages 96
4.3.2 Reasoning about timeliness 98
4.3.3 Reasoning about situated action 99
4.4 Control 100
4.4.1 Reasoning about customisation 100
4.4.2 Reasoning about direct control 102
4.5 Computational Accountability 103
4.5.1 Reasoning about accountability 104

4.5.2 Reasoning about motives 105
4.5.3 Reasoning about transparency 106
4.5.4 Reasoning about the nature of transparency 107
4.6 Social Accountability 108
4.6.1 Reasoning about acceptability 109
4.6.2 Reasoning about agency and entitlement 110
4.6.3 Reasoning about trustworthiness 111
4.6.4 Reasoning about risks 112
4.7 Key Takeaways 113

Chapter 5 Further Exploration through Video Scenarios 116

5.1 Implementation 116
5.2 Engagement with users 121
5.2.1 Participants 122
5.2.2 Focus group sessions 123
5.3 Coordination 125
5.3.1 Reasoning about convenience 125
5.3.2 Reasoning about food automation 127
5.3.3 Reasoning about recommendations 129
5.4 Control 132
5.4.1 Reasoning about loss of control 133
5.4.2 Reasoning about granular control 135
5.5 Computational accountability 137
5.5.1 Reasoning about ease of use 138
5.5.2 Reasoning about personal data legibility 141
5.5.3 Reasoning about terms of use 143
5.6 Social accountability 145
5.6.1 Reasoning about privacy 145
5.6.2 Reasoning about community sharing 150
5.6.3 Reasoning about embodied autonomous systems 152
5.6.4 Reasoning about inconvenience 154
5.7 Key Takeaways 159

Chapter 6 The Drill-down Study 163

6.1 Implementation 164
6.2 User Engagement 169
6.2.1 Participants 169
6.2.2 Interview sessions 171
6.3 Coordination 173
6.3.1 Reasoning about machine initiative 174
6.3.2 Reasoning about human autonomy 179
6.3.3 Reasoning about the human in the loop 182
6.4 Control 185
6.4.1 Reasoning about customization 185
6.4.2 Reasoning about human autonomy 187

6.5 Computational Accountability 190
6.5.1 Reasoning about accountability 191
6.5.2 Reasoning about algorithmic decision making 192
6.5.3 Reasoning about motives 195
6.5.4 Reasoning about personal data flows 198
6.6 Social Accountability 200
6.6.1 Reasoning about acceptability 200
6.6.2 Reasoning about risks 203
6.6.3 Reasoning about trustworthiness 207
6.6.4 Reasoning about agency and entitlement 210
6.7 Key Takeaways 211

Chapter 7 Discussion 217

7.1 Methodology: Lessons Learned 219
7.1.1 Breaching Experiments 220
7.1.2 Scenarios 223
7.1.3 Contravision 227
7.2 The Problem of Social Acceptability 230
7.2.1 Coordination and control 231
7.2.2 Computational accountability 234
7.2.3 Social accountability 236
7.3 Design Challenges 239

Chapter 8 Conclusions 247

References 249

List of Tables

Table 1 - System autonomy taxonomy, as proposed by Parasuraman et al. (2000) 33
Table 2 - Overview of the contra-vision scenarios 92
Table 3 - A summary of the scenarios produced for our second user study 119
Table 4 – Sample set of screenshots illustrating the aesthetics of our scenarios along with the respective contravision points 121
Table 5 – Study 2 participant overview 123
Table 6 - Study 3 participant overview 170

List of Figures

Figure 1 – Donald and his smart home.....92
Figure 2 - Collage of graphic storyboards depicting the early version of our scenarios.....167
Figure 3 - Collage of screenshots from the video scenarios developed for our third user study.....168

Publications Disclosure

The material presented in Chapter 4 (Storyboard-based Pilot Study) and Chapter 7 (Discussion) of this thesis is an extension of work that has previously been published as the following:

> *Nilsson, Tommy, Andy Crabtree, Joel Fischer, and Boriana Koleva. "Breaching the Future: Understanding Human Challenges of Autonomous Systems for the Home." Personal and Ubiquitous Computing 23, no. 2 (2019): 287-307. http://dx.doi.org/10.1007/s00779-019-01210-7*

The material presented in Chapter 3 (Methodology) of this thesis is an extension of work that has previously been published as the following. This publication was a recipient of the 2019 CHI Video Showcase award (Best Design Research Method).

> *Nilsson, Tommy, Andy Crabtree, Joel Fischer, and Boriana Koleva. "Creating Worlds: Exploring Animated Videos as a Tool for Contextualization in User Research." In Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems, p. VS11. ACM, 2019. http://dx.doi.org/10.1145/3290607.3311776*

The material presented in Chapter 5 (Further Exploration through Video Scenarios) of this thesis is an extension of work that has previously been published as the following:

> *Nilsson, Tommy, Joel E. Fischer, Andy Crabtree, Murray Goulden, Jocelyn Spence, and Enrico Costanza. "Visions, Values, and Videos: Revisiting Envisionings in Service of Ubicomp Design for the Home." In Proceedings of the Designing Interactive Systems Conference 2020 (DIS'20), ACM, 2020. http://dx.doi.org/10.1145/3357236.3395476*

"But I don't want comfort. I want God, I want poetry, I want real danger, I want freedom, I want goodness. I want sin."

— Aldous Huxley, Brave New World

Chapter 1 Introduction

The future ain't what it used to be. Technological progress was meant to make us wise and powerful. We were supposed to journey to the center of the Earth, to embark on a space odyssey, even engineer a modern utopia. Instead, it appears to have made us progressively more alienated from our own world, dependent and deprived of control over our lives. Indeed, there is much that seems backwards; cars are driving themselves, phones are talking to each other, TV's are watching *their* people. The material world around us which we have been collectively shaping and moulding to our needs since the dawn of civilization is suddenly awakening to a life of its own.

It has been suggested that a widespread adoption of such *autonomous* technologies would make our existence more economically efficient, convenient and healthy, perhaps even usher in a new industrial revolution (Schwab, 2016) resulting in a state of unprecedented prosperity. Given the current speed of technological proliferation, such a grand vision does not seem out of reach. In 2015 it was, for example, estimated that the global autonomous systems market reached a 15% annual growth rate (Bisset, 2015) with the potential to create £ trillions worth of economic impact (Lewis, 2014). All this progress, however, comes at a price. With too many opportunities opening up too quickly, there is a risk our technological capability will outpace our capacity for making wise and far-sighted design choices. Instead of trying to figure out how we *could* develop ever more capable solutions, we ought to pause and contemplate whether and why we *should*. Can we really harness the potential offered by autonomous technology without breaching a sense of social acceptability? What would the human costs be if we fail?

Rather than any material benefits brought about by future and emerging autonomous systems, their domestication creates a situation where it is a system's capacity to manifest socially-aware behaviour by catering to our human needs and values, inducing trust and preserving our self-actualization that ultimately determines its acceptability.

1.1 The Problem

Market analysts are suggesting that the domestic context has a realistic potential of becoming the main growth area for future and emerging autonomous systems, with experts estimating that the global value of the smart home sector is set to exceed £100 billion by 2022 (Markets&Markets, 2019). Systems such as the NEST smart thermostat (Yang and Newman, 2012) which predicts occupancy to regulate temperature or services built around the Amazon Dash Replenishment API (Lehmacher, 2017), are already finding their way into the everyday spaces of our homes with many more expected to soon follow.

It is, however, important to treat such predictions with a degree of caution as they are based on mere extrapolations from the current trends in technical progress. It is not beyond the realm of possibility to speculate that new non-technical pitfalls might yet show up and negatively impact - perhaps even preclude - the mainstream adoption. A novel technology has the habit of stirring up novel human problems, and domestic autonomous systems are no exception. Yang and Newman (2013) found, for instance, that users of the NEST thermostat have frequent difficulties in understanding its behaviour. Similarly, while probing prospective users of autonomous domestic systems, Rodden et al. (2013) came across substantial concerns pertaining to the loss of control, and the level of personal data harvesting enabled through widespread sensing.

Studies such as these suggest that social expectations regarding the intelligibility and trustworthiness of autonomous systems are as important to their uptake and use as

any perceived technological benefits. A novel solution deployed in a domestic setting will, in other words, not be successfully adopted simply because it works on a technical level, but equally because it fits-in with the complex expectations that govern the intricate social setting of the home.

Yet, in spite of this, people's expectations towards domestic autonomous systems remain a relatively poorly understood domain (Wilson et al., 2015).

1.2 Aims

The aim of this thesis is to address these limitations by exploring how potential users relate to the notion of having their homes augmented through autonomous systems and in so doing elaborate key points of friction. In order to reach this goal, a number of steps will have to be taken. To begin with, although autonomous systems are already in the process of gaining traction, as argued above, they are yet to reach a level of ubiquity where they would be intimately interwoven with the everyday domestic life. Consequently, envisioning such emerging settings, and making them available for prospective users to scrutinize, will inevitably require a degree of speculation. With this in mind, the first research question of this thesis has been defined as follows:

- R1: How can we assess the manner in which prospective users will respond and relate to future autonomous systems for the home?

As mentioned previously, a range of problem areas surrounding social acceptability of autonomous systems, such as the lack of intelligibility or trustworthiness, have already been surfaced by previous studies. Such research enquiries have however been generally focussing on existing solutions, rather than a future all-encompassing autonomous sociotechnical infrastructure. It seems reasonable to assume that as the ubiquity and sophistication of autonomous systems increases, known problem areas

will evolve and entirely new ones will emerge. In line with this reasoning, the second research question of this thesis has been set as follows:

- R2: What are the key areas of friction impacting the acceptability of future domestic autonomous systems? What implications do these frictions have?

For the research work presented in this thesis to have any practical value, it will be crucial to go beyond mere passive observation and to propose specific and actionable takeaways. Offering insights that could potentially be put to use in the design of socially acceptable autonomous systems is, in other words, the purpose of our final research question:

- R3: What are the key design challenges that ought to be tackled in order to improve the level of acceptability of domestic autonomous systems?

1.3 Objectives

Following Rohracher's (2003) call to engage prospective users from the very envisioning stage of development, we[1] have developed a novel workshop-based methodology that exploits utopian and dystopian scenarios as tools enabling sociologist Harold Garfinkel's (1963) "incongruity procedure" or "breaching experiments" (Garfinkel, 1964) as this procedure is more commonly known. The nature of breaching experiments will be elaborated in greater detail later on in this thesis (see section 3.1). For now we might summarize the core underlying idea as a research approach deriving insights from a deliberate violation of social norms. For example, Garfinkel (Ibid.) would have his students bargain for standard-priced merchandise in stores. The shopkeeper's reaction to such odd behaviour would in

[1] Although this thesis has been written by me as the sole author and reflects my independent research efforts, the presented ideas were conceived and iteratively refined collectively with input from my supervisors. To acknowledge their contribution, I have chosen to adopt the pronoun "we" whenever referring to the originator of the described research activity.

turn allow the researcher to study the social norms and expectations that apply in the given situation.

Drawing on such work, the purpose of our procedure is to provoke (as in call forth and make visible) the taken for granted background expectancies that organise familiar settings of everyday life, which in turn impact the adoption and use of new technologies within those settings.

- Accordingly, we will create a set of provocative contra-vision (Mancini et al., 2010) scenarios to probe and elicit social expectations that impact the uptake of autonomous systems in the home and subsequently explored those in workshops across three separate studies with a total number of 80 potential end-users.
- The data gathered through such user studies will fuel a subsequent qualitative analysis, which will help surface and identify key thematic concerns pertaining to social acceptability of domestic autonomous systems.
- Finally, upon reflecting on the outcomes of our work in the light of relevant literature, we shall translate these findings into a set of key design challenges and articulate practically applicable takeaways.

1.4 Chapters overview

The remainder of this thesis is organized as follows:

Chapter 2 provides a summary of existing research concerned with autonomous systems for the home. We start out by providing a brief outline of some of the key technical concepts and terms discussed in this thesis. This is then followed up with a summary of the technological advances that have enabled a range of computational solutions capable of proactive decision-making and actuation on behalf of its users. Many political reports and economic forecasts have it that the potential benefits offered by such autonomous technologies will lead to their widespread proliferation

and in turn have a dramatic impact on our daily lives. Such predictions are however not adequately factoring in the potential non-technical problems that might render prospective autonomous solutions socially unacceptable. Based on a review of relevant literature, we argue that these human challenges can be loosely categorized into three problem areas: "Coordination" speaks to the need for autonomous systems to gear-in with the intricate social routines present in the home. "Control" deals with the need to preserve human agency. And finally, "accountability" is concerned with rendering autonomous system behaviour intelligible to its users. Although efforts to mitigate these challenges are currently underway, we argue that much of the relevant autonomous technology was conceived with the workplace in mind, and that these problem areas, along with the relevant human expectancies, thus remain poorly understood in a domestic context.

Chapter 3 elaborates the methodological approach we have employed in this study, with a particular emphasis placed on describing and motivating its three key components in the form of "breaching experiments", "scenario-based design" and "contravision". Breaching experiments represent a method designed to disturb common sense understandings of everyday life in order to make visible the background expectancies that order everyday life. Scenario-based design is an approach to low-fidelity prototyping enabling the envisioning of interactions between prospective technologies and their users. We reflect on the potential use of scenarios to convey plausible futures and thus produce a degree of foresight. Finally, contravision represents a framework centred around the idea of conveying multiple contrasting scenarios with the goal of reducing bias and eliciting richer responses during user studies. Drawing on these three methodological components, we developed a series of provocative utopian and dystopian scenarios which we subsequently used to drive our focus group approach to qualitative user research.

Chapter 4 describes the initial user study and its findings. By carrying out two separate focus group sessions with a combined number of 32 prospective users, we were able to surface a range of relevant background expectancies that largely confirmed the three problem areas proposed in chapter 2. The nature of concerns

prompted us to introduce a fourth problem area in the form of "social accountability" which we argued extends beyond the intelligibility of system actions towards their "fit" with social expectancies and visible appropriateness.

Chapter 5 describes our second user study and its findings. Whereas the first study helped us converge on 4 key problem areas, the goal of our second study was to elaborate these areas further and reveal more of the background expectancies that speak to the role that autonomous systems ought to play in our lives. For this purpose we developed a set of animated video scenarios contrasting domestic autonomous solutions prioritizing the user's convenience with solutions that instead encourage a more healthy and sustainable lifestyle. Unsurprisingly, while evaluating these scenarios with 28 prospective users spread across 6 focus groups, the majority of participants voiced their preference towards systems encouraging healthy and sustainable lifestyles. By invoking values such as trust and human autonomy, the underlying reasons behind these preferences allowed us to gain a better understanding of the previously revealed problem areas.

Chapter 6 describes the third and final user study along with its findings. Whereas our initial two user studies focused on surfacing and elaborating the problem areas concerning social acceptability of autonomous systems for the home, the goal of our final study was to offer a hint as to how these problem areas might potentially be resolved and in-so-doing help us formulate specific design challenges. For this purpose we abandoned the traditional "binary" approach to contravision and instead produced three distinct scenarios, each postulating a different relationship between particular autonomous systems and their prospective users. In order to examine the given issues in greater depth, we also replaced our original focus group approach with interviews featuring 20 prospective users spread across 16 separate sessions.

Chapter 7 summarizes our key findings and places them within the context of existing literature. We review our methodological approach and reflect on the key lessons learned through our work. We follow this up by elaborating each of the 4 key problem areas and explain the manner in which our findings expand on

previous research. We then formulate a set of 7 key design challenges faced by future and emerging autonomous systems and illustrate how these could be translated into explicit design guidelines. Finally, we reflect on the novelty and potential limitations of our methodological approach.

A brief summary of our findings, along with the study conclusions will be offered in chapter 8.

1.5 Key findings

Our analysis of workshop participants' responses to the scenarios reveals background expectancies centring on coordination and the need to build the human into the behaviour of autonomous systems, control and the need to protect human agency, computational accountability and the legibility of autonomous system behaviour, as well as social accountability and the compliance of autonomous behaviours with social norms. Each of these key topics is formed by and brings with it a range of ancillary expectations that impact the adoption of autonomous systems in the home and open up design possibilities to enable developers to gear autonomous domestic systems in with the non-technical expectations that govern their uptake in everyday life. By extrapolating from these key topics, we argue that building accountability, context awareness, social competence, transparency, coordination mechanisms as well as control mechanisms into autonomous systems constitutes the key design challenges in this topic.

1.6 Original contribution

This thesis makes two contributions. The first one is of a methodological nature and focuses on the use of provocative utopian and dystopian scenarios of future autonomous systems in the home. These are used to drive an innovative workshop-

based approach to breaching experiments, which surfaces the usually tacit and unspoken background expectancies implicated in the organisation of everyday life that have a powerful impact on the acceptability of future and emerging technologies. The other contribution is substantive, produced through participants' efforts to repair the incongruity or "reality disjuncture" created by utopian and dystopian visions, and highlights the need to enable coordination and control, as well as building computational and social accountability into autonomous systems.

More specifically, we demonstrate the role of user input in keeping autonomous systems responsive and acting in a timely fashion to foster human-machine coordination. We elaborate the need for granular control and customization of individual smart features in an autonomous system in order to enable effective human-machine orchestration and synchronisation. We show that explanations of machine behaviour should not be concerned with the inner workings of autonomous systems but rather articulate the overarching motives and reasons. Finally, we show the importance of computational agency exhibiting social competence and autonomous behaviour complying with the differential rights and privileges at work in any particular home.

We follow this up by formulating 7 broad design challenges that we argue will play a crucial role in shaping the acceptability of future autonomous systems for the home and elaborate the manner in which these might help guide the design of real solutions.

Chapter 2 Literature Review

This chapter discusses the state-of-the-art and major challenges surrounding domestic autonomous systems. We shall start out with a brief overview of the key technological advances that have led up to the present situation. This will be followed up by a definition of the term "autonomous system", explaining its use in this thesis. Next, we shall look into key areas of friction emerging as a result of autonomous systems being increasingly frequently adopted in domestic settings. Finally, we shall examine the efforts to mitigate these frictions that have taken place so far and summarize the main challenges that have yet to be addressed.

Before we proceed, however, it would be productive to first briefly define a set of key terms that we shall rely on persistently throughout the remainder of this thesis. Although each of these terms will be elaborated in greater detail later on in this chapter, given that they are interrelated and partially overlapping, we reason that presenting such an overview here will help prevent any misunderstanding and improve the overall readability of our work.

- **Ubiquitous computing**, also referred to as "UbiComp" or "pervasive computing", stands for the notion of embedding computational resources into the fabrics of everyday life in order to make their potential benefits available anytime and anywhere (Sundar et al., 2013). While traditional personal computers took on the form of bulky and relatively stationary machines, the gradual miniaturization of the computer microchip made it possible to augment a growing spectrum of everyday objects with the capability of computing, thus contributing to its increasing ubiquity (Kuniavsky, 2010).

Everyday objects augmented in this manner are often described using the adjective "smart" (Ibid.). Examples of such UbiComp solutions include smart watches, self-checkouts in supermarkets and even smart showers (Kwon et al., 2018).

- **Internet of Things (IoT)** refers to the notion of linking up everyday UbiComp objects through wired or wireless networks (Perera, 2017). The idea is that such interconnection will enable exchange of data and in turn contribute towards better coordination and optimization of products and services (Ibid.). Examples of IoT include connected security systems or domestic thermostats.
- **Artificial intelligence (AI)** refers to the ability of a system to imitate human intelligence by performing learning and problem-solving activities (Luger, 2005). Typically, such level of intelligence is infused into systems using algorithms designed to enable identification of patterns in collected datasets and a subsequent inference of potentially important information (El Naqa and Murphy, 2015). A common example of AI is an online store monitoring the activity of its visitors in order to subsequently provide them with tailored shopping recommendations (Ahmeda et al., 2015).
- **Cyber-physical systems (CPS),** as used in this thesis, refers to any computer system capable of sensing environmental variables (e.g. temperature, pressure or sound) and responding to these through physical interactions with the environment (Rajkumar et al., 2010). An example of a CPS is an energy saving system capable of monitoring the occupancy of an office space and automatically turning off the lights and heating whenever the space is not in use (Kleissl and Agarwal, 2010). A CPS may or may not be endowed with AI, depending on its design.
- **Autonomous systems** refer to any computer system - virtual or cyber-physical - endowed with sufficient AI to act proactively and to dynamically adjust its course of actions in response to a changing situation, without requiring direct human input or oversight (Albus et al., 1998). Self-driving cars (Brown and Laurier) and auto-reordering (Fuentes et al., 2019) services are two prominent examples of emerging autonomous systems.

- **The home** setting has been described as a leading growth area for future autonomous systems (Crabtree et al., 2003) and one of the key hubs where the above described technologies are expected to coalesce (Ma et al., 2014). We acknowledge that the term "home" has a broad meaning spanning a range of concepts from student flats, shared housings to family mansions. When we use terms such as "home" or "domestic environment" in this thesis, we are thus partially brushing over this complexity and simply referring to any type of private residency which one normally only shares with family members and close friends. The distinction between 'public' and 'private' is of importance here, as it seems reasonable to assume that our expectations concerning privacy, security and other human values are elevated in the private sphere. It is this broadly conceived private living space then, that constitutes a key focal point of this thesis.

2.1 The Sociotechnical Context

The gradual reduction in size and cost of the microprocessor has been described as one of the most significant technological trends of the past century (Kuniavsky, 2010), effectively extending the application area of electronic computing beyond desktop devices into the realm of smartphones and eventually a growing spectrum of mundane everyday objects ranging from fridges to shoes (Chi et al., 2007).

An important factor making it practical to embed computational resources into the fabric of our everyday living spaces stemmed from another, largely congruent, technological trend in the form of rising sophistication and falling cost of sensory technology. When paired with embedded microprocessor devices, sensors made it possible to interpret and use various forms of data (such as video, audio, temperature, or humidity) extracted directly from the physical surroundings, thus enabling the sensing of environmental factors so that they could in turn influence computational processes and vice versa. When coupled with computing machinery,

such interconnected conundrums of context-aware sensors are sometimes referred to as *cyber-physical systems*, or simply CPS (Malekian, 2017).

It should be briefly noted here that our use of the word *system* is not markedly distinct from its traditional meaning as defined by the Oxford Dictionary of English (1989): a system is "a set of things working together as parts of a mechanism or an interconnecting network; a complex whole". Upon examining scientific literature concerned more specifically with computing systems, we find that the provided definitions do generally not deviate from this dictionary description. Sillitto et al. (2017), for instance, describe a computer system as "a complex whole, whose properties are due to its constituent parts, as well as to relationships among the parts". In line with these definitions, we can thus summarize our use of the word system as any composite computing solution made up of individual constituencies that are mutually interdependent in the pursuit of a task.

In the case of a CPS, this may entail a wide spectrum of such constituencies. Besides utilizing sensors to passively infer data about the physical world, a common practice is for example to also incorporate various forms of actuators. An actuator, such as an electric motor, can then empower a CPS to not merely sense and process data, but also to carry out direct physical actions in response to this sensed data, effectively allowing it to perform manual work tasks without the need for a human operator, or in other words *autonomously* (Bridgera, 2017).

Systems capable of such autonomous actuation have been suggested to hold the potential to revolutionize many aspects of our lives and are consequently in the process of attracting substantial interest from both the industry and government bodies (Steering Committee for Foundations in Innovation for Cyber-Physical Systems, 2013). Indeed, various forms of autonomous cars, robotic surgery systems, intelligent buildings, smart electric grids, smart manufacturing solutions, and implanted medical devices are just a few of the practical examples that are already gaining traction (Monostori, 2018).

The growing sophistication of the tasks undertaken by such solutions is in turn fuelling the rise in demand for various forms of artificial intelligence (AI) that would grant systems the ability to autonomously assess complex situations and independently choose the optimal action to be taken, rather than just act reflexively in response to each predefined stimuli in a predetermined manner (Obermeyer and Emanuel, 2016).

In an attempt to summarize the key idea behind AI, the UK-based Engineering and Physical Science Research Council (2019) converged on a somewhat broad definition, describing it as a set of technologies that "aim to reproduce or surpass abilities that would require 'intelligence' if humans were to perform them". The definition goes on to add that these abilities specifically include "learning and adaptation; sensory understanding and interaction; reasoning and planning; optimisation of procedures and parameters; autonomy; creativity; and extracting knowledge and predictions from large, diverse digital data" (Ibid.).

While there are numerous approaches to endowing computer systems with such artificial intelligence, one of the most successful techniques is based around the principle of "automated detection of meaningful patterns in data" (Shalev-Shwartz and Ben-David, 2014). This approach, also known as *machine learning*, utilizes algorithms capable of inferring potentially useful information patterns from datasets. The inferred information patterns are subsequently used as a basis for extrapolations, effectively enabling a system to calculate a set of predictions and act on them (El Naqa and Murphy, 2015). For instance, social networking services are frequently utilizing machine learning algorithms to infer media consumption patterns of their users and in turn proactively provide each user with recommendations and other content tailored to fit their own individual interests (Daffodil, 2017).

The arrival of such intelligent proactive systems is already in the process of enabling completely new services and business models, ranging from simple self-regulating thermostats (Min et al., 2017) to complex solutions predicting travel behaviour of urban commuters in order to help taxi drivers optimize their driving routes (Sun et

al., 2014). Chief Technology Officer of the High Value Manufacturing Catapult, Sam Turner (2017), elaborates the potential offered by such solutions, stating that "using digital technologies such as artificial intelligence, sensor technologies and automation makes companies more agile and better equipped to respond to, or even act ahead of, changing consumer demands".

The impact that autonomous solutions endowed with a degree of intelligence might prospectively have on daily life were discussed by the *2017 House of Lords report on the state of AI in the UK*, which explains that "just as electricity, or steam power before it, started out with particular, often somewhat niche, uses prior to gradually becoming fundamental to almost all aspects of economic and social activity, AI may well grow to become a pervasive technology which underpins our daily existence" (House Of Lords Select Committee, 2018). A report by the World Economic Forum (WEF) supports such views, claiming that there is a £100 trillion opportunity for both industry and society through the adoption of intelligent technologies (Schwab, 2016). In the UK alone this process is estimated to result in a net gain of 175,000 jobs throughout the economy over the following decade (Maier, 2017).

In the light of such potential, it is hardly surprising to see new autonomous products and services being conceived and rolled out into our everyday lives at a staggering pace. In 2017, it was estimated that around 6.4 billion so called "smart objects" were in use globally, with this number expected to reach 20 billion by 2020 (Schaffer, 2017). The scale and speed of this technological progress and adoption has prompted some scholars to label it the "fourth industrial revolution" (Schwab, 2016). The aforementioned WEF report goes a step further, arguing that when compared to previous industrial revolutions, "this one is evolving at an exponential rather than a linear pace while disrupting almost every industry in every country." The report concludes that "the breadth and depth of these changes herald the transformation of entire systems of production, management, and governance" (Ibid.).

It does indeed appear as indisputable that autonomous solutions are in the process of changing the way we work, live and relate to the world around us. The questions raised by their widespread uptake have been noted to already increasingly often

challenge the existing ideological positions that have defined politics in the UK and much of the developed world" (House Of Lords Select Committee, 2018). The transformative nature of the processes set into motion by this wave of technological innovations is likely to further increase as we move into the future, making it critical to study the nature of autonomous systems, their capabilities and the challenges associated with their adoption and use.

2.2 Defining an Autonomous System

What then exactly is an autonomous system? Answering this question is problematic insofar as there is no single universally accepted definition to draw upon. Williams and Scharre (2015) note that the far reaching impacts stemming from the proliferation of relevant computing technology has resulted in a situation where pertinent issues are not just being discussed from a technical perspective, but - amongst others - also from ethical, legal or business-centric standpoints. This, they argue, has in turn contributed to a state where the term "autonomous system" constitutes a difficult concept to understand, with its use in existing discourses being frequently inconsistent or even inappropriate. Williams and Scharre (Ibid.) add that in particular there is a tendency amongst scholars to use the words "automatic" and "autonomous" interchangeably. Vagia et al. (2016) follow a similar line of reasoning, arguing that understanding the proper meaning of these two terms is a necessary precondition for analysing the literature on this subject.

While interrelated, the adjectives *automatic* and *autonomous* are each referring to a different kind of machine behaviour and as such ought not to be conflated. The Oxford Dictionary of English (1989) lists five definitions of *automatic*: 1. operating by itself without human control; 2. (of a gun): able to fire continuously until the bullets run out; 3. done without conscious thought; 4. occurring spontaneously: automatic enthusiasm; and 5. (of a punishment): applied without question because of a fixed rule.

On the other hand, the aforementioned dictionary defines the meaning of the word *autonomous* as: 1. (of a country or region) having the freedom to govern itself or control its own affairs. 2. Having the freedom to act independently. 3. (in Kantian moral philosophy) acting in accordance with one's moral duty rather than one's desires.

While examination of the definitions above might easily lead one to the conclusion that automatic and autonomous are largely overlapping in their meaning, a closer look reveals that the latter possesses a particular quality of freedom that we normally associate with human beings. Being autonomous is not simply about performing a task "by itself" or "without question". Rather, it is about the notion of also being able to perform it independently, based on one's own free decision.

This distinction becomes more pronounced once we look into relevant academic literature concerned with computing systems. Fraser Dunlop (2009), for instance, argues that automatic systems have a "fixed choice points", meaning that they carry out a specific set of actions in response to predetermined inputs. An action carried out automatically, in other words, does exactly what it was programmed to do by executing predefined instructions without having any choice or possibility to act in any different way or to readjust itself in response to a changing situation at hand.

As explained by Albus et al. (1998), systems operating autonomously go a step further by "making their own decisions about their actions while performing different tasks, without the need of an exogenous system or operator". Autonomous system behaviour is thus typically based on techniques such as machine learning, allowing it to act in a manner that is dynamically adapted to a changing environment. Instead of blindly following a pre-programmed course of action, autonomous systems are thus endowed with the power to make choices free from outside influence in the sense that the data that's being inferred and acted upon may be distinct from what was anticipated when the system was originally developed. This distinction appears to be prevalent in existing literature. In his study of military drones, Bruce T. Clough (2002), for instance, declares at one point that "a cruise

missile is not autonomous, but automatic since all choices have been made prior to launch."

Another noteworthy attempt to elaborate the concept of intelligent actuating systems can be found in the work of Wooldridge and Jennings (1995), which helped popularize the term autonomous "agent" for denoting any computer system capable of acting independently on the user's behalf, irrespective of whether this actuation takes place in the real or virtual space. In an attempt to define what exactly the agency of such a system consists off, they offer a "weaker" and "stronger" definition.

The weaker definition settles for a relatively broad description, arguing that besides the ability to operate without direct human intervention, agency in computer systems is also characterized by a degree of social capacity, enabling agents to interact with other agents or humans. Moreover, agency is also reactive, allowing agents to perceive their environment and respond in a timely fashion to changes that occur in it. Finally, agency entails proactivity, in the sense of granting agents the ability of goal-directed behaviour and initiative.

The stronger definition of agency, on the other hand, narrows down this description by arguing that another criterion for agency comes in the form of a range of human-like attributes, including "knowledge", "belief", "intention" or "obligation". Such anthropomorphism of computer systems, they argue, may even be enhanced by employing visual representations, such as a cartoon-like graphical icon.

While the weaker definition of agency constitutes a fairly general description, thus making it more acceptable across the many fields concerned with autonomous systems, Wooldridge and Jennings (Ibid.) suggest that the stronger definition of agency is more precise and better captures the nature of agency within the context of the AI and machine learning discourse.

This then leaves us with a somewhat multi-layered description of actuating systems. While we have established that automatic behaviour generally refers to systems executing pre-programmed commands in a predictable manner, and that autonomous behaviour entails a level of independent volition, Wooldridge and

Jennings (Ibid) appear to introduce yet another - higher - level of system autonomy, one featuring more or less explicit social and human capabilities. Articulating a single coherent definition is consequently no easy task. The diverse nature of autonomous computer systems has prompted Scott Totman (2019) to argue that we ought to think of them as a continuum, rather than a fixed definable concept. Systems that were originally endowed only with the most elementary form of intelligence, operating on the basis of pre-programmed sets of inputs and outputs, may need to become smarter over time as their usage and the environment in which they operate change. As he put it himself, one could "take an automatic system and build in some autonomous capabilities, extending the useful life of the system and its overall applicability".

A major strand of academic literature on autonomous systems follows the same reasoning, elaborating the concept of system autonomy as a continuum in the form of a layered taxonomy. Each layer is then typically used to describe a different degree of system autonomy, starting from simple automation of repetitive tasks and extending towards independent decision-making and other forms of human-like behaviour.

One of the earliest of such attempts can be found in the work of Sheridan and Verplank (1978), who identified six basic categories of actions that either a human operator or a computer system can take on. These include "gets", "selects", "starts", "requests", "approves" and "tells". The degree of system autonomy, they argue, is dependent on the manner in which control over these categories is divided between the computer and the human user. Upon examining the various permutations this distribution of power could assume, Sheridan and Verplank identified 10 different levels of machine autonomy ranging from complete human control to a state where the human is completely left out of the loop.

A number of scholars have since then proposed various refinements and simplifications of this taxonomy. Mica Endsley (1987), for instance, proposed a somewhat trimmed down version, based around 4 different levels of autonomy at which a system can operate. The first level describes systems that issue

recommendations to their users, but otherwise refrain from taking any direct actions. The second level describes systems that carry out direct actions autonomously but have to ask the user for permission each time. The third level describes systems that carry out actions autonomously without having to ask for permission, but the user maintains the power to interfere with, or terminate, any particular system action. Finally, the fourth level describes systems acting fully autonomously, with the human user being deprived of any control.

The common denominator shared by the taxonomies outlined above is the notion that there is no single universally accepted optimal level of autonomy to strive for. Rather, different levels are needed for different purposes. It is likewise reasonable to assume that the ideal level often depends on a range of human variables, including the social context and the user's personal preferences. To complicate matters further - as will be discussed later in this chapter - various solutions implemented into a domestic ecosystem are frequently operating simultaneously while overlapping, complementing and having direct impact on each other. Sensing devices collecting data about their users, for instance, serve as a crucial precondition for any higher-level autonomous service by providing the necessary input. It is this conundrum of automatic and autonomous behaviours that allows computer systems to ingest and learn from growing datasets faster, and eventually more reliably, than what any human would be reasonably capable of doing.

Perhaps the most up-to-date - and best aligned to the topic of this thesis - description of autonomous systems can therefore be found in the work of Parasuraman et al. (2000), who offer what can be described as a re-evaluation of the previously described taxonomies in the light of the emerging AI-enhanced socio-technological landscape. As illustrated in Table 1, they identified 4 different categories of system behaviour where automation can be applied. Instead of presenting these categories as 'levels' along a continuum ranging from complete human control to full automation, Parasuraman argued that each category can in itself be automated to a varying degree. Whereas category 1 describes monitoring and data collection, category 2 is concerned with the processing and analysing of thus collected data.

Category 3 encompasses processes leading to the production of specific actionable decisions. Finally category 4 deals with the implementation of those decisions. Parasuraman et al. (Ibid.) go on to suggest a recommended level of autonomy for each of the outlined categories. Whereas system behaviour falling under category 1 and 2 can reliably be automated to a high degree, category 3 and 4, they argue, require a lesser level of automation in order to avoid interfering with human activities.

Behaviour level	Purpose	Description
Level 1	Information acquisition	Systems responsible for sensing, monitoring, and registering data.
Level 2	Information analysis	Systems responsible for the processing, predictions, and general analysis of collected data.
Level 3	Decision selection	Systems responsible for selecting between different decisions alternatives.
Level 4	Action implementation	Systems responsible for implementing their decisions or commanding new actions.

Table 1 - System autonomy taxonomy, as proposed by Parasuraman et al. (2000)

In line with the work above, a key premise of this thesis is thus the assumption that understanding the expectancies held by users towards domestic autonomous systems is only feasible by taking a big picture approach and studying system operation as a whole, rather than singling out any of its individual constituencies. Consequently, at this stage of our work, we feel compelled to adopt the somewhat broad description of system autonomy proposed by Parasuraman et al. (Ibid.), which allows us to examine the full spectrum of actuating systems. Although one could easily argue that for example the act of information acquisition (i.e. sensor networks passively collecting data) fits better under the definitions of *automatic* rather than *autonomous*, as we shall see in the following section, in an intricate social environment, such as the home, even such semi-intelligent mechanisms can have a powerful impact on the uptake of autonomous systems and as such will not be excluded from our enquiry.

2.3 Domestic Autonomous Systems and Key Challenges

In this section we elaborate the ongoing trend of adopting autonomous systems in domestic settings along with the key challenges emerging as a consequence of this process.

Against the backdrop of governments across the world gradually codifying internet access as a basic human right (Kay, 2012), we now see connectivity - hard-wired or wireless - turning into a standard feature of computing devices (Tanti and Buhalis, 2016). The introduction of 5G mobile network technologies, which has been touted by media outlets as bringing internet connection speeds so fast that they will "support an entirely new way of life" (CNN, 2019), is only further accelerating this process.

Perhaps most obviously, this development has helped enable an increasing number of internet-connected computer systems. A recent report by Gartner (2017), for instance, estimates that the total number of devices connected to the internet will reach 20 billion by 2020. The near ubiquity of such devices in our surroundings - in combination with the often staggering volumes of data being collected by them - is in turn raising the demand for systems capable of routing, exchanging and processing data, and - if necessary - also handling some of the associated actuation autonomously, without the need for constant human attention.

It is in this interconnection and autonomous coordination of such smart internet-enabled devices where many researchers believe lies their greatest potential (Rajkumar et al., 2010). The *Internet of Things*, as this emerging network of interacting and cooperating everyday objects frequently tends to be referred to, has been noted for its potential to improve the efficiency of existing services as well as enable whole new ones (Akyildiz and Kasimoglu, 2004). Drawing on the new potential offered by this technological landscape, a deluge of solutions is currently emerging with the

ambition to intersect and enhance nearly every facet of the mundane life. A smart fridge capable of sensing when milk is running out and sending a notification to its user (Suhuai et al., 2009), or a smart shoe measuring the number of steps taken and distance travelled and logging this information to the user's online fitness account (Kuniavsky, 2010) are just two examples illustrating this ongoing trend.

With such solutions beginning to find their way into domestic environments, the term *smart home* is starting to gain increased traction both in the industry and academia. A term which Frances K. Aldrich (2003) defines as "a residence equipped with computing and information technology which anticipates and responds to the needs of the occupants, working to promote their comfort, convenience, security and entertainment through the management of technology within the home and connections to the world beyond". In practice, this translates into a range of autonomous solutions, including controlling and regulation of indoor lights, air conditioning and heating appliances, media consumption recommender systems as well as various safety features, such as smart gates and locks (Felfernig et al., 2013; Gebhardt et al., 2014; Zanella et al., 2014; Mahesh, 2014). Market analysts have long been flagging that such solutions for augmenting domestic environments have a realistic potential to become a leading growth area in computing (Crabtree et al., 2003). Zion's Market Research (2017), for instance, estimates that while in 2016 the global value of the smart home business amounted to USD 24.1 billion; by 2022 this number will have risen to USD 53.45 billion.

While there is little doubt that autonomous solutions have the potential to do a lot of good in a domestic context, scholars have also been warning that the ongoing migration of smart technologies from workplace environments into the home may introduce a range of novel challenges. The fact alone that much of autonomous technology was originally designed with offices and other workplaces in mind, rather than people's homes, has sparked debates regarding the extent to which user requirements in these two distinct contexts overlap (Hindus, 1999). Bill Gaver (2001), for instance, argues that "there is a danger that as technology moves from the office into our homes, it will bring along with it workplace values such as efficiency and

productivity at the expense of other possibilities". In other words, Gaver argues that while efficiency represents a key priority in a work setting, we ought to not automatically assume the same has to apply in a domestic context. This suggests that designing acceptable solutions for the home will require a re-specification of user requirements and a refocusing of existing design goals and priorities.

From its inception, the notion of domestic autonomous systems has already led to a number of challenges being identified and elaborated. Here, we choose to group these into three broadly defined and loosely overlapping key problem areas in the form of *coordination*, *control* and *accountability*, each of which has a long-standing history in HCI and is likely to play a vital role in shaping the acceptability of future autonomous systems for the home.

Coordination constitutes a key problem area revolving around the need for systems to be capable of frictionless integration and synergy with the setting into which they are embedded. Compared to most workplaces, homes exhibit a markedly social character, whilst also featuring entirely new user groups, including infants, elderly or disabled people (Crabtree et al., 2003; Crabtree and Rodden, 2004). The domestic daily life is thus being governed by a set of intricate social routines that in turn raise the demand for autonomous systems to possess a sufficient level of social awareness ensuring a smooth integration into this space.

Control represents another long-standing problem area associated with autonomous systems. Although the prospect of delegating tedious and repetitive daily tasks to autonomous systems might often seem appealing, scholars tend to agree that some level of control must be maintained in the hands of the user, such as to correct system errors (e.g. Holden, 2009). Future interface mechanisms will thus have to allow the user to converge on a suitable balance between system autonomy and human control (Yang and Newman, 2012).

Accountability of autonomous systems constitutes another major problem area that is in the process of attracting growing attention from academia as well as the industry. Systems operating autonomously, without constant oversight of the user,

run an increased risk of being perceived as illegible and difficult to decipher by their users (Ashraf et al., 2018). In combination with the fact that autonomous system behaviour typically requires a degree of personal data collection, this has been suggested to cause a range of potential problems, including impaired privacy, security and trust (Crabtree et al., 2017). Significant efforts are consequently underway seeking to design systems capable of explaining their own behaviour to their users and thus render themselves accountable (Biran and Cotton, 2017).

The following subsections will be dedicated to examining the coordination, control and accountability problem areas, along with related research efforts, in greater detail.

2.3.1 Coordination

In this section we bring to attention some of the routines that are typically found in a domestic environment and the challenges they pose for future technology deployments.

The intricate nature of a domestic environment has led Edwards and Grinter (2001) to argue that avoiding - or at least minimizing – the frictions that might arise from novel technological deployments in this setting will only be possible if we ground our designs in realities of the home. Much of the pertinent research in this strand has consequently focused on developing a better understanding of the everyday domestic life, or – as Crabtree et al. (2003) put it – on mapping the "stable and compelling *routines* of the home".

As Tolmie et al. (2002) explains, routines "are the very glue of everyday life, encompassing innumerable things that we take for granted such that each ordinary enterprise can be undertaken unhesitatingly. This is especially pertinent in the home where the highly disparate priorities of family members have to be coordinated without the commonality of an orientation to some shared work objective to bind them together. Routines help provide the grounds whereby the business of home life

gets done. Routines mean that people can get out the door, feed themselves, put the children to bed, and so on, without eternally having to take pause and invent sequences of action anew or open up their every facet for inspection or challenge or to constantly have to account for what they are doing with explanations or rationales".

The fundamental role that routines play in the functioning of a household means that a technical solution deployed in this setting is highly likely to get intertwined with them at least to some extent. Consequently, it seems reasonable to assume that a detailed knowledge of these routines will be critical in order to design solutions that would blend-in and complement the domestic life in a frictionless manner.

Following this line of reasoning, a large volume of empirical enquiries have taken the form of ethnographic fieldwork studies, seeking to examine the social and cultural life in the home with the goal of generating better understanding regarding just what routines consist of and what their role in coordinating the domestic life is. In the remainder of this section, we shall take a closer look at a few such studies that are representative for this research vein. Although many of these research initiatives were undertaken with the goal of informing the design of general ubiquitous computing solutions, rather than exclusively autonomous systems, their concern with the mundane domestic routines has helped surface findings that are highly relevant to the subject of this thesis.

Grinter et al., (2005) for instance, carried out a study of the process by which households adopt complex networks of computer and audio-visual technologies, while paying particular attention to the impact these novel solutions have on the daily routines. The study found that successfully integrating such networks into a domestic environment requires significant, collective household effort. Not only do the household members have to figure out how to coordinate the use of the technology, but also its setup and maintenance. Grinter et al. (Ibid.) conclude that for any technology to be successfully adopted by household members, it is essential for such social coordination to be worked into the routines of the home.

The importance of supporting such coordination between individual household members was further elaborated by Crabtree et al. (2003) who examined the "socially organized production and consumption of communication" that goes on in a domestic environment. In doing so, they shed light on the use of key home-based communication technologies; including paper media, digital media or telephones, and the role these play in the ecology of the domestic space. Drawing on its findings, the study identified "prime sites" in the home for ideal placement of prospective ubiquitous computing solutions and elaborated a set of design questions, the answering of which, they argue, should be prioritized by existing and emerging research agendas, including "What might it mean to make the digital more prominent throughout the home?", or "How may a system best represent the ongoing 'work' to household members in order to support the management of activities within the home?".

Research enquiries concerned with domestic routines have not been restricted to the home itself - in the sense of a physical building - but have likewise encompassed those routines that transcend the domestic space. Hyland et al. (2018), for example, carried out an ethnographic analysis of the "methodical ways in which domestic shopping is socially organized". This entailed the full shopping and consumption lifecycle, including mapping out when items are used in the home, disposed of, arranged to be bought, found in a store, chosen, paid for, unpacked and placed back on the shelf in the home. By mapping this mundane conduct, the study surfaced a surprisingly intricate process of "need-anticipation" carried out collaboratively - and socially - by members of the household in order to converge on shopping lists of items that ought to be purchased. Monitoring of food items in the home was, for instance, found to be reliant on division of labour, with each household member being implicitly expected to make others aware whenever an item he or she uses runs out and needs to be restocked. E.g., adding nappy bags to a shopping list may for example be seen as the responsibility of the mother in a family. Furthermore, idiosyncratic factors, such various peer pressures, were likewise found to play a major role in need anticipation, with family members frequently being found to

purchase the same food items as others, simply to avoid feeling “food-envy”. Hyland et al. (Ibid.) conclude that the practice of need anticipation always relies on implicit local knowledge (unspoken rules) that can scarcely be inferred proactively by any system. As such, he argues, it might be unwise to attempt to automate domestic decision-making entirely. Rather, we should design proactive systems that grant their users the ability to inspect, influence and - if necessary - “take control of machine decisions on demand”.

The empirical work outlined above makes it perspicuous that any solution expected to operate viably in the context of homes will have to be designed to meet a range of unprecedented requirements, such as supporting social coordination involved in the often intricate domestic routines, refraining from interference with existing division of labour and respecting the often unspoken local knowledge, including various social rules.

2.3.2 Control

In this subsection we take a look at the changing role of user interfaces in the age of ubiquitous computing and autonomous systems along with some of the key pertinent challenges.

The term "interface" in the context of HCI has traditionally been seen as entailing two domains: human users working on various tasks, and parts of a computer system (Bowers and Rodden, 1993). As the ubiquitous computing paradigm continues to gain traction, this view is however being increasingly challenged. Indeed, since its inception, the guise and role of the computer interface have been evolving hand in hand with progress in computing technology. This relation was perhaps best described by Jonathan Grudin (1990) in his landmark paper “The Computer Reaches Out”, which examines the advancements that have taken place in user interface research and development from the first appearance of commercially available computers in the 1950’s to the dawn of ubiquitous computing in the 1990's.

Grudin argues that this evolution has been characterized by "a continuity in the outward movement of the computer's interface to its external environment". This process, he elaborates, can be roughly categorized into 5 key stages: "Interface as hardware", "interface as software", "interface as terminal", "interface as dialogue" and "interface as a work setting".

As Grudin explains, the early days of the computer saw the principal focus being placed firmly on the hardware, only to gradually shift towards software and eventually converge on the user interface. While the first computer users were engineers who required relatively full understanding of the hardware, the widespread appearance of interactive systems and non-programming end-users helped push the development of accessible interaction mechanisms with attention to perceptual and motor issues, such as the display and the keyboard. Eventually, as computers moved into social or work settings, it became necessary for software solutions to support not just a single individual, but frequently a larger group or organization.

Grudin (Ibid.) concludes his reflection on the evolution of user interfaces by declaring that "In a sense, the computer is colonizing its environment, or, less threateningly, computers are progressively learning more about the world around them". This shift, he argues, has brought with it a range of new challenges. Social, motivational, economic and political factors arose that did not affect traditional single-user applications. Computer solutions are now for instance expected to be controlled by groups of individuals with a wide range of roles, skills, backgrounds or preferences. This represents a substantial complication, for as explained by Grudin, group processes are often variable and context-sensitive, and usually unfold over time and in different locations. Designing interfaces that would meet the needs of such diverse groups of users is becoming of a key importance.

This line of thought was further elaborated by Bowers and Rodden (1993), who argued that the migration of computing solutions into complex social settings raises the need to redefine the objective of HCI as a field. Instead of its traditional narrow

interpretation as "the study of user interfaces", they argue we ought to adopt a much broader perspective. To illustrate their case, Bowers and Rodden describe one of their previous user studies which explored the use of computer technologies during GP appointments. The study surfaced a large number of occasions where doctors using a computer device, such as for the purpose of writing up a medication prescription, got interrupted by their patient once they hit the *return* button on the keyboard. The patients, in other words, appeared to assume that the sound of the *return* button signalled that their doctor had finished writing a paragraph and would therefore be more receptive and willing to answer their questions.

Bowers and Rodden explain that the *return* button did thus not only end up facilitating the interaction between the doctor and his computer, but likewise served as a medium helping to coordinate the interactions between doctors and their patients. This complex pattern of interactions that surrounds a computer solution deployed into a social setting prompted Bowers and Rodden to argue that the familiar notions of a user interface as a boundary between an individual user and a computer are beginning to lose ground to more “socialized concepts”.

Bowers and Rodden conclude that it is “interfacing”, rather than “interfaces”, that ought to be seen as the focal point of HCI. The most important issues are increasingly concerned with the properties of the system relevant to the coordination of people's work activity, rather than individual cognition. Given this conundrum of human users and technologies, finding a control mechanism that would accommodate the needs of the different parties involved is thus emerging as a central design challenge.

The arrival of autonomous systems appears to add yet another layer of complexity to this challenge. As Pattie Maes (1995) explained in her talk titled *Interacting with Virtual Pets and other Software Agents*, the home of the future will be "half real, half virtual”. The virtual, she argued, won’t just be a passive data landscape waiting to be explored by us. Instead, it would be full of software agents capable of sensing the physical environment and interacting with it. This development raises important

implications for HCI in that instead of users manipulating and controlling systems directly to accomplish their goals, they will increasingly often manage autonomous systems that manipulate things and accomplish goals on their behalf.

Lucy Suchman (2007) follows a similar line of thought, arguing that the idea of personified computational artefacts attributed with a capacity to exercise intelligent, interactive behaviour has materialized into computational agents that serve as a kind of personal representatives - or assistants - of their human users. The fact that machines are now increasingly capable of acting and interacting with others, Suchman explains, has resulted in a situation where computational artefacts are interactive in roughly the same way that we are. As Button et al. (1995) elaborates, this process has been further strengthened by the widespread tendency to try to overcome any technical limitations of autonomous systems by endowing them with more of the cognitive abilities attributed to humans.

Abandoning the traditional notion of a control interface in favour of such interactive social agents might seem like a necessary step towards the development of new control mechanisms that would be better suited for complex social settings, such as the home. Scholars have however also warned that such a paradigm-shift introduces novel issues that go beyond mere practicalities (i.e. finding interaction interfaces that would best work in this emerging socio-technical ecosystem), to instead encompass a range of ethical concerns that ought to be dealt with. Suchman (2007), for instance, argues the acceptance of autonomous social agents could end up tilting the balance of power from the user towards the developer and potentially lead to reshuffling of societal hierarchies in a detrimental manner.

Although the studies described above were principally concerned with traditional work settings, rather than homes, it is reasonable to assume that much of the social complexity applies to the domestic space too. Autonomous systems carrying out actions proactively in people's homes will inevitably impact and be felt by a wider range of people, beyond just the primary user. This potential collateral impact of autonomous systems will thus have to be factored in by any prospective control

solution. Furthermore, the work outlined throughout this section makes it clear that the challenges faced by future autonomous systems do not merely revolve around the need for interaction mechanisms that would support complex social settings. It seems equally important that autonomous systems are designed in a manner that would preserve the existing social order and protect the agency of individuals.

2.3.3 Accountability

In this subsection we present an overview of literature concerned with limited intelligibility of autonomous systems and explain why this constitutes a key human challenge to be solved by future autonomous systems.

Mark Weiser (1991), by many regarded as the forefather of ubiquitous computing, famously envisioned a future with our everyday environments being augmented through a wide spectrum of computational resources. Instead of subjecting users to a constant stream of complex information, he argued such computing technologies ought to operate calmly in the background and only reach out for our attention when and where we needed their assistance.

While Weiser's vision does share some common denominators with the socio-technical landscape we now see emerging in the wake of the widespread adoption of autonomous systems, increasingly many researchers are also warning that the calm system operation he championed might come at a potentially steep price. As Bellotti et al. (2002) explain, the absence of direct human control can result in a situation where users have great difficulty reasoning about and understanding system behaviour and its underlying causes.

It might be tempting to conclude that a key design goal should thus revolve around revealing the inner workings of autonomous systems to their users in the name of greater transparency. Experts, however, have warned that such an approach is not merely technically problematic to achieve, but could easily turn into a distracting

element, inadvertently swamping the user with information and making the coexistence with autonomous solutions untenable in the long run (Goetz and Sadoski, 1995). Much of this discourse has been influenced by Alvin Toffler's seminal book *Future Shock* (1970), which describes the feeling of disjunction and disorientation stemming from the rapid social changes brought forth by technological acceleration. Toffler argued that the increasingly dynamic nature of our world would eventually overwhelm us with information, leaving us confused, unable to understand the underlying processes governing our world and derail our chances of making informed and sound life decisions. Toffler termed this state "information overload".

The notion of information overload has since been adopted and avidly discussed in the context of computing, with scholars exploring and elaborating the often fine line between keeping users informed about the operation of their various computer systems and at the same time avoiding overwhelming them with information (e.g. Malone, 1986; Maes, 1995).

As the number and complexity of autonomous systems integrated into a daily life context expands, the amount of data collected, transferred and processed is indeed increasingly becoming too large for a human user to reasonably make sense of, let alone act on (Edmunds and Morris, 2000). Instead, a growing amount of data-processing and actuation has to be delegated to machines and carried out without direct participation of its users (Woods et al., 2002). Consequently, much of the computing activity surrounding a smart home is becoming increasingly concealed from the user's immediate attention.

Besides coping with the complexity and large volume of processed data, another frequently cited reason contributing to the impaired intelligibility of autonomous systems comes in the form of the limited understanding most end-users have concerning the technical workings and capabilities of such solutions (Rodden et al., 2013; Anjomshoae et al., 2019). The general consumer unfamiliarity with autonomous technology and the sometimes complicated control mechanisms have

been noted to represent a particular challenge in a domestic context (Gunge et al., 2016). Although an improved general awareness and knowledge of AI and its implications may be desirable, some scholars argue that there are limits to what the public can be reasonably expected to learn with regards to an often highly technical subject (Gao and Bai, 2014). While this problem is understandably the most pressing amongst the older generations born before the spread of digital technologies, recent studies suggest that even young people are often overwhelmed by the growing extent of AI in automated decision-making processes today (House Of Lords Select Committee, 2018).

While drawing on these findings, the *2017 House of Lords Report on the State of AI in the UK* (Ibid.) reaches the conclusion that "achieving full technical transparency is difficult, and possibly even impossible, for certain kinds of AI systems in use today, and would in any case not be appropriate or helpful in many cases". Indeed, it is, for example, reasonable to assume that users reading off a smart energy meter will not be particularly interested in knowing details regarding the network protocol that is being employed. Some level of black-boxing is therefore not merely necessary to avoid overloading users with information, but can also help filter away information that is not contextually relevant given the user's situation.

A key strand in the research concerned with autonomous systems has hence been revolving around the idea that it is not *transparency*, but rather *accountability* that ought to be the key requirement for intelligible autonomous systems. While the word *accountability* in ordinary English is often associated with justification, in HCI it generally refers to the legibility or intelligibility of computer systems to end-users (Dourish and Button, 1998). HCI researchers have consequently coined the term *computational accountability* (Crabtree et al., 2016) to distinguish this specific kind of accountability. Such a notion of accountability has since gained considerable traction in HCI and Computer Supported Cooperative Work (CSCW) literature, with scholars frequently putting significant effort into elaborating what exactly accountability consists of and how to best design it into future systems (e.g. Friedman, 1997; Bellotti and Edwards, 2001). Much of this work has been advocating

approaches that would allow the user to retrospectively verify the nature and meaning of particular system actions by relying on various forms of post-hoc explanations. Weitzner et al. (2006), for instance, have previously used the term "information accountability" to describe the user's ability to check "whether the policies that govern data manipulations and inferences were in fact adhered to".

In contrast, Dourish and Button (1998) have argued in favour of representational models that would make system behaviour intelligible as it happens, rather than in hindsight. This notion became a centrepiece of their *Technomethodology* framework. Technomethodology emphasizes the importance of a system's ability to give accounts of its own work in order to reveal more of the mechanisms unfolding under its hood. System designers are in this sense encouraged to develop solutions capable of giving accounts of some of the complexity implicit in their actions while balancing black boxing with transparency in order to achieve *translucency* of computer systems. Dourish and Button (Ibid.) conclude that such a translucent nature of a computer system is vital for achieving a sense of accountability, explaining that "just like social interactions in everyday life, we would like our interactions with computer systems to be organized in terms of abstractions that are supported by their own unfolding, rather than opaque black boxes".

In practice, this means that rather than simply revealing all aspects of their operation, computer systems ought to utilize a level of abstraction to produce representational models communicating relevant information to their users, while keeping much of the unnecessary complexity hidden. Any potentially important change of the system behaviour should then be accordingly reflected by a corresponding change in its representational model. It is this representational model, as Dourish (1997) explains, that becomes a vehicle for conveying accounts to its users.

The role of an account, he argues, would be to allow the user to see not only what their system is doing, but also why it is being done and what is likely to be done

next. This continuous offering of such accounts is then what lays the ground for an accountable computer system.

While the ideas presented by Dourish were mostly developed with traditional computer systems - such as the desktop computer - in mind, there can be little doubt that they contributed towards establishing a solid foundation upon which the exploration of interface metaphors and other accounts for explainable autonomous systems can be based.

2.4 Responding to the Challenges

In the previous section we focussed on laying out three broad problem areas spanning some of the key challenges hampering the acceptability of autonomous systems. In this subsection we shall drill deeper and look at some of the concrete steps that have been taken in an effort to tackle these challenges.

2.4.1 Coordination

Here, we focus on some of the key design challenges and principles pertaining to coordination that have been produced through previous research enquiries.

Aforementioned research work concerning domestic routines has shown that conducting mundane household activities constitutes a complex socially coordinated practice that involves multiple factors, including relationships between members of the household and their mutual coordination. Actual deployments and studies of autonomous solutions have helped further corroborate this fact as well as bring up additional challenges.

One area in this vein that has attracted a particularly strong interest revolves around autonomous systems for regulation of home heating, ventilation, and cooling

(commonly referred to by the acronym HVAC) along with the associated energy consumption management, no doubt at least in part due to the fact that HVAC systems represent one of the first categories of domestic autonomous solutions that is gaining traction amongst mainstream consumers (Ahmad et al., 2016). Research enquiries into the use of related technologies were able to document a large number of instances where even the most basic autonomous solutions failed to gear-in with the everyday domestic life without causing considerable frictions (Rodden et al., 2013).

Some of the evidence showing that supporting social coordination represents a non-trivial matter can for instance be found in a fieldwork study carried out by Costanza et al. (2016). The study sought to assess autonomous interactive systems that enable users to do their laundry at the cheapest possible price by automatically charging smart home batteries at times when electricity price is the lowest. It was found that in spite of its good intention, thus designed system introduced a number of tensions into the household, such as individual household members competing for a particular laundry time-slot. The study concluded that a frictionless use of any autonomous solution is often dependant on finding a "balance between utility and convenience". In other words, just because a particular solution is designed to improve the convenience of its users does not automatically mean that it is useful (or even usable) within the social context of a domestic setting.

Similarly, a study by Rayoung Yang et al. (2013) highlighted the frustration that users frequently feel when dealing with the autonomous Nest thermostats. The intelligent features were generally not perceived to be as useful or intuitive as expected, mainly due to the system's inability to correctly interpret the reasons and intent behind sensed human behaviour. Another factor hampering the adoption of smart thermostats turned out to be revolving around the difficulty that many users experienced when trying to understand how the Nest works.

This indicates that coordination between household users and their autonomous systems goes frequently hand-in-hand with the system's accountability. In this case, alleviating the limited intelligibility of a system is however not merely a question of

the system rendering its own actions explainable, but just as much a matter of social collaboration between the individual household users. This notion was advanced by Brézillon (1994), who argued that explanation of intelligent systems involves cooperative problem solving between human users and is thus at least to some extent a "coordination problem".

The problem of limited system intelligibility in a social setting was further elaborated by Rodden et al. (2013) in a study exploring the manner in which prospective users relate to future smart energy grids. Current power grids are largely centralized, distributing power from generators to consumers, who in turn use up the power irrespective of its fluctuating supply and demand levels in the grid. Smart grid technologies, on the other hand, have not just the capacity to monitor the situation in the grid, but can also proactively and dynamically optimise power consumption in a household, thus helping to generate savings for their users. In doing so, smart grid solutions will however inevitably expose users to many of the energy-related complexities that would otherwise remain hidden. Through a set of focus group sessions, the study sought to assess the extent to which users might be able to accept this complexity and adopt such proactive solutions. It was revealed that although prospective users see autonomous systems as helpful tools in dealing with complex energy infrastructures, their willingness to adopt such solutions is not merely held back by issues pertaining to coordination with other household members, but also by a lack of trust in the companies providing them. This suggests that acceptable domestic solutions would not merely have foster coordination internally between the individual household members and their autonomous systems, but likewise between the household members and relevant parties from the external socio-economic infrastructure, including the responsible service providers.

Rodden et al. (Ibid.) reached the conclusion that such a level of coordination is necessary to help foster a sense of trust. Based on their findings, they articulate four broad design principles, which they argue represent a step in the right direction. Firstly, the ownership, intent and permitted activities of autonomous systems must be articulated, making it explicit who owns and controls the system and its stated

aims and limits. Secondly, autonomous systems should promote and support an open infrastructure, so that if consumers do not trust a particular action, they should be able to easily replace this with an alternative. Thirdly, systems need to be designed from the outset to provide users with understandable accounts of their actions. And fourthly, systems ought to provide a strong dialog-based approach to consent as an on-going process, making sure users are always aware of their right to withdraw or renegotiate the levels of consent provided.

2.4.2 Control

In this subsection we examine the different interaction mechanisms that have been advanced by existing research in an attempt to come to grips with the proliferation of autonomous solutions.

While a key selling point of autonomous systems is by definition their ability to carry out our tasks for us, without requiring any direct assistance, relevant literature is generally in agreement that some level of human control is nonetheless desirable.

Whitworth (2005), for instance, stressed the need for at least partial human control over the computer, pointing out that computers have typically a poor understanding of context, which may result in faulty actions if left unattended. A degree of human control is in other words often necessary to correct any mistakes made by an autonomous system. Other studies have confirmed this, finding that human-robot teamwork has a strong tendency to yield better results than just autonomous systems alone (Johnson et al., 2014).

Similarly, Rayoung Yang et al. (2013) carried out a comparative study, examining the differences in people's interactions with smart thermostats in comparison to traditional thermostats. The study uncovered the often fine line that autonomous systems need to walk in order to pursue their objectives without annoying or alienating the user. The use of interfaces, such as the implementation of web or mobile applications, resulted in substantial improvements to the overall user

experience. The authors conclude that the use of such interfaces is often needed to converge on a balance between automation and user engagement.

One such approach that has seen some success in the domestic context revolves around the development of tools for end-user programming. Such control mechanisms typically allow the users to define some form of programming rules without requiring them to write any actual computer code. Dey et al. (2006), for instance, proposed the *iCAP system* - a visual, PC-based and rule-based system for building context-aware applications for the home without the need for any programming. iCAP is based around the use of *trigger-action rules*, allowing the user to predefine desired system responses for particular stimuli.

Similarly, Lee et al. (2013) developed the *GALLAG Strip* - designed as an "easy to use mobile and tangible tool that allows users to create context-aware applications without the need of programming experience." Other studies, (e.g. Ur et al., 2014) have since helped further validate the use of such end-user programming solutions in the domestic context. Arguably the first solution in this vein to achieve a widespread level of popularity comes in the form of the IFTTT website, which allows its users to build a range of customized programs (Ur et al., 2016).

Nonetheless, such tools have also been noted for typically only providing the user with a limited set of options (e.g. Amershi et al., 2014), which might pose a problem when exercising control over complex tasks, such as those requiring machine learning.

This problem was perhaps best elaborated by Eric Horvitz (1999) in his landmark paper on the principles of *Mixed-Initiative User Interfaces*, which adopted a different approach to tackle the tensions between the development of new interface metaphors and tools on the one hand and the development of systems capable of operating autonomously on the other. Instead of siding with either of these paradigms, Horvitz advocated searching for "valuable synergies" between both of them. This search prompted him to coin the concept of "mixed initiative", centred

around the notion that intelligent services and users may often collaborate efficiently to achieve the user's goals.

Horvitz went on to identify 12 factors which he argued are critical for an effective integration of autonomous systems with traditional user interfaces. These factors include for instance a system's ability to consider uncertainty about its user's goals; the ability to infer ideal action in light of costs, benefits, and uncertainties; or the ability to employ dialog to resolve key uncertainties.

Upon studying these factors in practice, Horvitz concludes that it is the level of uncertainty that constitutes the determining factor in converging on a suitable balance between automation and human control. To demonstrate his point, Horvitz analyses the application *Lookout*, which serves to automatically help the user sort email messages while utilizing various levels of automation. Whilst amongst the key advantages of its autonomous features is its ability to predict or estimate the actions a user is about to carry out and based on this prediction take proactive measures, there is frequently a great deal of uncertainty in these estimates. Horvitz essentially says that automation should therefore be applied when the uncertainty is low, and the user should be left in full control when the uncertainty is high.

Another alternative approach to the study and development of user interfaces comes in the form of the newly emerging field of *Human Data Interaction* (HDI). An axiom of HDI is the realization that both consciously and unconsciously, through our interactions with emerging technologies, as well as with each other, we are consistently generating trails of data about us and our activities (Crabtree et al., 2015). This in turn results in collections of digital data so vast that the possibilities they bring have been described by some scholars as "the next frontier for computing and society alike" (Mortier et al., 2016).

HDI responds to this challenge by seeking to provide mechanisms that would allow individuals - as well as groups of people - to interact explicitly with these data collections. The key concern of HDI has, in other words, little to do with management and use of any particular computing device. Rather its goal is to place

the human user at the centre of the relevant personal data flows that exist in our dynamic and unpredictable society.

McAuley et al. (2011) explain the importance of this shift, arguing that data is not so much an object in itself, as it is an object embedded in human relationships. Or as they put it themselves: "more often than not it does not relate as much to 'me' or 'you', but rather to 'us'". Much in line with the shift from the interface towards interfacing advocated by Bowers and Rodden (1993), a HDI approach thus allows us to focus on the interactions between individuals and the data surrounding these interactions, rather than on the interplay between a single individual and a computing device.

One of the earliest attempts to translate the key ideas behind HDI into a hypothetical solution can be found in the work of McAuley et al. (2011) who proposed the "Dataware" as an interaction model highlighting the role of key parties involved in any viable HDI solution. McAuley et al. argued that any interaction involving data collection fundamentally revolves around three such key parties: the "user" by or about whom data is created, the "data sources" generating and collating this data, and finally the "data processors" entailing any parties that wish to make use of the data.

While never actually deployed, the Dataware model proved helpful in exploring some of the central interactional and technical issues surrounding HDI, including the need to support users by surfacing and articulating their personal data trails. McAuley et al. acknowledge that a better understanding needs to be developed regarding what exactly should be articulated. It seems obvious, however, that per minimum this will entail visualization of the data being taken by processors from data sources as well as visualizations that would help users make sense of the data usage. This would in turn serve as a necessary precondition for enabling the user to responsibly negotiate permissions to data processors or to control the environment in which the data processing gets executed. A prototype solution under the name 'Databox' has been developed in an effort to put the Dataware model to practical use and is showing potentially promising results (Crabtree et al., 2018).

Given the scale of data collection in our society, it seems reasonable to assume, however, that users will not only need tools that would enable them to manage their own personal data sources, but also allow them to delegate control of their data sources to others without impeding a sense of accountability.

2.4.3 Accountability

In this subsection we discuss the ongoing efforts to make behaviour of autonomous systems intelligible to their end users.

Discussions surrounding the problem of intelligibility have a long standing history in the development of novel computing solutions. In his landmark book on the psychology of everyday things, Don Norman (1988) described two *gulfs* representing key obstacles hampering a state of perceived intelligibility of computer systems: the *gulf of evaluation* and the *gulf of execution*. The gulf of evaluation is concerned with the degree to which a system provides representations that can be perceived and interpreted by its users, allowing them to assess the system's state and capacity to support their intentions. The gulf of execution, on the other hand, entails any discrepancies between the intentions of the users and what the system actually allows the users to do, i.e. whether and how well it supports the user's actions.

Although Norman's gulfs capture the essence of intelligibility issues surrounding traditional computer systems, the emergence of AI is in the process of introducing novel accountability-related challenges revolving around numerous ethical and social concerns. Pattie Maes (1994), for instance, warns that users delegating their tasks to opaque autonomous solutions are often left with no choice but to fully trust in their system's ability to act responsibly in the user's best interest. As Muir (1994) warns, besides causing potential misuse, the failure of a system to repay such trust will also inevitably undermine its credibility and spark apprehension amongst its users. Lim et al. (2009) go a step further, arguing that besides undermining trust, limited system legibility can also easily diminish user's satisfaction and ultimately

degrade the prospects for acceptance of autonomous systems. Recent studies are indeed indicating that the perceived risks of relying on autonomous computer systems might in some cases be so grave that instead of choosing to trust them, users simply decide to refrain from their adoption altogether. As Doran et al.'s (2017) rhetorical question illustrates: "If you were held accountable for the decision of a machine in contexts that have financial, safety, security, or personal ramifications to an individual, would you blindly trust its decision?"

Such scepticism is by no means unfounded. As the political activist Lawrence Lessig (2009) noted, "code is law" and the architecture of systems and the algorithms that run them are having a powerful influence on personal liberty. With autonomous systems continuing to gain traction in our society and AI algorithms becoming responsible for increasingly consequential decisions in our lives, it seems inevitable that the degree of this influence will only continue to grow. Machine learning algorithms are now being employed to carry out a wide spectrum of tasks, including the handling of online review systems, educational evaluations, the operation of markets, the running of political campaigns and even the management of social services concerning areas such as welfare and public safety (Diakopoulos, 2014). In spite of this trend, scholars have been flagging that the general public still lacks clarity about how exactly AI algorithms use our personal data and exercise power over our daily lives (Pasquale, 2015).

The gravity of this matter has been further exacerbated by numerous high profile incidents, such as the Snowden revelations, making the public alive to the extent to which personal data collection could be misused for ulterior motives (Bauman et al., 2014). Although some of the negative societal impacts associated with autonomous systems can be described as incidental (Kiss and Berecz, 2019), others, such as the use of biased algorithms benefiting a specific group of stakeholders, have indeed been found to be intentional (Vakkuri and Abrahamsson, 2018). Media outlets are regularly fuelling such mistrust by pushing a narrative of computer systems being used for widespread exploitation of end users along with a variety of privacy

infringements, thus further undermining the trustworthiness of AI and computer systems in general (e.g. Lowe, 2018; Datta et al., 2016; Ashraf et al. 2018).

While information campaigns seeking to raise awareness of the security risks associated with the use of autonomous technologies are frequently employed in an attempt to help mitigate some of the potential risks associated with it (Trend Micro, 2018), researchers have also voiced concerns that this results in a situation where the public is being presented with overly negative or outlandish depictions of autonomous systems (House Of Lords Select Committee, 2018). This, in turn, could foster a general climate of mistrust towards all autonomous solutions, even in cases where it is unfounded, and thus diminish the prospects for their widespread adoption. This has led Malcolm Fisk (Ibid.) to argue that public engagement should take on a more positive guise by instead focussing on building trust in autonomous systems. Lahijanian and Kwiatkowska (2016) follow a similar line of reasoning, arguing that developing partnerships of trust between autonomous systems and their users ought to be our key priority. The authors go on to explain that this is no easy task given that trust is generally influenced not only by objective factors, e.g. location and visibility conditions, but also subjective factors that include personal motivation, emotional state, preferences and experience of the user.

The concatenation of the circumstances described above has brought forth novel ethical and legal challenges that are in turn paving the way for an increasingly prominent research strand seeking to find a way of making AI actions and decisions intelligible. A milestone in these efforts came in 2016 when the American *Defence Advanced Research Projects Agency* (DARPA) responded to this increasingly challenging socio-technical landscape by launching an *Explainable Artificial Intelligence* (XAI) program, based around the premise that making the learned models and decisions of AI systems understandable to the end users represents a key step towards trustworthy autonomous systems: “Dramatic success in machine learning has led to a torrent of Artificial Intelligence (AI) applications. Continued advances promise to produce autonomous systems that will perceive, learn, decide, and act on their own. However, the effectiveness of these systems is limited by the

machine's current inability to explain their decisions and actions to human users. [...] Explainable AI—especially explainable machine learning—will be essential if future war fighters are to understand, appropriately trust, and effectively manage an emerging generation of artificially intelligent machine partners" (Gunning, 2017).

Although warning that the development of such explainable AI systems will be no easy task, Biran and Cotton (2017) are quick to add that this emerging field has indeed a bright future ahead of it, arguing that "current efforts face unprecedented difficulties: contemporary models are more complex and less interpretable than ever; they are used for a wider array of tasks, and are more pervasive in everyday life than in the past; and they are increasingly allowed to make (and take) more autonomous decisions (and actions). Justifying these decisions will only become more crucial, and there is little doubt that this field will continue to rise in prominence and produce exciting and much needed work in the future. "

In line with DARPA's call, a growing number of experts from non-defence-related domains, including academia and the industry, are currently in the process of advancing our understanding of the manner in which autonomous system actions ought to be best articulated and explained to the human users.

Perhaps the most immediate challenge faced by such efforts is the absence of any clear consensus regarding exactly what set of criteria a system needs to meet in order to be considered explainable. Instead, a myriad of perspectives concerning what the attributes are that a system ought to communicate to its users - as well as the manner in which this information ought to be communicated - have been brought up and discussed.

Bellotti and Edwards (2001), for instance, proposed a relatively narrow perspective, arguing that systems should be made intelligible by informing users about "what they know, how they know it, and what they are doing with that information".

Doran et al. (2017), converged on a broader definition, distinguished between 3 categories of systems based on the degree of their explainability: *opaque systems* offer no insight into their algorithmic mechanisms, *interpretable systems* grant their users

access to analyse said mechanisms and finally *comprehensible systems* emit information that enables the user to understand the underlying reasons behind each AI decision. Within this taxonomy, Doran et al. (Ibid.) argue that only the final category - comprehensible systems - can be considered truly explainable from the perspective of the mainstream user.

Lim et al. (2006) provide a more detailed perspective on how such comprehensible systems ought to be achieved. Upon studying extensively the use of explanations in autonomous systems - in particular what they refer to as *reasoning traces* addressing the *why* and *how* of machine behaviour - they concluded that such reasoning traces can be provided through 5 types of explanations, or *intelligibility questions* as Lem et al. call them: *What* explanations inform the user of what a system does, *why* explanations inform the user of why the system did what it did, *why not* explanations inform the user of why the system did not carry out a different action instead, *what if* explanations tell the user what the system would do if X happens. Finally, *how to* explanations inform the user of how a system can be made to carry out a particular action.

Based on a series of user studies, Lim et al. (Ibid.) concluded that the effectiveness and importance of *why* and *why not* explanations generally exceeds those of *how to*, *what* and *what if*. The authors argued there is a particularly clear link between the use of *why* explanations and the improvements of user's understanding and trust in the system.

Lipton (2016), on the other hand, suggested that explainability goes beyond a system's ability to simply motivate and explain its actions. Instead, he argued, explainability revolves around 5 requirements that a system ought to satisfy in order to qualify as being *explainable*. These include *confidence*, *trust, safety, sound ethics* and *fairness*.

In contrast, Shafiq and Muhammad (2015) took an alternative approach, emphasizing the need to make the general rules behind algorithmic decision-making visible, rather than explaining every action individually.

Another branch of the XAI literature is less concerned with defining the criteria for explainability and instead takes a more practical approach by developing and evaluating specific mechanisms capable of translating the processes and decisions made by AI algorithms into accounts that would be comprehensible to human users.

One of the earliest attempts at developing systems capable of such intelligible self-representation and information-disclosure was carried out by Brian Smith (1982), who proposed a general architecture, called *procedural reflection*, as a method of making some aspects of the computation explicit "right in the midst of a computation, even if they were implicit a moment earlier". This practice has shown promise in certain types of autonomous systems, such as recommender systems (Herlocker et al., 2000). Recommender systems represent a particularly well researched domain in this context. In particular, a number of scholars have helped establish a clear link between a system's ability to explain the motivation behind its recommendations, and the degree to which users are willing to trust these recommendations (e.g. Hoffman et al. 2018, Antifakos et al., 2005; Dourish, 1995; Suermondt and Cooper, 1992). Unsurprisingly, a range of commercial applications in this vein have already seen implementation. The Amazon online retail store has, for example, designed such explanations into its product recommendations (Cong and Chen, 2019).

Another well researched area that has seen successful implementation of explanations comes in the form of knowledge-based systems. A knowledge-based system (KBS), also known as Expert System, is a computer system which generates and utilizes knowledge from different expert sources in order to support human learning, problem-solving and decision-making procedures in a particular field (Techpedia, 2019). By explaining its own reasoning process, a KBS helps its user understand the rationale behind any information or recommendation supplied by the system. Gregor et al. (1999) carried out a broad review of the nature and use of explanations in this context. Much in line with the other enquiries described throughout this section, the study found that a KBS capable of explaining its reasoning and justifying its behaviour has a strong tendency to improve the

performance and learning efficiency of its user, thus resulting in more positive user perception of the system. The manner in which such explanations should be conveyed has likewise seen some discussion, with researchers frequently advocating the use of rich visualizations along with textual descriptions for optimal explainability (You et al. 2016).

While most of the research described above places the burden of explanations squarely on the autonomous system itself, another strand in the relevant literature has also focussed on a range of external factors, such as the system provider's ability to communicate its adherence to the law or an ethical code as a means for making the behaviour of AI systems seem intelligible and trustworthy in the eyes of its users. Brown et al. (2019) have for example highlighted the importance of designing AI solutions that explicitly work in concert with policy changes, which in turn, they argue, helps positively impact community perceptions of system justice. Similarly, the British responsible technology think tank Doteveryone argues in favour of shifting some burden of explanation away from systems themselves, to their providers, suggesting that organisations relying on AI-centred services should be required to declare where and how they use it, "similar to declarations of the use of CCTV or telephone recording", and ready to explain AI functions and outcomes "in plain terms" (House Of Lords Select Committee, 2018). Legal experts have further elaborated this requirement, arguing that companies responsible for AI services must explain how these services work and whether - or at what stage in the process - a human is involved, as well as provide clarity on how AI benefits the customer experience, and finally inform customers about the steps taken to safeguard their data (Ibid.).

In the light of such arguments, the idea of greater governmental oversight appears to be gaining popularity. In a recently published report, the Institute of Electrical and Electronics Engineers (2017) has for example argued that in some cases regulators ought to have "the power to mandate the use of more transparent forms of AI, even at the potential expense of power and accuracy".

Against the backdrop of the ongoing academic debates described above, the private sector is also becoming actively involved in these efforts, with a number of high profile technology companies committing to develop explainable machine learning systems. Google's *Glassbox* framework for interpretable machine learning (Voosen, 2017), and Microsoft's development of *AI principles* for intelligible systems (Zeng et al., 2018) constitute two of the more notable examples.

2.5 Synthesis

The application area of machine autonomy is in the process of growing beyond the workspace, with the lucrative smart home market stimulating development of an increasing number of autonomous products and services designed to operate in the everyday living spaces of our homes. This trend, however, has come at a cost. As the research efforts outlined throughout this chapter demonstrate, major areas of friction are being stirred up and, unless dealt with, will run a real risk of undermining social acceptability of emerging autonomous systems.

While drawing on relevant academic literature, we have loosely categorized the key challenges that have surfaced in this domain into three broad and partially overlapping problem areas in the form of *coordination*, *control* and *accountability*.

Coordination is concerned with the often intricate routines that form the fabric of domestic life. Although much of the literature in this vein is concerned with traditional computing technology, a number of findings suggest that intelligent systems capable of proactive behaviour run an increased risk of negatively interfering with the established social order in the home.

Control is primarily concerned with interaction mechanisms enabling an efficient balance between the human and machine autonomy. Instead of revolving around traditional interfaces mediating communication between an individual user and a computer device, existing research suggests that the emergence of autonomous

systems raises the need for solutions that would factor in broader human-machine ecosystems and data flows, in turn giving rise novel practical as well as ethical concerns.

Finally, accountability (in the context of HCI) revolves around the need to balance black-boxing with explanations in order to render system behaviour intelligible to its users without overwhelming them with information. Due to frequently involving widespread collection of personal data and their ability to take potentially consequential decisions, designing accountability into autonomous systems is not simply a question of usability, but likewise involves a range of other critical issues, including privacy and security.

The challenges summed up by these three problem areas all contribute towards impaired trust and broad social apprehension towards autonomous systems. Although efforts to mitigate such challenges are currently underway, the existing frictions in the domestic context remain a poorly understood domain. We take it then that further corroboration and elaboration of such human challenges is of vital importance in order to help shape and improve the acceptability of future and emerging autonomous systems for the home. By adopting these problem areas as a starting point for our own enquiry, we seek to build on previous findings and examine the nature of the above described challenges more explicitly in a domestic setting.

It seems reasonable to assume that surfacing and correctly assessing more of the user requirements surrounding the domestic context will be of vital importance in order to inform the design of acceptable solutions. As Richard Harper (2003) points out, the concept of homes augmented through digital technologies has been around since the 60's, but failed to achieve any meaningful level of popularity. He goes on to argue that this is at least partially a result of "too much technology push, and little attention given to users or usability".

Indeed, the concerns summed up in this literature review have made it perspicuous that some of the most critical challenges surrounding design and adoption of future

domestic solutions are not primarily of a technological nature (in the sense of them having little to do with any particular technological limitation), but rather have a distinctly *human* quality. Sole engineering progress will likely not be enough to guarantee a safe and productive partnership between a user and a domestic autonomous system. Rather, it appears to be our understanding of relevant human expectancies that needs to be advanced.

This in itself constitutes a challenge. As explained by Crabtree et al. (2003), the fact that existing user research methods were largely conceived with non-domestic settings in mind raises the need to develop new conceptual as well as analytical tools that would be better adapted to this novel socio-technical context. Such conceptual and analytical tools, they argue, would allow designers to gain a deeper understanding of the domestic life and how novel technologies may best support everyday activities and values of the home. In the following chapter, we shall attempt to tackle this problem by proposing a novel methodological approach to user research tailored for this particular context.

Chapter 3 Methodology

Our approach to elicit background expectancies impacting the acceptance of autonomous systems in the home merges breaching experiments (Garfinkel, 1964) with scenario-based design (Carroll, 1995) and contra-vision (Mancini et al., 2010) to create provocative visions of the home of the future that intentionally disturb common sense reasoning and create incongruities or reality disjunctures whose repair surfaces taken for granted background expectancies that impact the uptake of future technologies in everyday life. Below, we describe each of these three methodological components in detail and explain the role they played in helping us engage with prospective users. Breaching experiments

Amongst the axioms of sociology is the notion that there is an external objective world that exists beyond an individual's subjective beliefs and which is accessible to others. The process by which members of a society elaborate each other's subjective beliefs in order to maintain and reinforce shared perceptions, cognitions, and descriptions of real worldly events was described by Melvin Pollner (1987) as *mundane reasoning*. As he put it himself, "the assumption of an intersubjective world taken together with the inferential operations for which it provides comprises what we shall term an idiom of mundane reasoning." Pollner adds that any well socialized mundane reasoner will assume a world which is not only objectively present but a world to which he also has continued experiential access and which others experience in more or less identical ways.

An axiom of the empirical enquiry to be presented in this thesis is the notion that it is such mundane reasoning that holds the key to better understanding the relevant expectancies that shape users' acceptance of prospective autonomous systems.

Whilst generally implicit, Pollner (Ibid.) argues that the process of mundane reasoning becomes visible through attempts to transcend the mundane, which forces individuals to explain away any discrepancies: "mundane reason provides its practitioners with a wide range of explanations which preserve mundane reasons' stipulation that reality is coherent, determinate and inter-subjectively accessible".

One potential source of such discrepancies making mundane reasoning observable comes in the form of novel technological deployments. As Andy Crabtree (2004) explains, the emergence of a range of novel technologies (such as mobile, ubiquitous or autonomous computing) have enabled a wave of new innovative systems that have little or no grounding in existing practice. Instead, he explains, these systems "create entirely new possibilities, and the practices for their use have yet to emerge". Such absence of any established practice naturally raises serious challenges for the design and development of novel solutions.

Much like Harold Garfinkel's (1967) portrayal of the sociologist, designers commonly approach this issue by selecting "familiar settings", such as familial households or workplaces, to make prospective technological (rather than social) systems accountable and to motivate particular forms of practice. Take Mark Weiser's (1991) much-cited *Sal scenario*, which pitches a future mundane life teeming with UbiComp devices, by way of example.

In line with Pollner's (1987) concept of mundane reasoning, Garfinkel (1967) observed that what lends our everyday life its orderly character are the numerous mundane interactions taking place between individuals. No matter what the sociologist (or the technologist) might have to say about familiar settings, it is then also the case that the members of society have their own common sense understandings of these familiar settings and employ them as a scheme of interpretation enabling concerted action and the ongoing conduct of social life. Common sense understandings furnish members with background expectancies drawn upon in situ, as contingencies dictate, to both recognise events as events-in-familiar-settings and to enforce compliance with expectations of action in familiar settings.

The latter point is of particular consequence to design as it speaks to the moral ordering of everyday life. It is not then that background expectancies are only used to passively interpret and make sense of events in familiar settings, but also to actively enforce their social acceptability too. As Erving Goffman (1963) described, this collectively constructed notion of social acceptability is leading to a state characterized by conformity, with individuals manifesting a high degree of sensitivity towards any deviations or breaches of the established social order. Consequently, one of the most common principles in all social situations is for the individual to actively try to *fit in*.

We take it that this applies as much to actual events *here and now* as it does to projected technological events that implicate people located in familiar settings, and that the design of prospective solutions may therefore be usefully informed through an investigation of the background expectancies that members know and use to understand and order familiar settings of everyday life.

Indeed, given the practical indispensability and moral importance of background expectancies, we might expect members to have much to say about them. Curiously, however, Garfinkel observes that the member is typically at a loss to tell us of what specifically the expectancies consist on any occasion of inquiry; "when we ask him about them, he has little or nothing to say". Instead, these taken for granted expectancies that govern our everyday social lives exist implicitly, as tacit unspoken rules. In other words, as elaborated by Maynard and Weathersbee (2007), not do individuals only engage daily in building up *rules* for social interaction, but they are also largely unaware that they are doing so.

Garfinkel (1967) recognized that surfacing and examining these unspoken rules and background expectancies would be essential in allowing him to analyse the methods that individuals employ when constructing the orderliness of everyday life. In his own efforts to get to grips with *Ethnomethodology*, as he called this analytical approach, Garfinkel tells us that for background expectancies to become visible, one must either be a stranger to its *life as usual* character or become estranged from it. Given that he wasn't a stranger, he developed a procedure of estrangement, which

he called the *incongruity procedure* (Garfinkel, 1963) or *breaching experiments* (Garfinkel, 1964; Garfinkel, 1967). As he explained in his own words, "since each of the presuppositions that make up the attitude of daily life assigns an expected feature to the actor's environment, it should be possible to induce experimentally a breach of these expectancies by deliberately modifying scenic events so as to disappoint these attributions. By definition, surprise is possible with respect to each of these expected features." (Garfinkel, 1963)

Garfinkel's experiments were thus designed to disrupt ordinary actions by confronting individuals with behaviour that is outside of social norms. Doing so allowed him to surface their background expectancies, making them available for study. Following this recipe, he would for instance instruct his students to negotiate and bargain for standard priced merchandise in grocery stores. While typically taken aback, the shopkeepers caught up in such experiments would eventually try to reason and explain the odd behaviour of the students in an attempt to restore a sense of social order, thus revealing the nature of their social expectations.

Any form of behaviour that violates - and to some extent tears down - the existing social norms, would in other words constitute a viable breaching experiment, offering the researcher an analytical lens through which to examine the background expectancies held by individuals, along with the mundane reasoning by which social acceptability is reconstructed and enforced.

The notion of deriving insights from frictional situations arising in social settings is not without precedence in the context of design. A number of previous efforts, commonly encapsulated under the loose umbrella term *Critical Design*, have sought to convey deliberately provocative experiences as a means for eliciting deeper reflection on the underlying assumptions and values that govern the society (Dune, 2006).

Such an approach has been noted to hold potentially important leverage when applied in social settings. Boer and Donovan (2012), for instance, argue that the conflicts and tensions revealed through ethnographic research can potentially be

valuable - rather than destructive - in the sense of helping designers better empathize and understand a given user context. Moreover, scholars have noted that there are benefits in applying such an analytical lens in complex environments featuring diverse categories of users. As elaborated by Jaffari et al. (2010), more actionable formats of ethnography that would examine a diverse set of prospective users analysing social frictions are worth exploring, since such approaches can "engage a wider range of stakeholders in a shorter time frame, create a common understanding of ethnographic findings and evoke a sense of ownership over a project." Much like Garfinkel's previously described breaching procedure, critical design initiatives thus operate under the premise that tensions emerging around an area of interest in a social setting enable collaborative analysis and intersubjective reflections that could in turn help inform future design work.

One popular materialization of this approach came in the form of so called *provotypes*, defined by Mogensen (1994) as "concrete artefacts to provoke current practice". Perhaps most notably, Steve Mann (2002) utilized such provocative artefacts in his exploration of computer wearables and surveillance technologies. His provotypes, such as a head-mounted camera worn in public spaces, were designed to elicit strong responses from the user's surroundings. By actively creating situations of uncertainty and anxiety, Mann sought to reveal everyday structures of surveillance, governance and control. Mann believed that his approach would help expose current power structures and in turn galvanize the public into scrutinizing or even reversing them.

Crabtree (2004) takes the ideas behind critical design and provotypes a step further, arguing that any novel technological deployment can be treated as a breaching experiment. Even an "uncontroversial" solution can in his view provoke practice and make the relevant underlying mundane reasoning visible and available for designers to derive insight from. It is in the nature of novel technology deployments, in other words, to force users to devise explicit ad-hoc practices that would make said technologies work "here and now".

While breaching experiments are generally associated with making trouble, as Garfinkel (1963) characterised them that way, proponents of the approach in design have demonstrated that trouble is not essential or necessary for their conduct (Crabtree, 2004). Rather, and to borrow from Garfinkel, we might say that as seen from the member's perspective, compliance with background expectancies turns upon a person's grasp on what are commonly seen and understood to be the *natural facts of life in society*. It follows from this that the firmer a person's grasp on the natural facts of life in society—and we assume that a firm grasp of such facts is key to one's social competence—then the more severe should be their *disturbance* when those facts are impugned as a depiction of his (or her) real circumstances.

To illustrate the notion of treating regular technology deployments as breaching experiments, Crabtree (Ibid.) points to the mobile mixed reality game *Can You See Me Now?* (CYSMN) which was centred around the idea of having *runners* situated in the physical streets of a city chase and catch online players moving through a virtual model of the same city streets (Flintham, 2003). Communication between the two groups was facilitated through handheld computers and GPS tracking. Although the initial prototype was designed to work according to a predetermined plan of use and the technology was expected to support projected forms of interaction, its actual deployment ended up surfacing unexpected patterns of use. Due to frequent GPS blackspots where connection was impossible, runners had to resort to collaborative diagnostics of these interruptions in order to handle them. Future iterations of the game ended up taking these inconsistencies into account and designed them into the experience (e.g. by rewarding runners for catching players in areas with poor GPS coverage).

Once analysed through the lenses of a breaching experiment, we can see that the social circumstances that innovative technologies turn upon have potential implications for design. In the case of CYSMN, they have helped inform the development of mobile location-based games for real world use. Crabtree (2004) concludes that the configuration of relationships between ethnography and design leads to a research and development model where technology becomes a vehicle for

social research and the results of that research, in turn, demonstrably propel innovation and design.

It is important here to recognize the distinct analytical position represented by breaching procedures in comparison to traditional technological probes. Whereas a technology probe is centred around working technical devices that are left with people in a context of use in order to study the way that technology may be appropriated into people's everyday practices (Hutchinson et al., 2003), the goal of a breaching procedure is not to refine the design of any technical device itself. Rather, as suggested by Jeffrey and Shaowen Bardzel (2013), the technologies deployed during a breaching experiment ought to be approached as *value fictions*, where a familiar product is used to express an unfamiliar set of values, thus provoking reflection on values that currently apply, and how they might change in the future.

Mogensen (1994) elaborates this distinction further, arguing that instead of focusing on *solving* problems through commitments and negotiations, a breaching procedure helps designers refocus their attention towards "what goes wrong, why, and how does it relate to other parts of the practice". Consequently, instead of solving specific problems in a given context, it is the context and the practice that is questioned. Mogensen concludes that a logical next step from studying isolated technological deployments would be the creation of longer coherent scenarios provoking the context from different angles.

It follows then from the methodological approach described above that *depictions* of real circumstances — of which there may be many forms including storyboards, scenarios, design fictions, low-fidelity prototypes—which breach the *natural facts of life in society* may throw the background expectancies that order everyday life into relief. These depictions need not necessarily make trouble. They need only *disturb* common sense understandings of everyday life. It may then be possible, as it became possible for Garfinkel, to *call forth* or make visible the taken for granted background expectancies that order everyday life. As detailed in the following subsections, our own enquiry thus seeks to exploit established design methods—scenarios and contra-vision in our case—to intentionally disturb common sense reasoning and

surface the usually tacit and unspoken background expectancies that order domestic life and impact autonomous systems for the home.

3.1 Scenario-Based Design (SBD)

As declared previously, a key objective of this thesis is to assess issues concerning social acceptability that might arise as a consequence of a widespread adoption of future and emerging autonomous technologies in a domestic setting. Or, as Lindley et al. (2017) would put it; our goal is to evaluate such technologies as they might exist "beyond their prototypical implementation". A major practical issue implicit in this objective is the fact that developing and deploying such futuristic socio-technical settings would naturally require vast investments, both in terms of resources and time. In search of a more efficient method that would allow us to convey both prospective technologies as well as the wider futuristic settings they will be embedded into, and to make these holistic visions available for prospective users to comment on and scrutinize, we converged on the decision to base our research enquiry around a scenario-based design approach.

Scenario-based design, or "SBD" as it is commonly abbreviated, is a well-established approach towards user-centred design, which emerged in HCI during the 1990s (Carroll, 1997). Scenarios are stories about people and their activities, consisting of a setting and one or more actors (Rosson and Carroll, 2002). A scenario describes a sequence of actions and events that lead to an outcome, such as a user manipulating a computing device to accomplish a given goal.

Perhaps the most notable use of scenarios was proposed by Mark Weiser (1991) in his landmark paper *The Computer for the Twenty-first Century* which sought to highlight what our daily lives might look like in the future once augmented by ubicomp systems. Weiser described a fictional character named *Sal* as she interacted

with a range of imagined solutions, such as windows displaying weather forecasts or systems for automatic location-sharing between relevant colleagues, all of which were seamlessly integrated with her everyday life.

One of the prime reasons for such scenarios is to help designers focus attention on users and their prospective behaviours that are often left implicit in most traditional technological renderings (e.g., software specifications) of systems and their application. Scenarios can be conveyed through a range of media, including paper, slideshows, videos or storyboards. They are construed of as "soft" prototypes that provide a degree of context, exposing not only the functionality of a proposed system but also specific claims about the user experience, which may then be assessed by potential end-users (Carroll, 2000).

Beyond simply being time and cost efficient, scenarios have a range of other advantages that made them into our tool of choice. As formulated by Jack Carroll (Ibid.), scenario-based design approach comes with a number of key advantages. Two particular strengths of this method stand out for being relevant to our work:

- Scenarios are free of practical concerns. As fictional stories, scenarios are not held back by the contemporary state of technical knowledge in the same way functional prototypes are. This in turn means that they can be used to evaluate ideas that are beyond current practical possibilities, including the all-encompassing domestic autonomous infrastructure that constitutes the focal point of our thesis.
- Scenarios are open to user participation. Users can easily get directly involved in shaping scenarios, which in turn helps designers to focus on finding solutions to genuine needs and concerns. A vivid representation of an innovative concept can in this sense promote innovative thinking and help raise relevant questions. In the case of Mark Weiser's aforementioned Sal scenario, for instance, this helped trigger a widespread debate on the potential of concepts such as calm computing and eventually led to the materialization of the UbiComp research field (Blythe, 2014).

While the former of these points provided us with means for assessing some of the hypothetical socio-technical infrastructures expected to arise around domestic autonomous systems, the latter allowed us to immerse prospective users into such visions and open them up to assessment.

Drawing on these strengths, in our own attempt to convey prospective autonomous technologies and tease out some of the broader concerns and general attitudes connected to social acceptability, we produced a set of visual storyboards (study 1) and short animated movies (studies 2 and 3) depicting fictional characters living through a range of daily life situations in a future teeming with autonomous systems. By allowing us to pencil out various prospective autonomous solutions, scenarios provided us with effective means to gain feedback and assess key human expectancies without having to invest time and resources into building and deploying any actual working solutions.

Developing similar design scenarios has a relatively long tradition in HCI (Briggs and Blythe, 2013) and has enjoyed a strong popularity in participatory design activities (Mackay et al., 2000). In particular, scenarios have been noted for their ability to enable rapid communication about usage possibilities and concerns among many different stakeholders (Rosson and Carroll, 2009). Given the diverse user groups typically found in the domestic environment (Crabtree et al., 2003), such flexibility represents another key factor in favour of relying on scenarios.

Scenario-based design has however not been embraced universally, with some scholars arguing its flexible and efficient nature comes at a cost. Scenarios have traditionally been employed to summarize the results of requirement analysis (Carroll, 1997), to illustrate findings from field studies or as input for creating design rationale (Carroll and Rosson, 2003). Blythe (2014) criticizes the simplistic nature of these efforts, arguing that "such scenarios often lacked detail and the imagined users were little more than a name". In a more recent paper, Blythe (2017) elaborates his argument further, explaining that at the time when "Jack Carroll advocated scenarios as a means of reflective practice, computers were largely confined to the workplace. Since then, digital technology has become integral in almost every

human activity and the stories we tell have become more complex". Blythe goes on to argue that scenarios traditionally employ the *overcoming-the-monster-plot*, with *monsters* signifying various forms of usability concerns, such as inconvenience or inefficiency of a given computer system. Blythe argues that although many such monsters are worth fighting, it is important to be aware of the plots in order to "develop a more reflective and critical practice". Blythe goes on to suggest a more robust plot ought to be employed, allowing the reader (or viewer) to "enter into a larger world", in order to accurately capture and evaluate the complexity of social settings involving emerging ubiquitous technologies.

The notion of positioning prospective technologies within a complex narrative structure is in the process of gaining popularity both amongst designers and authors of fiction. The sci-fi author Bruce Sterling (2005) was amongst the first to propose that science-fiction might be a useful tool in the context of design thanks to its capacity to imagine future technologies and settings. In an attempt to describe how design thinking impacted his writing, he subsequently coined the term "design-fiction". Julian Bleecker (2009) was the first design practitioner to catch on this term, turning it into an actual research method and a design approach. Seeing his ideas gaining traction, Sterling (2013) followed up on this by formulating a more precise definition of design fiction as the "deliberate use of diegetic prototypes to suspend disbelief about change."

Such use of fiction has a range of advantages over the more narrow use-cases traditionally conveyed by scenarios. Lindley and Coulton (2015), for instance, explain that design fictions represent "self-contained worlds", and as such extend traditional prototyping approaches by demonstrating both the concept and context simultaneously. Similarly, Dourish and Bell (2014) used science-fiction as a lens through which to view ubiquitous computing, pointing out that what traditional design scenarios typically left unsaid is the implicit social and political context of a design.

Against the backdrop of emerging technologies frequently giving rise to novel areas of friction within their broader social settings, it is hardly surprising to see a growing

interest in using such diegetic approaches to present and explore hypothetical prototypes in plausible near futures. Consequently, a common practice is to enhance traditional design scenarios through elements otherwise associated with fiction. Blythe (2006), for example, proposes the use of "pastiche scenarios" based around the idea of borrowing literary characters and placing them in new narratives in order to "create richer and more resonant descriptions of users and technologies". Similarly, Nathan et al. (2008) advocates the use of "fictitious value scenarios" as an extension of traditional scenario-based design to incorporate critical and wide ranging set of implications.

Recognizing the limitations associated with traditional scenarios, as described above, we did borrow heavily from design fiction when producing our own scenarios. Consequently, our approach does share commonalities with some previous work in this domain, such as Blythe's (2014) work on *Imaginary Abstracts* of fictional research papers, which seeks to link design fiction more explicitly to research, and Linehan et al.'s (2014) suggestion to employ fiction as a resource to explore potential design futures.

The distinction between design fictions and scenarios is generally fluid, prompting some scholars to simply treat them as two distinct subsets of *speculative design* (Lukens and DiSalvo, 2011). Nonetheless, two factors have contributed to why we here feel our work is closer aligned with traditional scenario-based design and why we choose to refer to it as such throughout this thesis. To begin with, our work focuses on a *plausible near future*, rather than traditional fiction. In pursuit of our goal to provoke social expectancies by exposing people to future and emerging autonomous solutions, we developed our scenarios by drawing on existing enabling technologies and by extrapolating from current technological capabilities. Secondly, instead of relying on the narrative structure typically associated with design fiction in the form of *exposition, climax* and *resolution* (Alexander, 2017), we here instead seek to focus our narrative around a series of interactions between users and near-future autonomous systems.

As Reeves (2012) explains, the focus on the future and on envisioning hypothetical technologies constitutes a "characteristic future-oriented technique for design thinking", and is a particularly prominent driver of ubiquitous computing research. It is hardly surprising then that employing scenarios featuring elements of design fiction, in the manner we do here, is not without precedence. While Weiser's (1999) aforementioned `Sal scenario' represents the perhaps most salient example, other scholars have since engaged with social expectancies more explicitly. Rodden et al. (2013), for instance, have employed animated sketches in workshop settings to explore the ways in which prospective users related to emerging energy infrastructures. Similarly, Wong and Mulligan (2016) examined the use of corporate concept videos as a tool to help surface values relevant to imagined futures, such as who has power and agency with these technologies.

A common denominator of all these research enquiries is the fact that they had to navigate the uncertain nature of the future. It seems reasonable to assume that the practical applicability of any insights gained though such enquiries is heavily dependent on the degree to which the conveyed hypothetical futures actually turn out to be correct in their predictions. This line of reasoning has led a number of scholars to argue that in order to evaluate futures that are indeed plausible, it is necessary to first develop a degree of foresight (e.g. Happ, 1994). In the following subsection we shall examine this practice in greater detail.

3.2 Foresight and Plausible Futures

Designers of future technologies have frequently no choice but to deal with volatile goals that are subject to constant change. A common approach for coping with such uncertainty is to explicitly re-envision and re-evaluate a prototype continuously throughout the design process until its final deployment (e.g. Wynn et al., 2005).

This approach is not entirely unproblematic, however. In their consideration of the challenges surrounding the evaluation of future-oriented prototypes, Salovaara et al.

(2017) described a blind spot in HCI's methodology which stems from the fact that any prototype deployed in a study still exists firmly in the present world. The validity of such research enquiries, they argue, can only be assured if the researchers "gear their studies such that the primary goal is to understand the prototype in plausible future states of affairs instead of the present." To complicate matters further, the ubiquitous nature of nascent autonomous systems creates a situation where the deployment of a novel technological system can in and of itself lead to restructuring of the future states of affairs and in turn cascade into entirely new problems that may be difficult to predict. As Caroll (1995) emphasized, the process of designing novel computer systems cannot therefore be seen merely as a search for solutions to existing problems, but also to problems that do not yet exist.

By ignoring the future broader context of use, along with such not-yet-existing problems, traditional evaluation methods, in other words, run the risk of misplacing their focus by assessing a hypothetical system in the context of existing settings, while largely omitting the many dynamic circumstances that the future might bring (e.g. various social, environmental, economic or political factors).

Ignoring such wider societal factors might be feasible in cases where researchers are concerned with narrowly defined aspects of a technology. Indeed, projections based on current trends, such as Moore's law, have proven useful for estimating the performance and capabilities of future solutions (Schaller, 1997). Even though the stated scope of this thesis might seem rather down-to-earth, in the sense that we too are focussing on mere extrapolations and juxtapositions of actual current technologies, rather than any far-fetched futuristic fantasies, our literature review clearly demonstrated that emerging autonomous solutions have a habit of getting entangled with the broader society. Much in line with Salovaara et al. (2017), we thus reason that developing and conveying a plausible futuristic social context as part of our scenarios will increase the validity of the insights gained from our enquiry.

The ability to plan and accurately predict such future circumstances is crucial for the success of any organization and is consequently of a longstanding interest in a range of disciplines (Gabrielli and Zoels, 2003). The sensitive nature of many

computational solutions appears to have made such future-orientation into a particularly common stance in domains concerned with computing technologies. The fact that new computing technologies tend to "introduce new security vulnerabilities, as well as reintroduce earlier ones" has, for instance, prompted Neumann (2012) to advocate the importance of "embedding long-term thinking into our planning". Such practices are sometimes labeled under the umbrella term "futures studies". Sardar (2010) defined this term as the "study of postulating possible, probable, and preferable futures and the worldviews and myths that underlie them".

Upon analyzing the approaches to future studies employed by private tech corporations while planning their business strategy, Gabrielli and Zoels (2003) found that the development of such "foresight" generally unfolds over 6 key stages; Initially, (1) these efforts take on the form of a dialogue between internal stakeholders (such as design, marketing, engineering, or executive decision-makers). This is then followed up with (2) synthesizing ideas about new technologies and user behaviours. The resulting ideas are then (3) visualized and conveyed to the public in a manner that would abstract complicated ideas. Insights gained during this process are utilized to (4) create alternative models and spark internal as well as external debates. This in turn helps (5) create openness to multiple alternatives and shape the corporation's strategic decision-making. Finally, (6) the decision-making converged upon in this manner is communicated internally and externally in order to get buy-ins.

While our own enquiry is not directly concerned with business oriented strategic decision making, the process outlined above holds some relevance to our own approach. In our efforts to come to grips with the problem of the future, we have once again turned to design fiction in search for inspiration. Coulton (2017) argues that the key factor that differentiates and distinguishes design fiction from other forms of speculative design is its novel use of world building to create a "magic circle", or in other words as Huizinga (1955) described this term; "a place dedicated for the performance of an act apart". In design fiction the magic circle is a fictional

world within which speculative prototypes, and their users, can plausibly co-exist, thus providing us with ample opportunities to develop scenarios simultaneously conveying both future technologies and future states of affairs, along with their mutual interplay.

What then do we precisely hope to learn from these futuristic scenario-based explorations? Blythe (2014) argues that unlike traditional scientific studies, research through design will scarcely yield conclusive evidence in the sense of testing a particular hypothesis. Much like in the case of breaching experiments, Zimmerman et al., (2007) argue that the goal is not to inform the design of a product, but primarily to evaluate how current and future technologies might affect people. Dunne (2013) follows a similar line of thought, stating that rather than to show how things will be, designers should strive to use scenarios, and other forms of speculative design, to open up space for discussion. Such a discussion does however not necessarily come in and of itself. As DiSalvo explains, if speculative design is to be more than just a spectacle and provocation, it will have to present a trope which relates ideas and objects in ways that can be interrogated and challenged (DiSalvo, 2012). In our effort to build such a trope into our scenarios and open up a space for discussion, we turn to contravision.

3.3 Contravision

While the use of scenarios and other forms of speculative design represents a popular and well-established approach in the world of ubicomp technology design, it has also attracted substantial criticism for allowing designers to effectively force down their own vision of the future on prospective users. Such visions have been suggested to lack social legitimacy in that they can be narrowly focussed, may not take into account the concerns of people outside of the ubicomp community itself, such as the general public, or miss concerns from other parts of the world (Bell and Dourish, 2007).

Some scholars have gone as far as arguing that scenarios are prone to being used as vehicles for the design practitioner's "confirmation biases", enabling them to convey ideas that are aligned with their own preconceptions rather than relevant preferences of the actual users (Sutcliffe, 2003). Stuart Reeves et al. (2016) follow a similar line of reasoning, warning that should we allow this to happen, we will risk a situation where an illegitimate design vision is accepted as a legitimate end goal, thus directing present design activities "in a way as to make it come to pass as a self-fulfilling prophecy". An ill-conceived or implausible vision will, in other words, produce misleading findings regardless of how vivid and well-crafted scenarios were used to convey this vision in the first place.

One could argue that - given the complexity of real-world social settings, in combination with the unpredictable nature of technological progress - producing a *legitimate* future vision, devoid of any bias, constitutes an impossible task. It seems reasonable to assume that no scenario can accurately account for the unforeseen obstacles that might emerge in the course of widespread adoption of novel technologies. Scenarios that attempt to do so have been suggested to end up conveying distorted future visions that brush over major technical or social challenges (Reeves et al., 2016).

In search for an example, we need look no further than Weiser's (1991) aforementioned scenarios. In spite of their popularity, it is conspicuous that few - if any - of the life-changing ubicomp innovations, along with the new patterns of user behaviour, predicted by these envisionings ended up actually materializing. Bell and Dourish (2007) summed up much of this problem by arguing that Weiser's future is in fact "yesterday's vision of the future". In other words, scenarios of future ubicomp solutions run the risk of appearing both miss and under predictive (typically optimistically), in turn potentially leading to misplaced design efforts and beliefs. Yet, in spite of these lingering problems, similar "grand visions" and projections continue to shape technology development, particularly so in the ubicomp community.

In an attempt to find a solution to the issues described above, Stuart Reeves et al. (2016) advocate the development of scenarios that derive a wider social legitimation from prospective users, rather than from the designer's own convictions. It is essential, they argue, that scenarios of future systems are seen as legitimate by the broader community, rather than just by a small group of practitioners. In practice, this entails participatory approaches, opening up scenarios to criticism and subsequent refinements through the engagement of the broader public. Although non-experts are hardly any wiser than practitioners about what the future brings, increased participation will introduce a measure of organizational accountability and generate greater social legitimacy than what experts alone can do.

In a separate paper, Reeves (2012) suggests that the creation of "multiple scenarios based on different envisionings, rather than working under the implicit assumption of only one particular envisioning" constitutes another viable strategy for producing socially legitimate scenarios.

Reeves et al. (2016) conclude that employing such approaches as a means for avoiding scenarios centered around a single grand vision that lacks social legitimacy, represents a precondition for a more responsible design practice.

Following this call for greater social legitimacy, we complemented our scenario-based design approach with a *Contravision* method. Originally formulated by Mancini et al., (2010), the idea behind contravision is to develop multiple contrasting versions of the same scenario, rather than just one. Contravision scenarios typically take on the form of futuristic videos, or other narrative media, with each one of them conveying either negative or positive aspects of the same proposed technology. The different versions of the scenario are symmetrical in the sense that they all tell the same story with the same scenes and characters, but with the characters' attitudes and reactions being different in each version.

Although producing multiple versions of the same scenario generally requires greater time and resource investments, a number of upsides have been argued to make contravision worth this extra cost. For one, conveying two (or more)

contrasting, yet comparable, representations of a technology can help elicit a wider range of reactions from prospective users than what could normally be gained by conveying only one perspective (Mancini et al., 2010). This approach has also been suggested to serve as an effective tool for reducing the impact of any potential bias held by design practitioners. Much like the aforementioned concerns brought up by Reeves et al. (2016), Price et al. (2010) argues that although many existing scenarios are detailed and well-crafted, a key concern is that the narrative and context in which the technology is portrayed may "bias the reported acceptability, usefulness and usability of the technology shown". Price et al. conclude that conveying multiple perspectives of hypothetical technologies to prospective users can in this sense help produce more objective responses by its prospective users. This view was largely confirmed by Murray Goulden and Alexa Spence (2012) who utilized a contravision approach in their exploration of potential interactions that might occur between *smart grid* technologies and practices in the home. They found that contravision provides a "highly engaging frame" for focus group user studies in that it helped participants think more "openly and critically".

Besides reducing bias of the conveyed design ideas and encouraging open critical reflection, another factor making contravision into our preferred method is its suitability for assessing relevant attitudes held by the general public towards future and emerging technologies with potentially impaired social acceptability. As Mancini et al. (2010) explain, contravision represents a particularly valuable approach when researchers have reason to believe that new technology is likely to raise "subtle personal, cultural and social issues that can potentially jeopardize its adoption", such as when exploring the range of users' reactions when developing future technologies that may be controversial.

By way of example, Price et al. (2010) conducted a contravision study based around video scenarios of wearable device designed to assist with people's healthcare and wellbeing. They found that "the viewers of the two videos reported reactions toward issues such as privacy, identity, and adoption which overlapped but also spanned a much wider range". This brought them to the conclusion that the contravision

approach is “useful for eliciting user reactions to the more subtle and emotionally driven effects of future technology".

Given our interest in social acceptability of emerging autonomous systems, these unique qualities introduced by contravision made it into a staple of our engagement with prospective users.

3.4 Synthesis

Bellow we consolidate the key takeaways from the methodology chapter and elaborate our use of breaching experiments, scenario-based design and contravision as a coherent research approach.

Based on the arguments presented throughout this chapter, we centred our enquiry around three sets of contrasting (e.g. utopian and dystopian) contravision scenarios (one set for each of our 3 user studies), each containing elements that would intentionally disturb common sense understandings of everyday life. From our own perspective, portraying technology in a purposefully dystopian manner conveys a vision of the future which is essentially broken. Whereas, portraying it in a purposefully utopian manner conveys idealised and idealistic visions of the future. One might argue that our approach shares some common denominators with the “split testing” approach (also known as “A/B testing”), commonly employed in usability testing to help designers determine which of two solutions is regarded as preferable by prospective users (Xu et al., 2015). A closer look, however, will reveal that our approach differs in a number of crucial ways. Of the contrasting visions that we present, neither is right, as in correct, and nor are they meant to be. Rather, both constitute polar extremes (absurd contrasts) which are intentionally designed to disrupt common sense reasoning and surface the unspoken background expectancies that order what Garfinkel (1967) called “an obstinately familiar world” and in turn impact the adoption of future and emerging technologies in everyday life (Lindley et al., 2017). Rather than determining which of the presented visions is

preferable, our goal is thus to open up space for discussion, enabling us to stimulate and analyse mundane reasoning of our participants.

In more prosaic terms, following the call for greater social legitimacy (Reeves et al., 2016), we exploited the contravision approach as a tool for provoking our participants into commenting and critiquing our scenarios. In addition to unearthing the values people bring to bear in their judgments of potential future technologies, our goal was to invite participants to change (in the sense of renegotiate) the potential technologies to make them better reflect their own concerns and values and in-so-doing find a way for them to *fit-in* with their own lives.

By combining contravision with the breaching procedure, our work thus presents a dialectical approach to scenario-based design. With every presented *thesis* being counterbalanced by an *antithesis*, we open up room for our participants to converge on a *synthesis*. Instead of forcing our visions down on prospective users, we seek to provide them with the means to express their own preferred visions in a bottom-up manner. By employing such an approach in a focus group setting, the process becomes collaborative, effectively encouraging interactions between strangers, which in turn provides us with a way of examining people's *mundane reasoning* (Pollner, 1987) through a social or collaborative lens (Brown and Bell, 2004).

Using scenario-based design and contra-vision to enable breaching experiments in this manner provides us with the means to call forth taken for granted background expectancies impacting the uptake of autonomous systems in the home and provides a novel early development stage framework for producing foresight of the future, which in turn enables us to assess prospective technologies at minimal costs while eliciting rich end-user insights.

Chapter 4 Storyboard-based Pilot Study

This chapter describes the design, implementation and findings of our initial user study. We start out by elaborating the process and rationale behind the creation of our scenarios depicting future autonomous systems for the home. We then follow this up by describing the manner in which these scenarios were used to engage prospective users and drive focus groups. Finally, we present our findings in the form of the central thematic concerns that emerged as a result of the study.

4.1 Implementation

Although our scenarios were centred on hypothetical domestic situations, insofar as they envision and project autonomous systems that are not yet in existence, we felt it important that we not drift too far beyond what could reasonably be expected from such systems if our findings were to be of any practical use. This in itself was challenging, as the projected nature of domestic technologies is continuously shifting. The grand vision of the 'smart' home of the 1990s, for example, proactively managing every aspect of our lives, has been replaced in practice by the more mundane *connected* home (Harper, 2011), around which sensing and automation technologies now coalesce.

In approaching the development of scenarios for autonomous domestic systems, we found ourselves, like Rodden et al. (2013) before us, confronted by a technology that has "yet to be realised". However, this is not to say that we were starting with a

blank slate. As Reeves (2012) highlights in examining the origins of ubiquitous computing, technological projections or envisionments are grounded in a "milieu of existing and developing socio-technical infrastructures and innovations, drawing upon developments in diverse technologies". Recognising the problematic status of the future (Reeves et al., 2016), we therefore took great care to develop our scenarios by drawing on existing enabling technologies and extrapolating from current technological capabilities.

Much of the initial stage of our scenario development thus involved reflecting on existing research on autonomous systems and technologies that might impact the home, some of it based on research carried out by our lab (e.g. Fischer et al., 2016), some based on research being done by others (e.g. Yang and Newman, 2013), mixed with a degree of speculation on our part about where such developments and technologies could be heading.

Following this review, we also constrained the scope of our scenarios to 4 partially overlapping areas where autonomous systems are expected to make a substantial impact in the home:

- **Temperature regulation:** The use of wireless sensor networks for optimisation of indoor temperature is a well-documented area in ubiquitous computing (Stojkoska et al., 2014) and one that is well within the reach of current technological capabilities. Studies of emerging solutions have demonstrated the technical viability of regulating indoor temperature automatically by monitoring environmental variables, such as occupancy (Scott et al., 2011). Yet research has likewise found that such autonomous systems run the risk of rendering their operation unintelligible to users, which in turn can result in confusion and impede a sense of trust (Yang and Newman, 2012). Our scenarios thus featured an autonomous system that not only controls heating, but also ventilation and air conditioning. This helps us

explore the tension between autonomous regulation driven by sustainability and its impact on inhabitants.

- **Auto purchases:** Various auto-purchasing and reordering mechanisms represent another form of autonomous technology with a potentially strong impact on future smart homes. Semi-automatic solutions, such as the *Amazon Dash* button are already showing promising results (Lehmacher, 2017) and studies suggest fully AI controlled restocking solutions in the future (Market and Markets, 2019). If coupled with auto-delivery systems, these technologies could, for example, be used to prevent essential food items from running out by keeping the fridge stocked. On the other hand, studies have also been flagging that physical sensing of products in the home, computational learning and the prediction of need is problematic, making it difficult for fully autonomous systems to achieve desirable results on a consistent basis (Hyland et al., 2018). Our scenarios explore this tension between auto-reordering and consumer choice and control.

- **Energy regulation:** Many households are already starting to benefit from microgeneration technologies, such as solar panels, that not only produce energy in an environmentally friendly manner but also help keep the electricity bills low. The fluctuating availability of alternative energy sources, such as solar power, requires continuous monitoring and regulation of consumption. The delegation of such activities to autonomous systems represents a viable alternative that has already been explored in previous studies (Rodden et al., 2013; Costanza et al., 2014). Studies have at the same time also been suggesting that the contingencies associated with everyday life make a seamless operation of fully autonomous solutions problematic (Alper et al., 2016). Our scenarios explore the tension between autonomous energy regulation in the home and its impact on inhabitants.

- **Lifestyle monitoring:** Sensors are capable of monitoring multiple parameters of our lifestyles, ranging from biodata, such as heart rate or sleep patterns, to environmental variables such as humidity and air pollution. This in turn enables a range of solutions promoting a healthy lifestyle by giving the user visual or auditory reminders and clues that are intended to influence their behaviour (Peek et al., 2017). Such solutions have been described by previous research as representing a double-edged sword. Yang et al. (2014) for instance have found that such systems might succeed at least temporarily in adjusting user's patterns of behaviour but are prone to causing new problems, such as inadvertently providing poor or misleading recommendations. Our scenario explores this tension between autonomous regulation of one's lifestyle and individual autonomy.

In practice, this meant that our initial scenarios revolved around a range of autonomous solutions, including microgeneration of energy and the smart grid (e.g. Ecuity, 2013), automated scheduling of energy infrastructure use (e.g. Rodden et al., 2013), home automation of electrical appliances, including turning lights, heating and security systems on and off (e.g. Alper et al., 2016), or diagnostics of faulty domestic appliances (e.g. Seabra et al., 2016), all of which were framed by what we perceived as general concern in the literature with the 4 aforementioned areas, and the promise that the smart and indeed autonomous home of the future could deliver on this in manifold respects (e.g. Morollo, 2015).

These initial ideas were subsequently developed into textual outlines for future scenarios, each revolving around a fictional character named *Donald* who has just moved into a new smart home and tries to grasp the various autonomous systems inhabiting his domestic space. Given that a key goal of this thesis is to assess the acceptability of prospective autonomous systems, it appeared as a reasonable starting point of our enquiry to develop a basic understanding of the general level of acceptable automation. Consequently, following Yang and Newman's (2013) emphasis on finding an acceptable level of balance between system autonomy and

user engagement, we developed two outlines per scenario: one focusing on *full automation*, the other on *human in the loop*. The following examples illustrate the outcome of this initial stage of scenario design:

- **Example scenario 1 - full automation:** *Donald has just moved into a new smart home. He has never bothered about ecology and believes that all the talk on global warming and environmentalism is nothing but superstition. One morning when leaving his house to go to work, he's horrified to find that his car has disappeared overnight. He rushes back into the house frantically searching for his phone to call the police. Suddenly he hears a voice coming from the speakers of his smart home IoT system. The voice announces that Donald's car has been found to be producing an unacceptably high level of CO2 emissions. To prevent further damage to the environment, the autonomous system has sent the car to a local recycling company. As a replacement, the system has ordered a bike, which is waiting for Donald in his garage. He is told that not only is a bike far more environmentally friendly, but riding a bike will also help him address his increasingly high cholesterol levels.*

- **Example scenario 2 - human in the loop:** *One morning after waking up, Donald's smart phone starts reading out a voice message informing him that his car is producing a dangerously high level of CO2 emissions. He is recommended to have his car recycled and get a new one. The application offers him a list of nearby recycling companies along with contact details. Donald laughs out at the message and decides to ignore it. While preparing a breakfast, his smartphone delivers another message. This time it urges him to do something about his high cholesterol levels, stating that his food is not particularly healthy. While getting ready to leave his house, Donald lights a cigarette. This triggers another voice message and sends it echoing across the house. It explains that smoking is not only unhealthy, but also pollutes the environment. Angrily, Donald slams the doors shut and departs to his work. Donald endures multiple similar messages throughout the day. Nonetheless, in the long run*

he finds it difficult to ignore all of these recommendations. His lifestyle is gradually and subtly nudged towards a greater sustainability.

It was in the course of reviewing these scenarios through internal evaluations, when it became evident that the absurdities (Philosophy-index, 2019) they contained might be something we could leverage more productively if we adopted a contravision approach to render provocative utopian and dystopian visions instead of simply contrasting full automation versus human in the loop.

The scenarios were thus subsequently revised, while drawing on feedback from other members of our research lab - who were engaged in smart home research and IoT device construction - to ensure the feasibility of what we were proposing. The result of this revisioning process was that we ended up with a new set of scenarios covering the four areas outlined above. An elaboration of these scenarios can be found in table 2.

Scenario 1	Scenario 2
Scene 1: Temperature regulation	
Temperature in Donald's home is regulated automatically, without the need for any intervention. Everything is taken care off in the background so that Donald can spend his precious time on more important things.	The windows in Donald's smart home are being opened automatically to ventilate and regulate indoor temperature. The system operates fully autonomously without explicitly notifying its user of the variables that are being taken into account to determine whether to open the window or not. This in turn results in a situation where from the user's point of view the windows are being opened seemingly randomly.
Scene 2: Auto-purchases	
The smart home is also monitoring food items available in the fridge. Whenever a particular item is running out, the system reorders it to make sure that critical items are always available.	There are frequent mismatches between what the system determines is needed and what Donald actually wants. Moreover, since the system spends money on his behalf, Donald is slowly losing control over his expenses.
Scene 3: Energy regulation	

The domestic system automatically optimises energy consumption in Donald's home in order to keep electricity bills low while contributing to a sustainable society.	Electrical devices, such as lights, are being turned off without warning, sometimes even in the middle of being used by Donald. This leads to chaotic situations and even puts Donald in danger.
Scene 4: Lifestyle monitoring	
Through a network of sensors, the autonomous system monitors a range of variables pertaining to the lifestyle of its users. The information serves as a basis for recommendations and even fully proactive actions, such as regulation of air humidity.	The system's goal of enabling a healthier environment leads to catastrophic consequences as it proactively decides to send away Donald's old car for recycling and replace it with an electric one in a bid to reduce the household's carbon footprint.

Table 2 - Overview of the contra-vision scenarios

The utopian and dystopian characters of these scenarios allowed us to provoke and disturb common sense reasoning, which in turn created a *reality disjuncture* (Pollner, 1975), i.e. incongruous experiences of the world that motivate efforts at resolution and repair. It is in repairing the reality disjuncture that the taken for granted background expectancies that people use to make sense of and order familiar settings and scenes in everyday life are "called forth" (Garfinkel, 1967) and thus become available to our consideration. To better catalyse this process and to further increase the accessibility of these scenarios to prospective users, we translated the textual outlines into graphic storyboards (see Figure 1 for example illustrations).

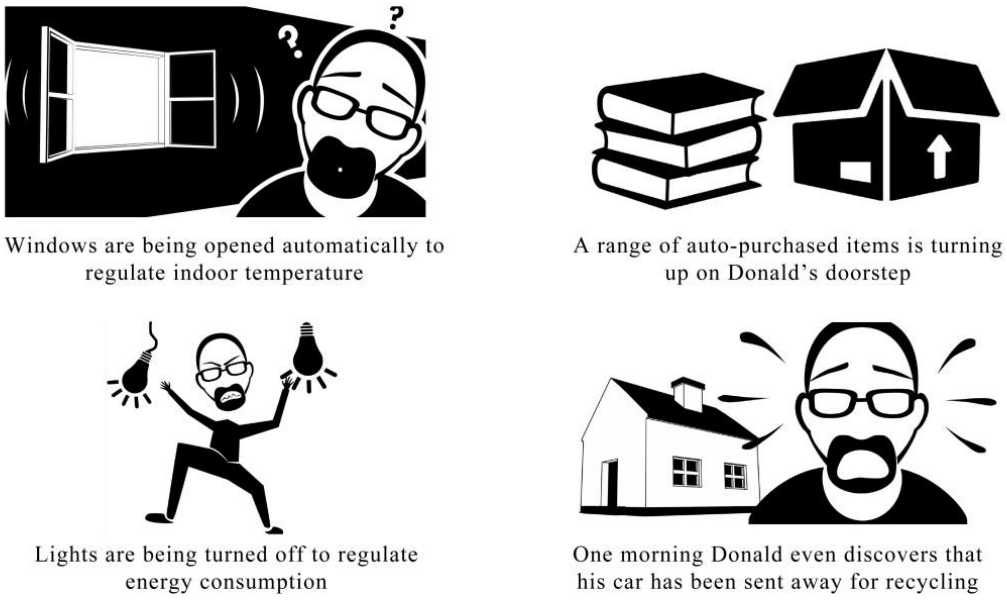


Figure 1 – A sample selection of illustrations featuring Donald and his smart home

4.2 Engagement with Users

We adopted a focus group workshop approach to assess our contra-vision scenarios with potential end-users of domestic autonomous systems. We chose this approach because group dynamics can help people to explore and clarify participants' views in ways that would be less easily achieved, if at all, in a one to one interview (Kitzinger, 1995). Below we detail recruitment of users and the organization of the focus group studies.

4.2.1 Participants

After obtaining ethics approval from the university's ethics committee, we advertised our study using a range of social media channels and sought to recruit adults of all ages with diverse backgrounds. Thirty-two people agreed to take part in 2 focus group workshops; 20 were staff members at our place of work (a university) who volunteered on the basis of a free lunch (pizza), and 12 were attendees at the 2016 Mozilla Festival, who volunteered out of interest in the topic. The participants were a mixed bag of designers, IT professionals, enthusiasts, PhD students, researchers and academics, male and female. As Twidale et al. (1994) put it, "… a possible objection to this would be that the small number of end-users might be unrepresentative. However, in one sense, no end-user is unrepresentative in that all end-users' viewpoints and requirements reflect a context in which the system may have to function".

The authors (Ibid.) also note that the "limited size" of a study does not stop "major problems" being "discovered and generalised", a matter we will return to in discussion.

4.2.2 Focus group sessions

Each of our focus group sessions ran for approximately 75 minutes and consisted of three key stages:

- **The utopian vision:** We initiated each session by presenting the utopian version of our scenarios. While constituting one of our contra-vision points, this stage also served as an overview of key technology concepts and of how these might ideally be implemented into the home of the future to improve everyday life. Participants were free to interrupt our presentation at any time.

- **The dystopian vision:** Participants were subsequently introduced to the dystopian vision, depicting a broken smart home infrastructure. The aim here was to open up utopian visions to scrutiny and encourage critical discussion and debate. Just as with the previous stage, participants were free to react or raise questions whenever they wanted.

- **Open discussion:** Finally, participants were also encouraged to reflect on and discuss what they saw. We initiated this stage by asking a broad question, such as "would you be willing to live in a smart home like the one depicted in our scenarios? If not, what would make it work for you?" Participants were given leeway to comment and respond to each other's thoughts and reflections, with occasional interruptions from us to ask additional probing questions around topics raised by participants.

The scenarios were presented via a slideshow and initially met with enthusiasm. The notion of augmenting domestic life with a range of technologies that operate

autonomously was generally well received, and the issues postulated by our dystopian scenario were frequently waved aside as temporary technical limitations that would eventually be sorted out. It wasn't until participants begun to relate the proposed technologies to their own lives and to the various social situations that might occur in their homes that more serious problems started to emerge.

Both focus group sessions were recorded on audio and subsequently transcribed. The transcriptions were used to examine the topical concerns that populated our participant's talk and their reasoning about autonomous systems in the home. Strictly speaking, we did not subject the transcripts to thematic analysis in order to "identify patterns within the data" (Braun and Clarke, 2006) but instead adopted an ethnomethodological orientation to identify the different orders of mundane reasoning exhibited in participants' talk and the endogenous topics elaborated by them (rather than by us as analysts) in the course of their conversations (Pollner, 1987). It is towards explicating the different orders of mundane reasoning and the topical concerns that characterise them that we turn below and elaborate through sections 4.3 to 4.6.

4.3 Coordination

The first major problem area to surface during our focus group sessions had to do - perhaps unsurprisingly - with concerns pertaining to the often questionable ability of autonomous systems to coordinate their actions with the existing routines of the domestic setting. It quickly became clear that our participants have strong background expectancies regarding what is and isn't acceptable in their homes and hold autonomous systems accountable to them. In order to be perceived as acceptable, an autonomous system would need to gear-in with these expectancies and operate in tune with the routines of the home. Unlike traditional computer applications, many autonomous systems are designed to work on the periphery of user's attention. Instead of simply responding to instructions obtained through

manual input from the user, these systems typically coordinate their actions with the user autonomously based on sensed data collected from their surroundings. Not all of these coordination processes are delegated to the background, however. Instead, more direct forms of communication are often required. A number of such mechanisms were brought up in the course of exploring our scenarios, including traditional visual dashboards, voice-based messages, virtually embodied agents and even humanoid robots. While the latter were generally seen as too futuristic, the former was seen as relevant to enabling coordination.

4.3.1 Reasoning about voice messages

Given the current turn towards voice-based interfaces (e.g. Amazon's Alexa and Google Home), as is evident from the (anonymised) conversational extracts below, much of the discussion about coordination inevitably revolved around the potential of automated voice messages, though not unproblematically.

> ***Researcher:*** *Another type of interface is the voice-based interface, which allows you to interact with a computer in a very natural way by speaking. It can basically wake you up or contact you whenever, it basically just shouts out recommendations or it could use something like motion sensors to deliver relevant messages whenever you're near a device.*
>
> ***Tom:*** *Personally I find voice a little intrusive. I don't even like phone calls. And the prospect of this thing going "Hey Tom, you need to eat carrots!" just while I'm talking or doing something else is not something I'm interested in.*
>
> ***Sarah:*** *Imagine if you have visitors in the house, so embarrassing!*
>
> ***Tom:*** *It's not what I would want.*
>
> ***Henry:*** *It depends if this is about all these things shouting at me or is this a single agent I'm interacting with by voice. I sit down in my living room,*

and I'm like "hey voice-thing tell me the things I need to know about things", and that's useful.

Fred: *Let's just paint a picture where you sit with your friends and you're eating some pizza, and it says "Hey, can I chat with you in another room?" And you go there and it says "On recommendations from the NHS [National Health Service] blah, blah, blah. You might want to do this. Do you want to do it?" You say yes or no. So it could be like a communication between you and your house.*

Mike: *I'm just visualising my fridge saying "You guys had an awful lot of beer tonight, maybe ..." you know?*

Claire: *What keeps coming up is that it's a relationship. Like the house is – the house becomes a housemaid. It might have your best interest at heart, but if it's socially awkward, or just does things wrong, brings in a bad mood, then it is a bad housemaid. So I think until you can have some form of natural feeling interface which isn't just going to be insulting you, or telling you to do stuff or muttering at you or treating you like a child then that's gonna be a big issue. Making a house that can do something is easy. Making a house work out whether it should is, I think, a lot more difficult.*

As our participants' talk makes clear, coordinative automatic behaviours (e.g. opening windows) are socially accountable to background expectancies regarding appropriateness. Context awareness is key. As Tom puts it, voice-based interventions would be inappropriate "while I'm talking" or "doing something else". Such interventions are seen as being not only potentially inappropriate but "embarrassing" if the underlying systems are not aware of the social circumstances at hand e.g. that "you have visitors in the house". This, of course, is not to say that voice-based coordination could not be useful, enabling autonomous systems to be seen as an "agent" or "housemaid" working in your "best interest", but that

usefulness needs to be balanced with a sensitivity to social context, as underscored by Claire.

4.3.2 Reasoning about timeliness

An issue that is intimately bound up with social context and coordinating autonomous systems with people is the timeliness of coordinative actions.

> ***Matthew:*** *If I just got home and I'm carrying 15 shopping bags, I don't want a thing to say "Matthew, have you bought carrots?" It's a real problem, a system that is going to interrupt me. A real person looking at me carrying 15 shopping bags would not say "Have you bought carrots?" I would be angry!*
>
> ***John:*** *Like here, if the system is telling me "Do you need a car?" that is not something you need to be told while you are in the car.*
>
> ***Henry:*** *It's fundamental that this is about asking for information at a time when it is, you know, not out of tune, and there is something you can do about it.*

Evidently, an autonomous system that seeks to coordinate with users will be seen as problematic if its actions lack sensible timing. The ability to "interrupt" at the right time thus becomes a vital component of acceptability. As Henry puts it, it is "fundamental" that system behaviour is "not out of tune" with human behaviour and occurs when there is opportunity to do something about it. Achieving coordination, in this sense, thus becomes a matter of orchestrating or synchronising system behaviours with the user's situation.

4.3.3 Reasoning about situated action

As the following extract makes perspicuous, there is more to synchronising system actions with the user's situation than being aware of what is happening "here and now".

> ***Researcher:*** *Would you accept a house that tries to change you and your habits?*
>
> ***Sarah:*** *When it comes to healthy lifestyle, like you said, you're not going to do it unless you want to. The house can try, but it should never force you. You should opt in to wanting to change. So you should have thing like "I want to have a more healthy lifestyle, but I don't know how to do that, can you help me house?"*
>
> ***Researcher:*** *So you're gravitating towards a recommendation system rather than a totalitarian smart home?*
>
> ***Sarah:*** *It definitely has to ask for information first, because if it starts recommending you stuff without knowing your situation – like for example, I could be eating a lot of junk food because I'm working 60 hours a week, and healthy food right now is just not making me happy; bad lifestyle to get me through my long work schedule. And although the smart thing here would probably be for the house to suggest that I don't work so hard, here what it's doing instead is making me more miserable.*

Context awareness is not just about being sensitive to social circumstance and timing interruptions to gear in with the current situation. That is to say, context awareness is not simply a momentary concern for people, but something that has a temporal horizon—e.g. "I'm working 60 hours a week". Coordination thus extends and turns upon gearing in with such things as a "long work schedule". This will require that

autonomous systems are furnished with "information" about the user's situation to enable the effective coordination of system behaviours.

4.4 Control

The perceived problems pertaining to coordination eventually ushered in the second major theme to surface in our focus group sessions – namely control. In spite of dealing with autonomous systems, i.e. solutions designed to carry out tasks on the user's behalf, it became increasingly evident as the conversation unfolded that manual control mechanisms enabling users to assume direct control and manually customise or even terminate inappropriate system actions would greatly contribute towards the overall acceptability of autonomous systems in a domestic context.

4.4.1 Reasoning about customisation

The need for autonomous systems to be furnished with information about the user's situation to enable effective coordination raises the issue of customisation as a distinct form of control.

> ***Sam:*** *I must say you have painted a really stupid autonomous system. You paint autonomous homes as something negative.*
>
> ***Dave:*** *It's intentionally provocative. The point is that it is extreme because if we talk about less extreme, the responses would be less interesting.*
>
> ***Sam:*** *But it's not realistic.*
>
> ***Dave:*** *It is realistic. If you live in social housing, they will put solar panels on your roof without asking you. They can turn off your electricity without much notice. You've got central heating which can shut off if there is not enough demand. There are all these things, OK. If you own your own home*

and are in full control of it, you're fine, but if you imagine a managed house, which a lot of people live in, then it's not impractical to suggest that some of this might happen. So it's not too far removed from what the reality is.

Sam: *I can't imagine anybody programming an autonomous system in a house that is this stupid.*

Dave: *Really? How long have you been doing computer science stuff?*

(Group laughter)

Sam: *I have faith in the fact that we can build working systems so there's a level of autonomy, a level of smartness, that's the thing. I think in this realm of setting things and turning lights off and stuff, I think it should not do that unless you want it.*

Researcher: *OK, so another problem that Donald ran into was the energy regulation issue. In the first scenario, the smart home started to shut down all the devices automatically. Whereas now it needs some input from the user. So a dashboard would probably look something like this. It would show a popup message encouraging you to do something and offering to take the appropriate actions on your behalf.*

Jack: *I kind of like the negotiations in the message interface.*

Claire: *But there's an awful lot of complexity in that …*

Sarah: *I'm sorry to interrupt you, because we're saying that all these things would be ongoing if not scheduled for a specific time. Why would it not be possible that he schedules exactly the time he wants all these events? You know, so that - of course there might be emergency so he'll have to change the schedule, but this does not happen every day I guess -- so he could schedule like "every day I want this to be done every time right before I go to bed or before I visit someone or do something else".*

The dispute between Sam and Dave concerning the realism of our scenarios ultimately helped surface the expectation that an autonomous system should not do things "unless you want it". This triggered discussion of dashboard interfaces as a means of users configuring "appropriate actions". The expectation here is that customisation should extend beyond system messages encouraging users to do things to enable users to "schedule events" themselves. In speaking about scheduling events "right before I go to bed" or "before I visit someone" or "do something else", there is a strong sense in which Sarah speaks of customisation as a means of gearing the behaviours of autonomous systems in with the rhythms and routines that are implicated in the user's everyday life.

4.4.2 Reasoning about direct control

On a final point of note concerning control, our participants also expected that they would be able to exercise direct control over the behaviour of autonomous systems. This was surfaced at various points throughout their discussions but perhaps none so pointedly as here:

> ***Paul:*** *You know there is a lot of ideology in the structure of the problem itself. A lot of things are biased ideologically to start with. I see it as a move towards a consumerist attitude that I don't see very likely to work. They can sell you a car that won't start if you drank too much. You might think it's a very good idea until the day that you have somebody who is bleeding to death and you drank a couple of beers. You want to run to the hospital, but the car doesn't start. The most important thing on a machine is the off button. Of course, you could use it if you drink a couple of pints too much at the pub, maybe, if you're irresponsible. But it's your responsibility, so you must have an off switch.*
>
> ***John:*** *How do you turn these systems off basically? I mean I appreciate your point, I do think technology should have an off switch, I'm just wondering, what form does this off switch take? If your magical smart*

> *future car has an off switch, you don't want that off switch to occur when you're driving 70 miles an hour down the highway.*
>
> ***Claire:*** *It's also about all the smart features, I mean how dumb do you make it. If you had a smart on and off, well now I can exceed 70 miles per hour and I can drive drunk, but I can't use cruise control or I can't use automatic gearbox, because I made it into a totally dumb car. So I think it's definitely a sliding scale of how smart do I want this car to be.*

The ability to exercise direct control and shut down an autonomous system is evidently an expectation that people hold, but it is not at all a simplistic on/off expectation. While some circumstances may warrant an "off switch" (e.g. overriding an autonomous system when "you have somebody who is bleeding to death"), others that may put users at significant risk (such as switching off while "driving 70 miles an hour down the highway") do not. The issue of control is not only one of balancing risk with human autonomy, however. Bound up with this is the expectation of granular control, of a "sliding scale" of smartness that can be determined by human beings to enable a better fit between autonomous systems and the particular human circumstances in which they are embedded.

4.5 Computational Accountability

As we elaborated in section 2.3.3, the term "computational" accountability originally was introduced into HCI to describe the *legibility* or *intelligibility* of computer systems to end-users (Dourish and Button, 1998). The importance of designing systems that are accountable to their users has long been recognised in this field (Friedman, 1997; Bellotti and Edwards, 2001).

As can be seen in conversational extracts from the focus group sessions below, issues surrounding computational accountability of autonomous systems are not only

concerned with accounting for the system's behaviour, but also with what is involved in providing an account, and particularly with making the motives and reasons for the actions of autonomous systems transparent along with the kind of transparency required to provide a sufficient account.

4.5.1 Reasoning about accountability

In considering the idea that autonomous systems could take proactive actions to improve the ecology and sustainability of the household, our participants exhibited a concern in their talk with the legibility of the smart home's behaviour to Donald.

> ***Jack:*** *It depends how the whole system is being framed for Donald. So if he wants to get a really good deal, if he's been told his house would do things to minimise the amount of energy he uses, he might think "fine", because he wants to keep his bills down. But if he is told "your house will do things because it is trying to be sustainable and environmentally friendly", he might feel like "well, I don't care; I'd rather it do what I want it to do for me."*
>
> ***Jill:*** *I wouldn't like to live in a house where I could not account for things that were happening.*

This simple interchange between Jack and Jill reflects broad concern amongst our participants with the accountability of autonomous system behaviours. While seeing the potential benefits of autonomous systems, no one wanted to live in an environment where what happened there—the automatic modification of heating levels, the turning on and off of lights, the opening and closing of doors and windows, etc.—could not, as Jill puts it, be accounted for. It seems clear then that a critical background expectancy impacting the uptake of autonomous systems is that their behaviours will be accountable to end-users. But what exactly does accountability consist of or require?

4.5.2 Reasoning about motives

A key background expectancy closely bound up with computational accountability became perspicuous when participants were considering the scenario in which the autonomous system in Donald's home decided to replace his car with a newer one. It did this to reduce Donald's carbon footprint without asking him for permission or giving any other prior notification.

> ***Phil:*** *But who is it that's telling you that you need to change your car? It could be some kind of a company that you pay to service it and they recommend you to change the car. Is their motive that you're paying for their service? Or is their motive like "let's sell cars based on some dodgy grey statistics that we made up."*
>
> ***Alan:*** *People get usually very emotionally attached to their cars.*
>
> ***Sally:*** *You might imagine you're in one of those schemes where you get the latest thing, and the latest thing is forced on you, but you have sort of a nostalgic attachment to the previous thing. "This is obsolete now, we'll give you the newest one automatically." You would not like that kind of autonomy.*

As quoted, Phil makes visible, "motive" matters in making the behaviour of autonomous systems accountable. Of particular relevance and concern is the "telling" of specific motives that drive specific autonomous behaviours, and on whose behalf they are acted upon. The "telling" of just what, just why and just who is important and speaks to the active provision of an account as a preface to autonomous behaviour rather than a bolt-on explanation delivered (like Donald's car) after the fact, for as Sally puts it, "You would not like that kind of autonomy". Accountability thus becomes an important part of the conduct of autonomous system behaviours, rather than a narrative that is subsequently appended to action

and speaks more generally of the need to build transparency into the behaviours of autonomous systems.

4.5.3 Reasoning about transparency

The background expectation that transparency should be built into the behaviours of autonomous systems became apparent when participants were confronted with the scenario of temperature regulation in Donald's house, which the system does (in part) by automatically opening and closing windows.

> ***David:*** *If you live in a climate where it gets really hot and you need to get cooler air in the evening or morning – if these windows open when I would open them – then great, why not. But why does the slide say the windows are being opened seemingly randomly? That means the system is faulty, the system is not opening them when he would like to open them.*
>
> ***Sarah:*** *I guess it kind of works because the windows are opening when it's too hot and closing when it is too cold.*
>
> ***David:*** *Who defines when it is too hot? For him? Donald at some point wants the windows open because for him it's sometimes too hot or too cold. So if it is seemingly random to him, then the system is not behaving properly.*

Like Phil, in raising the matter of motive, David clearly begs the question, on whose behalf is an autonomous system acting? With that goes the expectation that their actions will serve end-user need. But whether acting on Donald or some other unidentified actor's part, David also makes it perspicuous that should an autonomous system's actions appear "seemingly random", it would not be seen to be "behaving properly", where "properly" clearly means its actions are intelligible

rather than discernibly "random" and something that has to be guessed at as Sarah has to do.

So not only is it expected that an autonomous system operates in tune with the preferences of its users, but it is also expected that the grounds upon which it operates on any occasion are visible to end-users. It might otherwise be said that the telling or provision of an account extends to incorporate both the reasons for action as well as the motive, which is to recognise that motive alone does not adequately account for action; motive may account for what occasions action (e.g. temperature regulation), whereas reason also accounts for particular actions done with respect to motive (e.g. opening windows to cool the environment). Background expectancies regarding transparency thus extend the accountability requirement to include the in vivo accomplishment of autonomous behaviour, i.e., the doing of autonomous behaviour and provision of reasons that account for what is being done.

4.5.4 Reasoning about the nature of transparency

A further background expectancy became apparent in participants' reflection on the previous temperature regulation scenario, which concerns the kind of transparency considered necessary to understanding the behaviour of autonomous systems.

> ***Researcher:*** *Would you like to live in a house like this?*
>
> ***Megan:*** *Not with Donald.*
>
> *(General laughter)*
>
> ***Harry:*** *The question is whether you would be willing to live in a house where you cannot fathom the things going on around you. Where the house is behaving in ways you cannot understand.*
>
> ***Megan:*** *But that happens all the time. My house is too cold and I don't know why.*

> ***John:*** *Well, we all live in houses we don't understand, unless you know what's going on in a gas boiler.*
>
> ***Harry:*** *But you do understand the thermostat, because it is really simple.*

It would be easy to think that in expecting the reasons for an autonomous system's behaviour to be made transparent, end-users are also expecting designers to open up a black box of complexity. The above extract disabuses us of this notion. Just as most of us do not understand "what's going on in a gas boiler" and would not understand even if the underlying reasons were carefully spelt out to us, then so it will be with autonomous systems: end-users do not expect to understand the detailed reasoning implicated in their inner workings, and they do not expect to have that level of insight because they have no practical need of it.

A rather more "simple" level transparency is instead required for the practical purposes of end-users, of the kind provided by a thermostat, for example. Other examples discussed by our participants included articulating the "agenda" of an autonomous system and the underlying "business model", instead of the intricate technical processes going on under the hood, though as noted above, such matters will need to be manifested in the conduct of autonomous system behaviours and not merely as after the fact appendages.

4.6 Social Accountability

While the discussion amongst our participants was initially oriented to technical considerations of autonomous system's behaviours and their intelligibility, it became increasingly apparent as the conversation progressed that they were also holding the autonomous systems depicted in our scenarios accountable to the broader social context within which they were prospectively situated. To accommodate for this larger dimension of accountability, we here choose to introduce a fourth problem

area in the form of "social accountability". As the extracts below make evident, social accountability extends beyond expectations concerning the legibility of autonomous behaviour to expectations concerning the appropriateness of autonomous behaviour within a social setting.

4.6.1 Reasoning about acceptability

Concern with social accountability as distinct to computational accountability became evident in participant's consideration of the automatic temperature regulation scenario:

> ***Rachel:*** *You wouldn't like the windows opening when you're not at home, regardless how hot it is.*
>
> ***Megan:*** *Depends what kind of window it is; not with that kind of window (points to the large open window in the presentation slide).*
>
> ***Ben:*** *If you live close by the road, with cars going by while you try to listen to something …*
>
> ***Adam:*** *There are also these big bees outdoors ...*

The extract makes it perspicuous that understanding why an autonomous system is behaving in the way it is (e.g. why it's opening windows) does not automatically mean that end-users will find this behaviour acceptable. On the contrary, people evidently entertain background expectancies regarding the appropriateness of a system's behaviour. Opening windows "when you're not at home", or "while you try to listen to something" or with those "big bees outdoors" all point to the need for autonomous systems to not just be intelligible in their own right but also in relation to the specific social context within which they operate. Their behaviours, in other words, need to be intelligibly appropriate given the specific social conditions in which their actions are embedded "here and now" (no one is at home, people are busy, there are hazards outside, etc.).

4.6.2 Reasoning about agency and entitlement

Background expectancies bound up with the appropriateness of system behaviour also turn upon considerations of agency and entitlement, as became perspicuous when participants were confronted with the notion that people could order products simply by talking to the autonomous home through a voice interface:

> ***Sam:*** *I can imagine with the voice thing, if it would actually start buying stuff… that would be really bad!*
>
> ***Sue:*** *If you've got your five-year old kids in the house alone they might go like "chocolate and popcorn!" and the voice replies "OK!" and off we go!*
>
> ***Jill:*** *It should be able to know who is actually going to end up paying for it, because there are those examples of kids that run up hundreds of pounds on app purchases because their parents left out an iPad.*

As the extract makes visible, it is not only the social conditions in which autonomous systems are embedded that impact the appropriateness of their behaviour, who is implicated in those behaviours also counts. Clearly, not everyone has the same rights and privileges in the home and our participants held autonomous systems accountable to mundane expectations regarding how everyday life in the home is socially organised. Thus, and for example, an autonomous system should not reorder "chocolate and popcorn" just because it has run out, and the kids want more. Rather, any such action should be accountable to the party who is "actually going to end up paying for it". An autonomous system embedded in the fabric of domestic life must effectively become a social agent whose actions comply with the mundane expectations that order domestic life. Put simply, the system needs to act in a manner consistent with that which would be expected of a responsible human agent.

4.6.3 Reasoning about trustworthiness

A key social accountability expectancy concerns the trustworthiness of autonomous systems and whose interests they serve, as become visible when participants were discussing auto-replenishing systems capable of reordering worn-out household items:

> ***Researcher:*** *So the next problem that Donald encountered was the problem where random products were being delivered to his doorstep automatically. He didn't order anything and everything was paid for without his consent.*
>
> ***Jill:*** *So that's what I wanted to mention that we're actually assuming that you trust your house or you trust your personal assistant. We assume that they do the right thing.*
>
> ***John:*** *Say your washing machine gets broken and your house has purchased this specific spare part from one specific supplier. But that particular brand is maybe not in your best interest, because you know, it spoils quicker. It's not in your best interest, but it's in the best interest of the guys who made the washing machine, or some other company out there.*

As John makes visible, even if the actions of the autonomous home are intelligible and the end-user knows why it has done what it has done (e.g. bought a replacement part because the washer was broken), it is not evident that doing so is in the end-user's interest. On the contrary, it might "be in the best interests of the guys who made the washing machine, or some other company out there". There were, as we have already seen, multiple instances where our participants questioned whether the behaviour of an autonomous system could really be trusted and its actions carried out on the end-user's behalf. Our participants expected that external interests would exert considerable influence on what goes on in the autonomous home and found this deeply problematic. It is critical then that autonomous systems not only

articulate motives and reasons but that their actions are clearly accountable to end-user interests. But how?

4.6.4 Reasoning about risks

Clearly, autonomous systems raise the risk of widespread exploitation from an end-user viewpoint, championing the interests of external parties at the expense of those who live in the autonomous home. Another closely related key area of risk concerns the potential for widespread data harvesting and the risk this poses to privacy.

> ***Simon:*** *The thing about AI technology is that all of this stuff is data and it's stored somewhere and it knows all these things about you and potentially that makes you a target for an attack.*
>
> ***Researcher:*** *You never know who else is watching?*
>
> ***John:*** *Or who else has a relationship with your house.*
>
> ***Wilma:*** *If we start encouraging smart homes to have their own server, so all your data goes back to your cloud and that sends a limited amount of data to the people that need it, once that becomes a viable option, stuff like the legal requirements around how you store people's data, and how you make this clear to them, is going to make it so that if a company is asking for detailed data from your life, then that might be something why you choose not to do business with them.*
>
> ***Martin:*** *You need an on-off switch and an awareness model so that you can close your data from going out and you're aware of what data is going where. Everything is happening, I mean the Internet of Things, things that are in your home sending data of yours without you realising it. Or providers giving you a service without actually explaining their business model is another problem.*

In considering the implications of our scenarios, our participants' discussion reflected widespread social concern about the impact of intelligent technologies on everyday life and "who else has a relationship with your house" enabled through current cloud-based autonomous infrastructure. Our participants expect that "what data is going where" be made accountable to them by the autonomous home. They posit various mechanisms enabling this, including local data storage mechanisms that make end-users aware of data flow and which allow them to "limit" the flow of data and "close" it off, as well as "legal" mechanisms that enable end-user's choice. Such mechanisms might in turn make the behaviours of autonomous systems in the home accountable to end-user's interests.

4.7 Key Takeaways

Our scenarios have provided means, motive and opportunity to (gently) prod and provoke the expectations at work in domestic life that are ordinarily taken for granted and left unsaid. In doing so, the participants in our study have elaborated distinct challenges for future technological visions. These include the expectations that the behaviour of autonomous systems gears in with the behaviour of end-users, that people can intervene in their operations, that behaviours of autonomous systems will be accountable and that their behaviours are responsive to the social context. More specifically, we can say that our participants expect that coordination and control are enabled and that autonomous systems will be computationally as well as socially accountable. Below we provide a brief summary of the key takeaways for each of these themes.

Coordination speaks to the need for systems to gear-in with the routines of the home without stirring up too much friction. Our participants made it evident that the means by which a domestic autonomous system coordinates itself with its users can themselves be a source of such friction. It is therefore imperative that coordination mechanisms, such as a system reaching out to its user through a voice message, are

both contextually appropriate as well as timely relative to the user's situation. We were able to establish that such a level of synergy between autonomous systems and their users can only exist if the system correctly assesses not just the user's momentary state (e.g. being busy carrying multiple shopping bags), but also the big picture situation (e.g. working long hours and does not have time to eat healthily).

Our focus group sessions made it clear that an autonomous system could scarcely be trusted to correctly assess all the complexities implicit in a domestic life on a consistent basis. Consequently, the need for a degree of user control emerged as another major problem area. By unpacking the expectancies held by our participants, we were able to show that user control should not be reduced into a binary mechanism, e.g. an on/off button. Rather a user ought to be given a more fine grained choice over the individual features of a system. Control over an autonomous system, in other words, must be treated as a spectrum, rather than a dichotomy.

Being able to account for things happening in one's own home represents another major expectancy raised by our participants. Although transparency of domestic autonomous systems is crucial to meet this expectancy, our findings suggest that this does not mean autonomous systems should have their inner workings spelled out and made fully visible to the user. Rather, a "simple" level of transparency is desirable, providing the reasons and motives behind system actions, including on whose behalf they are taken. Our participant responses indicated that such accounts of a system's actions ought to be provided as a preface to the given action, rather than as a post-hoc appendage.

While the discussions surrounding computational accountability were initially concerned with the technical aspects of a system, the reasoning of our participants made it increasingly apparent that autonomous systems need to be held socially accountable as well. A system can provide accounts explaining its actions all it wants, but if these actions are not seen as intelligibly appropriate relatively to the context-specific social expectancies that surround it, the system behaviour will still come across as unintelligible to its users. It is apparent, in other words, that

autonomous actions must not only be intelligible in their own right, but also visibly appropriate to the end-user's interest. Respecting the roles and hierarchies of household members and making personal data flows visibly appropriate, were brought up as two examples of such socially accountable system behaviour.

The first three of the thematic concerns surfaced through our user study are generally in agreement with the problem areas laid out in our literature review chapter. The fourth one, which we choose to label *social accountability*, represents a novel addition to these. As evident from above, these themes are still superficial and more detail is needed before any conclusions regarding their impact on the acceptability of domestic autonomous systems can be made. Consequently, in chapter 5 we will focus on elaborating these themes further.

Chapter 5 Further Exploration through Video Scenarios

The pilot study largely confirmed the gravity of the key problem areas identified in our literature review, namely "coordination", "control" and "accountability". In addition, the expectancies voiced by our participants prompted us to introduce a fourth problem area in the form of "social accountability". At this stage, these four categories of concerns brought forth by our participants still constitute loosely defined - and to a certain degree overlapping - themes. More data is needed before we can start teasing out any conclusions and articulate the key design challenges at play. The study about to be presented in this chapter is thus designed to expand on our previous study by going deeper and probe prospective users for more detail concerning the aforementioned themes. We will start out by describing the development process and rationale behind a new set of scenarios. This will then be followed up by an outline of the structure of our second user study, explaining the manner in which our scenarios were used to engage our participants. Finally, we will present an analysis of the collected data.

5.1 Implementation

Our initial set of scenarios were employed as a tool for taking a relatively high-level look at various hypothetical applications of emerging autonomous technologies (e.g. automatic shopping or health tracking) and the various ways in which this affected

the overall acceptability of hypothetical domestic autonomous systems. This in turn resulted in the surfacing of 4 broad thematic areas.

The objective of our second study was to go beyond merely surfacing the existence of particular problems, to instead elaborate the overall role autonomous systems ought to play in our lives in order to remain acceptable. A frequently touted selling point of autonomous systems is the notion that they will carry out some of the tedious tasks typically implicit in daily life on our behalf and in turn allow us to live more conveniently (Bellotti et al., 2015). Our findings so far, however, seem to provide us with a prima facie indication that convenience alone cannot be seen as a valid design goal. Instead, it appears that a range of other values and preferences might have to be satisfied in order to render an autonomous system socially acceptable. Which exactly are these values and preferences? When is it acceptable – or even desirable – for a system to carry out tasks on our behalf and when do we prefer to do things on our own?

In search of an answer, we once again found ourselves iteratively sketching out and gradually converging on a pair of contravision scenarios pitching a number of daily life situations set in a future teeming with autonomous systems. At the heart of these scenarios, we ended up focusing on mundane activities which every prospective user could reasonably relate to. The particular use-case areas postulated by our new scenarios thus did not markedly deviate from those already identified and outlined in section 4.1 of this thesis. A strong focus was for example placed on food consumption and food management in the home.

Once again, we developed our scenarios by drawing on existing enabling technologies, extrapolations from current technological capabilities and specifications, policies and existing socio-technical systems. The scenarios were thus grounded in relatively familiar solutions such as mobile apps, grocery delivery services, “click and collect'” consumption, recipe recommendations, and barcode scanning.

Following the contravision framework, the two scenarios were again intentionally filled with contrasting visions of potential future autonomous system uses. The first of these new scenarios (henceforth, we shall refer to it as "scenario 3" due to it being our third scenario overall) focused on conveying hypothetical autonomous systems prioritizing a sustainable agenda and the health of its users, with trade-offs in the form of less convenience. For instance, to reduce food miles, grocery orders were not delivered straight to the user's home, but rather to a central pickup point where they had to be manually collected. Much of the services provided to the user were facilitated by a fictional company which we named "Greenfood".

On the other hand, the second of our new scenarios (henceforth "scenario 4") depicted an autonomous system which prioritizes the convenience of its users above anything else. In practice this meant that the user barely had to do anything, with food management and cooking all being taken care of on his behalf by the autonomous system. The conveyed downside to such a solution was its high cost, not just in financial terms, but also in terms of environmental impact. The services depicted in this scenario were facilitated by a fictional company which we named "Zanamon".

Both of these scenario versions were purposefully exaggerated, with multiple points taken to an extreme in order to provoke reaction. A summary of both scenarios is provided in table 2 below.

Scenario 3	Scenario 4
Scene 1: Ordering food	
Paula is returning home from her work using public transportation. A mobile app presents her with a list of recommended meals that she might have for dinner. The recommendations are based on the ingredients that are in season and locally available.	Paul is returning home from his work in a self-driving car. While still in the car, Paul is contacted by his personal robotic assistant who enquires about his choice of meal for tonight. Paul chooses an expensive fish imported from across the globe.
Scene 2: Managing food waste	

At home, Paula is notified by her smart home system that some of the groceries in her fridge are approaching their use-by date. She is encouraged to either consume them or donate them to a local food bank. She chooses the latter.	Paul does not have to care about food waste. Any food item that has gone out of date is automatically disposed of by his personal assistant.
Scene 3: Food delivery	
Paula is notified that her food order has been delivered to a central collection point. She cycles to the delivery point, leaves her leftover food in a food bank, grabs her new food ingredients and returns back home.	The food ingredients that Paul ordered are delivered by a drone straight to his doorstep. From there, everything is taken care by his robotic assistant.
Scene 4: Monitoring food dates	
Once back at home, Paula scans her newly acquired food items into her smart home system. This is to keep the system up to date on the available food and its use-by dates.	Every food item in Paul's fridge is being monitored by cameras and sensors. Whenever a particular food is running out, Paul is notified by an automated message and the food gets automatically restocked.
Scene 5: Cooking	
Once all the food has been scanned into the system, Paula proceeds by cooking and enjoying the dish she was looking forward to, safe in the knowledge that she is cutting down her food waste and sharing with the local neighbourhood.	Paul is relaxing on his sofa while watching some TV. Meanwhile, his home robot cooks and serves him the dinner he ordered earlier that day. The scenario ends with Paul's robotic assistant appearing on this TV screen in order to inform him that his bills for the service are due to be paid.

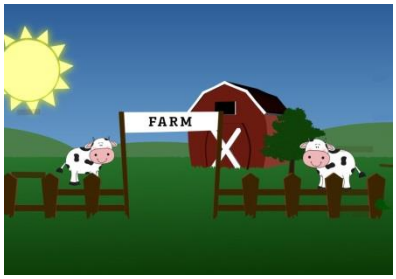
Table 3 - A summary of the scenarios produced for our second user study

Unlike the graphic storyboards that served as a focal point for our engagement with users during the first user study, this time we developed our scenarios in the form of a set of short animated videos[2]. As Young and Greenlee (1992) note, the use of video scenarios makes it possible to illustrate a design vision of prospective systems more effectively than written documents or static sketches. Moreover, Muller (2009) argues that videos can help more accurately simulate the use of not-yet-developed tools and technologies to explore new possibilities and that they thus help facilitate a

[2] Scenario 3 is available online under the following link: https://vimeo.com/238701035
Scenario 4 is available online under the following link: https://vimeo.com/241688455

"fuller understanding by focus group members, leading to a more informed discussion".

We chose to style these scenarios as 2D cartoon animations. This was largely for practical reasons, with cartoonish elements being easier and quicker to produce than more realistic visualizations, whilst also allowing us to relatively easily depict the operation of systems that are not yet in existence (which would have been more problematic had we instead relied on real-world recordings). A sample scene selection illustrating the overall visual style of our video scenarios is provided in table 4 below, along with the key contravision points featured in the two contrasting scenarios.

Scenario 3	Scenario 4	Contravision point
		Greenfood VS Zanamon
		Local VS Global
		Public transport VS Self-driving car

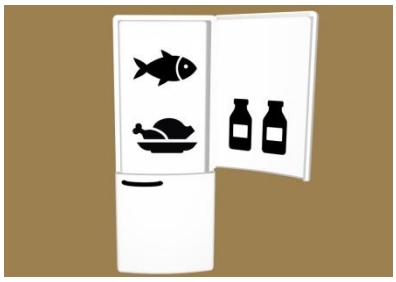
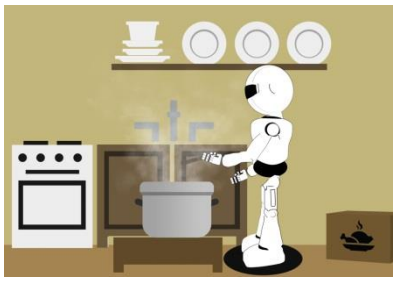
		Sustainability VS Convenience
		Manual pick-up VS Auto-delivery
		Manual scanning VS Auto-restocking
		Manual cooking VS Robot cooking

Table 4 – Sample set of screenshots illustrating the aesthetics of our scenarios along with the respective contravision points

5.2 Engagement with Users

We once again adopted a focus group workshop approach to assess our scenarios with members of the public. Below we provide an overview of the participant recruitment procedure and the organization of the focus group sessions.

5.2.1 Participants

After obtaining ethics approval from the university's ethics committee, we hired an agency to recruit adults with diverse backgrounds from all ages. 28 members of the public eventually took part in our study (14 female). Eleven participants were in the 20s, three in their 30s, seven in their 40s, three in their 50s, and four in their 60s. The median age was 39. Six had children aged 11 years of age or younger, three had a teen-aged child, and three had adult children. Our participants had a varied range of occupations, only four of them were students. An overview can be found in Table 4 below.

The focus groups were held in a lab setting. Each session lasted approximately 2 hours. Upon finishing, each participant was debriefed and provided an incentive in the form of £40 cash.

Participant	Group	Gender	Age	Occupation	Marital status	Children's ages
Fiona	1A	Female	25	Homemaker	Widowed	8 and 4
Francesca	1A	Female	24	Factory Worker	Cohabiting	2
Fabia	1A	Female	57	Part time Bar Steward	Married	None
Malcolm	1A	Male	47	Careers adviser	Single	7 and 2
Martin	1A	Male	64	Medical Advisor	Married	None
Gilbert	1B	Male	67	Project Manager	Married	None
Peter	1B	Male	47	Tutor/Community Worker	Single	None
Simone	1B	Female	19	Part time student	Single	None
Zikmund	1B	Male	24	Personal Trainer	Cohabiting	None
Amy	2A	Female	55	Private Therapist	Married	None
Jim	2A	Male	32	Senior Administrator	Cohabiting	6 months
Petra	2A	Female	23	Customer Service	Single	None
Salman	2A	Male	46	Reactive Engineer	Married	24
Bertha	2B	Female	37	Part time student	Single	None
Jane	2B	Female	26	Student	Single	4 and 2
Keith	2B	Male	49	Trainer	Single	None
Lana	2B	Female	36	Manager	Married	11 and 7

Max	2B	Male	60	Audit Manager	Married	27
Amanda	3A	Female	24	Account Management	Single	None
Carl	3A	Male	49	Student	Married	Empty nester
Jack	3A	Male	65	Civil Servant	Married	Empty nester
Melwin	3A	Male	47	Delivery Driver	Married	15
Pauline	3A	Female	55	Community Food Worker	Married	19
Amanda	3B	Female	25	Recruitment	Single	None
James	3B	Male	26	CAD Technician	Single	None
Mary	3B	Female	19	Student	Single	None
Robert	3B	Male	24	Factory Packer	Cohabiting	None
Samantha	3B	Female	41	Solicitor	Married	13

Table 5 – Study 2 participant overview

5.2.2 Focus group sessions

We adopted a semi-structured focus group approach to show and discuss the two contrasting scenarios. We chose this approach to help our participants explore and clarify their views in ways that would be less easily accessible in a one to one interview. Pope et al. (2000) argues that this is particularly the case when researchers wish to encourage their participants to explore the issues of importance to them, in their own vocabulary, generating their own questions and pursuing their own priorities. Adams and Cox (2008) argue that focus groups "can provide a flexible and participatory method that contextualises users' perceptions and experiences". Whereas focus groups hold the risk of a single individual taking over and dominating the discussion, we attempted to mitigate this through moderation, making sure each participant was able to express their opinion.

The study was carried out over three two-hour long sessions, each featuring two parallel groups of 4-5 individuals. After providing informed consent on participating in the study, members of the first focus group in every session were initially introduced to the sustainability scenario, which was then followed up by the

convenience scenario. To balance any bias, the parallel focus group in every session was introduced to our scenarios in a reversed order.

The animated scenarios were shown separately with a pause in between during which participants were invited to give feedback and discuss what they just saw. An approximately 30 minutes long semi-structured group discussion followed each screening, allowing participants to comment on the concepts introduced in the scenarios.

After watching and discussing both of our scenarios, every focus group was asked to hold a final discussion comparing the concepts in the two scenarios with the two parallel focus groups presenting and debating their respective views with each other. This was an extra precaution intended to help cancel out any bias that might have arisen as a result from the order in which the scenarios were presented or any other peer-pressure that might have developed in any of the individual focus groups.

All sessions were audio and video recorded. The group discussions were subsequently transcribed in full and served as a basis for our analysis. The data was then coded following the same approach we employed in our first study, namely using an ethnomethodological qualitative analytical approach.

During the coding process we paid a particular attention to statements in which participants made reference to their own lives. This allowed us to gain better understanding of the personal values and preferences invoked during the focus group discussions. We then iterated over the coded dataset, grouping the data more broadly around the core problem areas converged upon in Study 1. The participant statements were often drawing on multiple problem areas at the same time as these are – as previously noted - overlapping and not mutually exclusive. The headlines that we use below to group and present our findings are thus merely labels created through our own analysis, for presentational purposes, rather than the participants' own categories (Crabtree et al., 2018). Our ethnomethodological orientation thus commits us to exhibit members' mundane reasoning, rather than our own.

5.3 Coordination

Coordination once again surfaced as a key area of concern. Our scenarios postulated a wide spectrum of different autonomous systems and a variety of ways by which these could gear-in with the intricate domestic routines. The reasoning exhibited by our focus group participants that we unpack below has helped uncover major areas of friction.

5.3.1 Reasoning about convenience

Many of the autonomous solutions featured in our scenarios (particularly in scenario 4) were depicted as carrying out tedious tasks on their user's behalf in order to help them live more convenient lives. Our participant's responses made it explicit, however, that such a degree of convenience is a double-edged sword. Just because a system is technically capable of substituting for its human user and carrying out tasks on their behalf, does not automatically mean that it should. A range of aspects of the domestic routines turned out to be highly sensitive, suggesting that delegation of work to autonomous systems ought to be limited.

> ***Max:*** *These things give us more time to do things.*
>
> ***Lana:*** *I think people… I like to do things that keep me in touch with… that's my routine, that's my life, that's what my children see me doing. That system is taking the whole element of it away. And the older I get, the less... I'm gonna rely more on it I feel, because it's gonna do it for me. You're gonna withdraw yourself away from people.*
>
> ***Max:*** *When you mentioned that... my mum has dementia, and she used to be on her own at home, before her proper diagnosis and she's looked after herself at all, and I thought she did brilliant.*

> ***Lana:*** *I suppose there are no put pros and cons. But my mind-set is that with my children and my family at the moment it would be intrusive to my family-life.*
>
> ***Jane:*** *I feel like I'd have more time with my children. I don't get home till very late and then I'm cooking meals. And then we're not sitting down at the table till like 7. So having this, I can spend quality time with my children before putting them to bed. So I personally would benefit from this.*

Much of the initial discussions in the focus groups revolved around debating the nature and benefits of convenience. While some participants suggested that convenience goes hand in hand with freedom of choice and the ability to choose exactly what you want, others adopted a different stance, arguing that convenience is about not being forced to spend time on undesirable or boring tasks. A smart fridge could for example monitor stock and quality of food items on our behalf, thus freeing up more time for us to spend with our families or on other pleasant activities.

Conversely, the more our participants started to relate the hypothetical autonomous systems to their own life, the more worried they seemed that autonomous technologies carrying out tasks on our behalf might also affect our lives in a detrimental manner. Some participants suggested that autonomous systems would effectively make some of our skills obsolete and contribute towards their erosion. This in turn, they argued, would render us overly reliant on our technology and make us increasingly passive and vulnerable to any malfunctioning.

Moreover, a system carrying out tasks on our behalf would make us increasingly socially isolated. As illustrated by Lana in the vignette above, an over-reliance on autonomous technology might diminish the need for inter-human communication and in turn divest people of their social skills and erode a sense of family cohesion.

Some of the autonomous system behaviour that our scenarios suggested could bring us greater convenience turned out to be repulsive in the eyes of our participants. For example, the embodied domestic robot was strongly disliked in our convenience scenario; most participants frequently expressed their discomfort at the idea of

sharing their homes with an anthropomorphised robot and having it, for instance, cook their meals for them. Doing the cooking manually or hiring an actual human chef was generally seen as less troublesome.

It seems clear that any acceptable autonomous solution would need to gear-in with the mundane domestic routines in a manner that would respect the red lines outlined above and avoid interfering with them. Convenience of use is thus an important factor in the acceptability of autonomous systems not for convenience's sake, but rather as it warrants technology use by increasing quality of life, such as to free up time for valued activities like spending more time with loved ones.

5.3.2 Reasoning about food automation

Our participants were able to acknowledge that even the most mundane household tasks, such as tracking and restocking of food items in the fridge, can often become too overwhelming to manage given the hectic nature of everyday life. Small wonder then that the proposed autonomous food monitoring and reordering system depicted in scenario 4 was initially greeted with enthusiasm. As the discussions deepened, however, problems began to emerge, with participants increasingly questioning whether the different variables relevant to food management could - or indeed even should - be monitored and assessed by fully autonomous systems.

> ***Berta:*** *Well, for example I'm vegetarian, so it would need to know my dietary of carbons.*
>
> ***Keith:*** *Frequency at which you eat, like how often you tend to have your meals…*
>
> ***Max:*** *Yeah and you mentioned about types of food, whether you're eating healthy or not.*
>
> ***Keith:*** *And how long it takes you to get through your stockpile of your provisions.*
>
> ***Lana:*** *It needs to know, it would have to have so much portions… I eat less than my husband, he's not fat, but I eat less than him and the kids, they're different ages. The*

amounts of information that it would need… you know what I mean, if you're ill one week, it would need to be updated, coz you're not eating because you're being sick. Or this week it's Ladies Night, so I want even more chocolate, do you know what I mean? It would need that much updating constantly. So all the time that you're saving, you would just spend on updating a robot.

[...]

Peter: *I was sort of thinking maybe non-food stuff you get… something like bog roll, if you kinda got a pack of six delivered so you never run out of bog roll. I guess if you run out of bog roll, that's kinda stupid.*

Researcher: *So if there's no skill involved in doing that, you wouldn't mind automate it?*

Zikmund: *Yeah, if it helps keeping my home organized, I'd like it to purchase more.*

There is a great deal of information a system would need to know about its users in order to manage their food items efficiently. Such a service would for example require detailed knowledge concerning the user's eating habits, including whether they are vegetarians or whether they suffer from any particular form of food allergy. As Lana explains, different members of a household have frequently very different needs and an acceptable autonomous system would thus need to be able to identify the patterns in consumption and preferences of each individual in order to make sure everyone is satisfied. To make matters worse, as the extract above illustrates, such needs and preferences do not merely differ from person to person, but likewise across time. Our needs are subject to constant fluctuations in response to various external factors, some of which - such a sudden sickness - appear near impossible to predict. The inevitable presence of such contingencies made our participants question whether any autonomous system could really predict their consumption with sufficient accuracy to make autonomous food reordering reliable.

Besides the largely unpredictable nature of the everyday domestic life, another factor contributing to this sceptical sentiment came in the form of privacy-related concerns and a heightened self-awareness stemming from food monitoring (more on that under social accountability). Many of our focus group participants voiced the opinion that food consumption represents a highly personal matter and any significant degree of oversight and delegation to a computer system would make them feel uncomfortable. It seems apparent that a socially acceptable autonomous system needs to walk a fine line between collecting rich enough data to enable accurate coordination with the household members, while avoiding situations where this monitoring activity would infringe on such user preferences.

As Peter made perspicuous, not all household consumables, however, suffer from the problems outlined above. Bog rolls, were brought up as a specific example of a product that is being consumed regularly, without much fluctuations, as well as being devoid of personal significance.

As an alternative to a fully autonomous system, some participants also suggested that users ought to be able to manually schedule future automated food purchases. While generally agreeing this might improve the acceptability of an autonomous system, other participants argued that such manual coordination would essentially eliminate the key selling point of an autonomous system. For, as Lana put it, instead of saving the user's time, users would likely end up spending all their time on "updating a robot". In such cases, the majority of our participants ended up advocating that rather than automatic physical actuation, an acceptable system ought to merely provide the user with recommendations.

5.3.3 Reasoning about recommendations

We have already established through our previous user study that an autonomous system coordinating its actions with its users by proactively reaching out to them, such as by using voice-based messages, runs the risk of stirring up potential frictions

stemming from questionable situational relevance and timing of such interactions. In our second user study we went a step further, postulating various autonomous systems utilizing a range of food recommendations as a tool for coordinating automatic food reordering with their users. While scenario 4 depicted technologies providing their users with food recommendations based purely on their preferences, scenario 3 restricted the choice to local and seasonal produce. Perhaps unsurprisingly, choice and autonomy became some of the most persistently discussed issues throughout our focus group sessions, whenever food recommendations were brought up.

> ***Petra:*** *And with any real app... for example even those recommending available seasonal produce... you can always program an app to tell them "I'm allergic to this", they will start seeing patterns in the food that you prefer, so I would assume it would make suggestions depending on your past patterns. Whereas that [scenario 3 food recommending application] is giving you a limited choice, it's obviously… I think it appeals to a different type of person, doesn't it?*
>
> ***Amy:*** *But I got a friend who only eats Hawaiian pizza. It's about the only thing she'll eat. So how can she be encouraged to eat a little bit more variety of stuff? And she's in her 20's now, she just chooses one pizza bowl. If it only was about people's preferences, you know...*
>
> ***Salman:*** *Yeah, it should suggest more healthy meals, more healthy alternatives to…*
>
> ***Amy:*** *...but her mum has been doing that for years.*
>
> ***Petra:*** *If there's still a choice out there, it can make all the suggestions it wants, but they will usually go back to their own preference. While if they say "This is the only stuff that is available, you have to pick one in this field", then they don't have a choice… their choice is limited.*

A view that was largely shared by our participants was that the amount of choice available to us is often overwhelming. Being able to choose between a wide range of

items which are available to buy was suggested to, for example, create a situation in which more food ends up wasted, because people might over-purchase. Paradoxically, as Amy explained, an overabundance of choice might also lead to a situation where people start making less varied decisions, such as ordering pizza for lunch every day.

Consequently, the majority of our participants were in favour of autonomous systems that would utilize recommendations as a means for preventing such "choice overload". While pushing the user towards a more healthy and sustainable diet was, for example, generally seen as a commendable goal, participants were also able to recognize that by limiting the user's freedom of choice, autonomous systems would effectively be exercising control over their consumption habits. In order to preserve the autonomy of the human users, it would be imperative for the user to be able to specify precisely which criteria are used to put a constraint in place. Various criteria would be desirable depending on each individual's preferences and situation. As Petra put it, different types of constraints would "appeal to a different type of person".

Furthermore, a technology making uninvited suggestions (e.g., by altering recipes in order to use up certain items that are approaching their use-by-date) could be seen as patronising as it implicitly tells a person what to do. For instance, in scenario 4, a robot decides when to discard unused food items. One participant argued that technology should not take away the choice to manage our own lives. As the participant explained, he would rather decide for himself what to cook and when to discard food items. Similar arguments were brought up by other participants who argued that they are frequently cooking for themselves and value doing so. Participants thus emphasised the importance of being able to pick and choose for themselves when to rely on recommendations produced by an autonomous system and when to do things on their own. A deluge of recommendations and directions, they argued, would eventually deprive them of much of their freedom to *think for themselves*.

> ***Researcher:*** *But what about... because the system is also in your fridge, and it makes recommendations of stuff you could cook from that. So you would still do the cooking, but it would be giving you advice.*
>
> ***Amanda:*** *Then you're not thinking anymore, are you?*
>
> ***Samantha:*** *I can just see a different norm developing there. Lavish meals and social dinner parties not being as personal anymore.*

As the vignette above makes perspicuous, there are situations where freedom is desirable, rather than overwhelming. Being able to enjoy personal autonomy when managing household tasks, such as when cooking, might for example be desirable. Similarly, as Samantha suggests, simply accepting a recommendation, rather than converging on one's own free decision, might result in food purchases that feel less original and personal.

While our participants generally valued technology that aids user choice, particularly when it comes to making more sustainable or healthy choices, the above statements show that there is a fine line between helping guide human decision-making and constraining it in a desirable manner. A system's failure to reach such an equilibrium results in a situation where it comes across as patronising the user, and worse, potentially restricting choice to a point where the users orient to the technology as intruding into one's personal freedom. Protecting human autonomy, in this sense, demands mechanisms that would assure that the user always stays in control.

5.4 Control

As made perspicuous from participant's comments surrounding the issue of coordination, a substantial level of user control is frequently desirable in order to make sure the user remains in charge. To a great extent, this sentiment appeared to

be fuelled by our participant's concerns of losing control over their own intimately familiar domestic environment. While our scenarios were vague when it comes to depicting specific control mechanisms, the need for control, and the risks associated with the lack of it, still emerged as a major point of discussion in all of our focus group sessions.

5.4.1 Reasoning about loss of control

The mere fact that a system has the capacity to act on its own, without direct human oversight, sparked mistrust amongst some of our participants, many of which theorized that due to malfunction or malevolence, a system could slip out of control and turn against its human users. This became particularly apparent when debating the physically embodied robot depicted in scenario 4.

> ***Lana:*** *But isn't it the same thing with that robot, you don't know who's controlling it? You know what I mean, it could start a fire in your own home, you don't know, because you don't know who is controlling it. So it's the same thing.*
>
> ***Keith:*** *You kind of just see, is there anything in my house where it could do physical harm? I mean as long as it's easy to access and switch off…*
>
> ***Lana:*** *Yeah, but that robot may go rogue when you're at work, because he knows where your whereabouts are. I may be a bit extreme here, but you know what I mean.*
>
> ***Berta:*** *I think when looking at Roomba… it would not bother me. Coz I'm controlling it and…*
>
> ***Lana:*** *Yeah but you're not…*
>
> ***Jane:*** *But I am coz I can switch it off...*
>
> ***Lana:*** *...software is controlling it.*

> ***Jane:*** *Well the software... if someone was gonna hack into it, then someone is gonna hack into it… there's nothing I can do.*

A prevalent underlying theme of our focus group discussions was the notion that we are in the process of losing control over our lives as a consequence of technological progress. Debating the various autonomous systems proposed by our scenarios appeared to be only further amplifying this concern and steered the conversation into increasingly dystopian directions. At one point, participants suggested that the adoption of fully autonomous home systems would constitute a final nail in the coffin of our freedom, resulting in a state where the user would effectively become controlled by technology, rather than the other way around. A smart home, they suggested, would end up dictating every aspect of our lives, from the food that we eat to what we do and how we manage our free time. One focus group participant (Peter) even argued that much of the control we think we have is in fact an illusion. Although we might feel as if we are in control, he explained, when commanding a system to carry out a specific task, the commands at our disposal have in fact all been determined by the service provider.

Besides depriving the user of freedom, this lack of control turned out to be a major cause behind a range of trust-related issues. Participants were frequently questioning whether systems that operate fully autonomously are not in fact being someone else's puppets and whether there is any guarantee that the decisions they make are in the best interest of the user. Some participants went as far as suggesting that an autonomous system, if left unattended, could cause physical damage to their household, such as by starting a fire. Such negative scenarios painted by our participants were typically accompanied by a degree of defeatism, with participants arguing that they would be defenceless had an autonomous system in their home gone out of control. As exemplified by Jane in the vignette above, should a hacker attack occur, there's nothing she can do.

When prompted to propose a remedy to such control-related problems, our participants were generally suggesting that the user should always have the means to disable the system or any of its features. One such feature that got specifically brought up were autonomous data collection activities. When discussing the camera monitoring system installed in user's smart fridge in scenario 4, multiple participants suggested that a socially acceptable solution would need to let the user turn it off, not just for the sake of being in control, but also out of privacy related concerns.

Overall, participant's responses made it clear that future autonomous solutions will need to be designed in a manner that would grant the users a sense of ultimate control over their systems, even if this means "pulling the plug" and terminating the system operation. Our study suggests that in the case of domestic autonomous systems, the solution to such a design challenge could be as simple as implementing a visible off button. While not directly alleviating the underlying limitations in artificial intelligence and power distribution, our findings indicate that the knowledge of such an emergency control mechanism would help give users a much needed peace of mind.

5.4.2 Reasoning about granular control

While an off button might serve as an effective assurance for the user and thus help improve the overall acceptability of an autonomous system, the practical utility of such a feature would likely be limited. Just as in our first user study, participants once again suggested that a more granular control mechanism would be needed in order to better tackle some of the more mundane challenges. As illustrated by the extract below, which captures a debate concerning autonomous systems cooking meals for their users, different levels of manual control are desirable in different situations.

Amanda: *I also think that it [a fully autonomous system] would appeal to a different type of audience. So like the normal person would want to keep, you know, the sentiment of cooking for themselves and buying things for themselves. But there are people that are actually busy, they own businesses or are doing two or three different jobs, traveling and whatever… this thing would be amazing for those types of people. Then again you have those people that don't have that type of lifestyle and they have more time on their hands to actually cook. So I think… umm... I don't know how that would, you know, work in terms of merging the two groups of people together.*

Researcher: *So you might make a split in...*

Amanda: *Yeah…*

Samantha: *Or maybe you could have different levels of what you want it to do for you and you could choose what complexity of services you want.*

Mary: *I think it's a bit sad like if you're not cooking for yourself there [refers to scenario 4], do you know what I mean? Like you would do some stuff for yourself, like obviously there's the convenience of it, which is great, but I don't know if it does everything you don't or if you can pick and choose when it does stuff for you.*

Samantha: *Or switch it off on a Sunday.*

Amanda: *A Sunday roast!*

As the participant responses outlined so far have made perspicuous, the practice of delegating our daily tasks to autonomous systems comes with both risks and opportunities. On the one hand it can free up our time and make our lives more convenient. On the other hand, by doing so, it can also impact our lives negatively, such as by impeding a sense of social cohesion, or making our daily lives feel less enjoyable and meaningful. As one participant noted, being in manual control and doing something on our own, such as cooking a meal, gives us satisfaction. Being in full control of such tasks in this sense becomes a tool for our creative expression and

self-actualization. By the same token, giving up control over such aspects of our lives would drain our motivation to do anything meaningful.

All such views expressed by our participants point towards the conclusion that an ideal autonomous system would need to allow its users to "pick and choose" the degree to which they prefer to rely on the assistance of an autonomous system and when to do things manually. Unsurprisingly though, there was little agreement amongst our participants regarding where precisely this point of equilibrium between full automation and full user control lies. There is, in other words, no such thing as a one-size-fits-all approach to control. Rather, a deluge of variables pertaining to the individual users and their momentary situations must be taken into consideration. As Amanda explains, people with busy lifestyles might for example benefit from greater degree of delegation to autonomous systems. Similarly, the level of technical skills possessed by the user was brought up as another important factor. Digital natives might for instance benefit from a greater level of control over their autonomous systems than their older peers.

An acceptable autonomous system would thus need a highly flexible control mechanism, allowing it to accommodate for the varying needs and preferences of diverse users.

5.5 Computational accountability

Much like in our first user study, computational accountability surfaced as another major problem area, with our focus group participants debating which aspects of autonomous system behaviour ought to be explained to their users, and which are better to remain black-boxed. As we unpack below, a number of potentially relevant personal values and stances were invoked throughout these debates.

5.5.1 Reasoning about ease of use

The adoption of novel technologies typically requires the user to invest time and effort into getting acquainted with it to properly learn its use. Correspondingly, a novel technology that would be intertwined with nearly every aspect of our daily life, such as the one postulated by our scenarios, was seen by many of our participants as requiring substantial effort from its user. Some even went as far as suggesting that the hardship associated with learning to use such a technology would outweigh the benefits from its use.

> ***Amy:*** *Depends on how complicated the system is. I think that with recycling, I feel now, having recycled for years and being really keen on it, I feel guilty that it got so complicated about what you can and can't put in. And occasionally I think I can't be bothered to read whether it is the one and I just chuck it in the main bin, I have done this a few times. And I feel so guilty, but it's because of the whole thing.*
>
> ***Salman:*** *It's over complicating, that's the key. It's not to have it overcomplicated.*
>
> *[...]*
>
> ***Fiona:*** *You'd have to be tech savvy to be able to use it. Like older people wouldn't be able to use it. The elderly who can barely even access the internet and who still have basic phones, they wouldn't be able to use it... and it would probably be really good for someone like that. Who needs more help in the home as well. So it would probably like need to be really, really well explained... and taught to people. And then if the system goes down - there needs to be like another way... with the device... or app... give everyone a diary so that not everyone is caught by surprise when their food goes out of date.*
>
> *[...]*
>
> ***Carl:*** *You know, how this system needs sell-by dates or use-by dates is quite important, because at the moment we haven't got anything that does that and the supermarkets have tried to do that in supply chains and so you know it's not… it's*

> *not a great stretch of imagination to think technology could have an additional code that would contain sell-by date, and you could have that and introduce it across the world, I mean there's no reason why not and barcodes are standard, they're... you know, the standard across the whole world, so you know there's no reason why you couldn't have another code, maybe a QR code or something that…*
>
> ***Jack:*** *I'm with you, I wouldn't have a clue about the technology, I'm, you know... again this is an age thing I suppose.*
>
> ***Amanda:*** *No, I would just trust it works. Just assume that it does, I don't know how it does it but it does it.*

When confronted with the autonomous systems depicted in our scenarios, multiple focus group members voiced their concern that such technology might be too difficult to understand for the average user. This became particularly apparent when debating scenario 3, which postulated an autonomous system directing the user to donate or recycle various food items that were going out of date. Participants argued that some of these tasks are already overly complicated the way they are, hence taking the easy way out and "chucking" old items in the bin was frequently seen as a preferable option. Having such household activities augmented - and to a degree directed - by an autonomous system would only be acceptable should it avoid introducing additional complexity into the daily domestic life. As Salman notes in the extract above, not overcomplicating things is the key.

Our participants were thus in general agreement that an acceptable system would not only need to be able to explain particular directions to its user, but also how to avoid taking these directions in cases where they are perceived as too complicated. Instead of merely spelling out the steps to be taken (e.g. take an item to the recycling station), a system should also explain how to best configure and customize the system to avoid having to take such steps. Perceived ease of use goes in this sense hand in hand with a system's ability to explain and make its functions intelligible to the user.

Whereas a technically skilled individual might find a domestic autonomous system intuitive and accessible, our participants were concerned that too much complexity would be prohibitive for non digital natives, particularly the elderly users. As Fiona put it, an acceptable system would therefore "need to be really, really well explained... and taught to people".

Fiona went on to suggest that users should develop an alternative analogue way of operation, such as a physical diary tracking food items in the fridge. Not only would that make the relevant procedures more understandable to non digital natives, but it would also serve as a backup solution in case the system would for whatever reason malfunction and go offline.

Understanding of the relevant technical procedures did indeed turn out to be difficult for our participants. When prompted to reason about the inner workings of the depicted systems, we came across numerous instances where they struggled to describe how they thought the technologies functioned. For instance, the process of detecting and tracking use-by dates of food items in the fridge was seen as impossible to achieve using existing technology. Some of our participants, such as Carl in the extract above, responded by trying to fill-in such perceived gaps in our scenarios and invoked their own hypothetical solutions (such introducing qr-codes storing use-by date information). The debate that followed Carl's monologue reveals that understanding the inner workings of a system is however not typically the user's priority. As Amanda's stance reveals, a system that is perceived as working properly should not need to make all the details concerning its practical aspects explicitly transparent for the user. Instead, as we shall see in the following subsection, other aspects of an autonomous system's behaviour, rather than its inner workings, are in need of explanation.

5.5.2 Reasoning about personal data legibility

Multiple aspects of autonomous system behaviour were brought up by our participants as being in need of transparency. The one that appeared to stand out the most during our second user study was concerning monitoring and collection of personal data. Whereas all our participants understood that some level of personal data collection is necessary for the functioning of the depicted systems, the degree to which they saw such practice as being acceptable depended heavily on whom such data would be made intelligible to. As the vignettes below show, this became particularly apparent when debating the tracking of shopping habits as well as community food exchange.

> ***Fiona:*** *If you're happy to shop online, they track your order that way. So... you should be happy about that, but…*
>
> ***Martin:*** *Even if you're searching for things... it comes up because you've searched for toothpaste it'll come up and say...*
>
> ***Fiona:*** *Yeah, Facebook will tell me that my partner has been looking at god-knows-what on amazon. And stuff like that. Everything tracks everything.*
>
> *[...]*
>
> ***Researcher:*** *Would you like to know where your food has gone then? Not just who you're getting it from?*
>
> ***Robert:*** *I suppose so, if it's like you say, if it's going to, you know, pigs or feed some animals or something… that would be interesting, to find out where it ends up.*
>
> ***Mary:*** *Now that I think about it, would I want people to know where I live, would I like my random neighbours to be like "pay this money, these are my details" like update me like that.*
>
> ***Researcher:*** *So what happens if you have an avatar, like probably your fake email and stuff, certain things maybe like… if it was your username rather than your real…*

> ***Mary:*** *...You don't even trust them, if you knew who they were... But then I don't think I'd feel comfortable giving away my details to random people like so they can come find me if they wanted, you know. Changing that sounds safe in that way…*
>
> ***Samantha:*** *Makes you think about confidence, doesn't it, it's like the whole thing shifts.*

Our participants were generally in agreement that the user should be able to see precisely what aspect of their activity is being monitored and also be able to review any data that gets stored about them. A system that is collecting a range of metadata about our shopping habits in order to produce recommendations or take other proactive actions must articulate the particular information based upon which the recommendations or proactive action was taken. Doing so will in turn help it improve the overall perceived transparency and legibility of the system.

As the dialogues above shows, in a social setting, such as the home or a local neighbourhood, the ideal level of personal data legibility becomes somewhat multi-layered and heavily dependent on the various parties involved.

Even the most mundane information, such as a person's diet, could give rise to negative consequences if falling into the hands of the wrong people. Amanda, for example, explained that if the insurance company got access to information regarding the food a particular person eats, it could end up raising the insurance costs should the person be deemed to be eating too unhealthily. Similarly, whereas our participants generally argued that they wish to have access to stored data that relates to their own activity, having the activity of other family members conveyed to them could easily result in awkwardness. Likewise, sharing food with neighbours or the broader community would only be acceptable should all personal data be anonymized. Multiple participants also argued that they would prefer all such data to stay within the given community.

It seems reasonable to conclude that acceptability of future autonomous systems will be conditioned by achieving such a selective legibility, with personal data being made legible only to the appropriate people. Properly safeguarding the desired level of personal data legibility will require the user to be able to see not just their own data, but also the existing flows that their data is a part of.

5.5.3 Reasoning about terms of use

While an autonomous system's ability to explain its behaviour to the user is of great importance, we did also come across instances where the burden of explanation appeared to shift towards the system provider, with participants raising expectancies regarding the manner in which the provider ought to communicate with its customers.

> ***Mary:*** *It's like Apple, isn't it? I hate Apple for being so good almost. My family has all got Apple because you can do FaceTime. If you haven't got Apple, you can't. I didn't have it for a while, I was just the odd part. Then I got one of my own. So you know… I would pick Apple products now.*
>
> ***Researcher:*** *Yeah, so you resent that you don't have any choice?*
>
> ***Mary:*** *Well they're good products, but you still feel like you owe them that loyalty. It just suits your family life, so you feel tied into choosing Apple again.*
>
> ***Amanda:*** *But I think it's good to have the option of choosing, like for example with this I'm sure it'd probably be like a lifetime subscription, like you subscribe to the robot thing, or the different things you have to buy. You have to buy the robot and you have to buy the thing that's gonna turn up on your TV or your car, everything. And I personally don't subscribe to things that I know I can't get out of when I want to. So for example, like the gym, I subscribe because I know that anytime I can just cancel my subscription to go to the gym. And we get contract phones because I know at the end of the 12 months I can cancel it and get something else. If EE were to tell*

> *me this is a lifetime subscription, I wouldn't get it. I think I'd just do the… what's it called… pay as you go. I don't know about getting locked in… I don't like it.*

The focus group debates made it apparent that most participants would need a great deal of clarity regarding the manner and degree to which the system provider would limit their freedom of choice. The exchange above shows that our participants expect the different smart devices partaking in the functioning of a system to be intertwined into a coherent ecosystem. This in turn prompted participants to question whether the compatibility constraints resulting from such ecosystem would prevent the consumer from acquiring new smart household tools as they liked (e.g. a new smart fridge), instead forcing them to purchase devices of the same brand as their autonomous system.

The notion of automatic food restocking services turned into another point of discussion. Participants felt like it wasn't fully clear what role the provider plays in what food items ought to be reordered. For instance, if such a service would be provided by a major food chain, such as Tesco, what impact would this have on the brand of food items being restocked? Would Tesco exploit their position to prioritize reordering of their own products at the expense of other brands or would they reorder items that the user genuinely needs? A reoccurring view amongst our participants was that instead of just determining what food items ought to be purchased and ordering them autonomously, an acceptable system should rather track available food items in local stores and maintain a catalogue for the user to browse. The actual order would then be decided upon and placed by the user, rather than the system.

In summary, there was a broad consensus amongst our participants that all ramifications stemming from being subscribed to an autonomous system service ought to be made clear to the customers before they sign up. This would also entail communicating the conditions under which the user can get out of a subscription.

Any infringements on the user's freedom, in short, must be made legible and available for users to scrutinize.

5.6 Social accountability

In line with our initial study, social appropriateness of autonomous machine behaviour quickly emerged as a dominant theme across all of the focus group sessions. The picture that surfaced this time was however noticeably richer and more nuanced, spanning more direct concerns raised by participants trying to relate the depicted technology to their own life, as well as broader concerns pertaining to the wider societal impacts of autonomous systems.

5.6.1 Reasoning about privacy

Both of our scenarios were emphasizing the collection of personal data, such as through cameras and other sensing technology, and its role in the functioning of an autonomous system. It is perhaps not surprising then that a majority of the comments brought up in relation to social accountability were dealing with privacy in various forms.

> ***Amanda:*** *Oh this is so scary! I wouldn't be comfortable with that to be fair. I mean the convenience is really, really good, like "oh yeah you haven't had a fish for a two weeks, maybe it's a good idea to have fish today" but then… like having all that data... like you know the google [storing of search keywords] is fine, coz that's just my shopping. But like the intake of my food, what's running out in my fridge... I think that's a bit personal. I think maybe if it was like um… giving someone a shopping list and then they deliver the shopping to your house, I think that would be fine, but turning up on your TV? [Referring to the closing scene in scenario 4] I know it's an exaggerated scenario but yeah… I think it's a bit too much.*

> ***Samantha:*** *If someone hacked into the camera in my fridge to see what I have...*
>
> ***Amanda:*** *This way it also takes away the choice, well it wouldn't take it away, but you feel uncomfortable if you looked at the meats and it went out of date yesterday, I'm going to smell it and see if that's alright. Yes, it's fine, I'm going to cook it and eat it. Personally, I always use my judgement whether something is in date or out. I'd hated it if it was decided for me.*
>
> ***Robert:*** *You would feel quite judged though wouldn't you?*
>
> ***James:*** *Big Brother is watching you! [...] I was thinking, that robot is obviously owned by a company that has you pay them some subscription, so it's like Netflix, or whatever, account. And so that company is quite blatantly keeping track of everything. It's gonna be a service somewhere around the world, you don't know where it's gonna be, and they're just logging all of your data. So in a similar way that Facebook knows who you're friend with or they know where you've been on that particular day. This takes it even one step further and they know what you like to eat, they know what you're allergic to. And it's someone that you never met and it's probably really focused on… well their bottom line is the money really, isn't it? So they're not too bothered by taking care of all the information you have. So I think in some ways it's really good that the convenience is there, but you're kinda sacrificing like privacy for convenience.*

Our scenarios hypothesized that future autonomous systems could employ cameras and other sensing technology to monitor the food items in user's fridge. By keeping track of whether any food item is running out or approaching its use-by date, our scenarios suggested autonomous systems could help their users live more convenient and sustainable lives. While our participants agreed that consuming food before it goes out of date is important, they were not fully convinced that this should be done through technological means. As Amanda points out, it is reasonably easy

to figure out whether a particular food item is still safe to eat simply by smelling it. A similarly sceptical stance towards monitoring technologies was prevalent throughout much of the focus group discussions. By unpacking the comments made by our participants, it became clear that such an attitude had several underlying reasons.

Most prominently, the discussions were occupied by the risks of privacy invasion as a consequence of having sensing technology integrated into home appliances. The prospect of being constantly monitored made the majority of our participants feel uneasy. A few participants attempted to disperse these concerns by arguing that the monitoring technology proposed in our scenarios would in fact not mark a significant shift from the current state of affairs, arguing that companies such as Google are already actively tracking our browsing history in order to present us with personalised recommendations prompting us to buy various products. The general consensus was however that there is a limit to how explicit and intrusive monitoring activities could be while remaining acceptable in a domestic setting. Scenario 4, for instance, depicted a virtually embodied assistant turning up on a user's TV in order to provide them with recommendations or reminders. Amanda, as seen in the vignette above, felt strongly that such a practice would represent a step too far.

The data collection and processing that enabled the various solutions in our scenarios, such as meal recommendations or sustainable lifestyle directions, were seen as intruding into matters that are "a bit personal", making participants feel uneasy about being "judged" and held in other ways socially accountable for actions they perform with food items (and other activities). Important here is that the level of perceived acceptability seems specifically tied to certain objects, in that participants appeared to accept that information from their shopping list may be used, but not if the same data is taken from their fridge. Also, as illustrated above, the view tended to change from delivery to the door (OK) to being on the TV (not OK). The reasoning here changes from what may be seen as the private space of the home (and the objects within), to the world outside and the objects acceptable for sharing with that world (such as a shopping list).

As James put it in the vignette above, by allowing corporations into this private world of the home, we are in a way "sacrificing" our privacy for greater convenience. By allowing companies to monitor the private spaces of our home, we would essentially be giving away our personal information to complete strangers. As James elaborates, this is seen as particularly bothersome because the companies responsible for storing our data are not primarily concerned with the user's well being, but rather with maximizing their own financial profit. Consequently, participants were generally in agreement that the providers of autonomous systems cannot be fully trusted to serve the best interest of their users.

To make matters worse, at multiple occasions, participants appeared to suggest that this mistrust towards companies responsible for personal data collection is currently on the rise.

> ***Researcher:*** *I mean would it bother you that… so obviously it has to collect a lot of data about you, about what you like and what you don't like. Any thoughts there?*
>
> ***Simone:*** *I don't think I would trust people with data as well, especially after the past few years, with like the government having like access to our lives and personal data and stuff like that and I think now as well. So you store so much information about yourself in the cyberspace, it's like all you get is in the cloud, things like that. I mean I wouldn't say for someone with like a degree in computing and stuff like that... in IT... it wouldn't be hard to trace and find whatever stuff even if it's not on the internet and actually runs on a computer. So I think if it puts you at risk... I mean I don't want to put that much information about myself out.*

A series of recent events, including the Snowden revelations, appears to have increased people's awareness of the dangers associated with monitoring and personal data collection. Indeed, some participants confessed that they are already

opting out of existing services, including Facebook and other social media platforms, out of security and privacy concerns.

All of this is contributing towards a situation where people are wary of giving away too much information about themselves. Having a few selected activities monitored in exchange for customized recommendations is one thing, but participants were decisively against the notion of granting any service provider a blank cheque enabling them to monitor just about anything in their lives.

The responses provided by our participants made it perspicuous that a greater sense of trust will need to be built between the providers of autonomous systems and their customers. As elaborated in section 5.5.3, one precondition for such trust would be that service providers give out clearer information regarding the personal data that is being collected as well as the reasons for this collection. Multiple participants reflected on their personal experience and expressed frustration over being asked to approve collection of personal data without being informed of what the information is going to be used for.

Similarly, our participants argued they would need to be reassured about the security aspects of the data collection. Participants voiced concerns that should the security aspect of personal data storage be neglected, any individual or organization with sufficient technical skills could easily get access to any of their personal information, including where they live and what they are doing.

It is evident that a sense of privacy is dependent on a multitude of locally organised factors, such as the physical implementation of the technology in this case, the nature of sensors, their location and the data being collected, as well as more global factors, such as the perceived trustworthiness of the service provider. Together, such factors can make all the difference between acceptable and unacceptable.

5.6.2 Reasoning about community sharing

A closely related expectancy expressed by our participants had to do with the sharing of data within a local community. Scenario 3 approached personal data collection from a somewhat different angle compared to scenario 4. Instead of focusing on large corporations harnessing personal data in order to provide their users with relevant services, more emphasis was placed on the use of personal data to coordinate sharing and exchange of goods between members of the local community. Monitoring technologies were in this sense depicted as being used to enable collaboration between people in a neighbourhood. When realising she is running out of sour cream, a user could thus for example request a donation from a neighbour who is in a different situation (e.g., who has an excess of sour cream approaching its use by date), thus allowing the community to achieve a more efficient distribution of food resources, as well as eliminating waste.

> ***Salman:*** *yes I don't think I want my neighbours to know what I have in my fridge. But I'm happy for them to know what I have donated to the pot as it were. And you can pick and choose, I have got this fish, does anybody want it? And if anybody wants it they can...*
>
> ***Petra:*** *I like the idea of it being totally anonymous because of that whole judgment thing again. You don't want to know what's in the fridge, what's out of the fridge or more. They don't want to know, well, and that we put all of this in the pot, all of these people have taken and not contributed, it might just be that they are more efficient than other people. But there is always that level of "oh we have done more than that…"*
>
> ***Jim:*** *I don't think there is any need for putting your name on it. You know, why does it matter who put it there?*
>
> ***Petra:*** *It's quite like that... big thing in the media at the moment... where it is about body shaming. If someone is slightly overweight they don't want to*

> *be judged going for that piece of chocolate cake from the fridge, do you know what I mean?*

Upon a superficial reflection, most of our participants appeared to be more open towards such local collection and sharing of personal data. There was a general understanding that food exchange between community members serves a good cause and that monitoring and data collection would play a crucial role in making such exchange mechanisms feasible, such as by allowing a community member to verify that a particular food item offered by another member was in fact in an edible state before submitting a request for it.

However, as illustrated by the vignette above, upon a closer inspection, perceived problems started to emerge. In spite of the system having good intentions, our participants argued that sharing information about their food with the local community would make them feel self-conscious and judged by their peers. This could lead to various pressures on their behaviour, with users fearing to be looked down upon for leaving food go past its use-by-date, for example. Petra went as far as suggesting that this peer pressure could also lead to body shaming, with overweight users feeling guilty about consuming food when knowing that the community is watching.

In other words, participants felt uncomfortable about someone else being able to see their potentially wasteful - or in other ways disagreeable - behaviour. Once again, a dominant feature of the expectancies listed above is the uneasiness that people feel about being held accountable for how they behave in their own homes.

While utilizing autonomous systems to help facilitate collaboration between community members has its potential, more needs to be done to make systems respect existing hierarchies and relations between members of the community. The failure to do so will in this sense cause tensions by putting pressure on people to justify their lifestyle.

5.6.3 Reasoning about embodied autonomous systems

Besides issues pertaining to sharing of personal data with various external parties, another expectancy that emerged was more concerned with the nature of the technology carrying out autonomous actuation inside of the user's home. In particular, the notion of a physically embodied humanoid robot as a personal assistant sparked considerable debate.

> ***Keith:*** *The idea of a robot servant, which has been a staple of science fiction for ever but has never really actually happened in that way. Um… it would be exactly like having a servant, which most people are not used to. And you don't know… I would not know how to approach this extra presence in my environment. Like, do I just ignore them, because I'd be in my rights to do that, wouldn't I? Or do I humanize them, or do I start feeling in debt when I am thanking them and say please and all that.*
>
> ***Max:*** *I'm sure they could be made look more human.*
>
> ***Researcher:*** *So if it was less sci-fi, if it's automated still, but more in a kinda human form, if it was just...*
>
> ***Lana:*** *No, there's no emotion in it, there's no connection, there's no relationship to build with a robot. I get the drone, I think that's coming in, I get the fact that he has a self-driving car, coz we're partially there, but the fact of having a robot in your house is I think quite emotionless, there's no relationship to build with it, because they haven't got any feelings such as a human to say thank, to say these things, because it won't be appreciated, because it's... not human.*
>
> ***Researcher:*** *If it wasn't a humanoid robot but like you know those remote Roomba ones that are just like a disc, if we say it didn't make any attempt to look human-like, would that change it at all…?*

> ***Lana:*** *...It doesn't matter what it is, whether it's sci-fi is or whatever. Just that object telling you "you need to pay this" and "this is what you can have for tea". You're kind of like losing your…*
>
> ***Jane:*** *I quite like it. I think if someone had like a very busy day, you just want to come home and know there is something that they don't need to have an emotional attachment to.*

When confronted with our scenarios, multiple focus group participants expressed dismay over the fact that technology was pervasively implemented into practically every aspect of the user's life. This dismay appears to not simply have been caused by privacy or security related concerns, but rather by a sense of alienation that this technology introduced into the otherwise intimately familiar home environment.

This became particularly apparent when discussing the idea of a humanoid robot serving as a domestic personal assistant, as hypothesized in scenario 4, with multiple participants suggesting that its deployment would essentially become an alienating element in the existing social structures of their household. As the vignette above illustrates, participants generally argued that the presence of such an anthropomorphic agent would leave users in a state of cognitive dissonance. On the one hand, they would feel obliged to treat it as a human being by for example manifesting polite behaviour towards it. On the other hand, they would also feel aware of its non-human nature, making it impossible to develop any true social bond or genuine social relationship. As Lana exemplified, any attempts of non-human agents to develop human-like relationships with their users were making our participants feel uneasy.

On a similar note, another participant claimed that he would never entrust a robot with managing the household. Much like in the case of other autonomous technologies, there is no way, he argued, of being absolutely sure whether the robot is actually doing as it is told, or whether it is in fact following some ulterior motives. This sentiment appeared to be amplified due to the perceived human qualities of the

robot, making it appear as more capable than more traditional embodied autonomous agents, such as Roombas.

Moreover, we actually tend to enjoy some of our household tasks and delegating them to other agents – human or nonhuman - would not be desirable. A robot could in this sense be handy when taking care of some of the less intrusive tasks, such as provide recommendations and guidance, however tasks with social ramifications and relationship-building implications - such as cooking a dinner for the family - should remain in human hands. Jane was the sole participant of our focus group study who was willing to fully embrace the idea of a household personal assistant, arguing that by having it carry out household tasks on our behalf, we would in fact free up more time for the family members to spend with each other, thus strengthening the social bonds in the household.

5.6.4 Reasoning about inconvenience

Many of our participants' comments pertaining to the "coordination" theme suggested that autonomous systems ought to complement domestic routines in a manner that would not merely maximize convenience of the user, but also help support existing social bonds in the household. Here, under "social accountability", we revisit this strand of our focus group discussions once more and elaborate the potentially detrimental implications of autonomous systems on social bonds and cohesion in more detail. While some level of convenience was found to be desirable, a number of predominantly social factors contributed to our participants arguing that preserving some level of inconvenience would be equally important for the acceptability of future autonomous systems.

> ***Simonne:*** *I think you're gonna just be thinking like… how inconvenient is it really to pick up something in your fridge and look at it, I think everyone… coz technology, we obviously do use it and it's useful. I think again it's just not a very good thing to incorporate it into everything. I*

> *mean there's nothing… it's not that hard just pick up… it's not that vital that you need to be told when something's gone off. It's not like you're gonna forget eating and cooking and you're gonna die. Like, you'll be fine, you don't need to be told things like that, you don't need it to wave in your face that something has gone off date in your fridge, I'm sure you'll find out. I just think like... if there's not a bit of inconvenience in our life... everything can't be like super easy, coz we need to be challenged.*
>
> *[...]*
>
> ***Researcher:*** *So what about a system that deliberately puts bumps in the way, I mean like glitches in places that you'd have to deal with?*
>
> ***Zikmund:*** *But then it wouldn't be worth the money…a broken device.*
>
> ***Researcher:*** *Well you said that if it was too easy then there's no satisfaction from it.*
>
> ***Gilbert:*** *I wouldn't get any satisfaction from any of that technology.*

Our participants felt that technological solutions designed to make our lives perfectly convenient would in the same sweep also make them bland, monotonous and boring. There can be such a thing as too much free time or a too easy life. Delegating all the labours of their daily life to technology would leave a large empty hole in their lives, prompting our participants to ask themselves what would they do with all the newly gained free time. Keith suggested that if we would not be occupied by traditional daily life tasks, most people would end up wasting their time on meaningless activities, such as watching TV.

In spite of being well designed and reliable, an autonomous system could in this sense still end up affecting the user experience detrimentally. We found a broad consensus amongst our participants that technology should not be omnipresent and designed to spoon-feed us everything. Rather, some limited level of hardship and inconvenience should be seen as a desirable component of our lives, giving us a

sense of purpose and accomplishment. As Simone put it, "we need to be challenged". The extract below suggests that such "challenge" should not be produced artificially by the autonomous system, but rather results from its absence, by having it deliberately stay out of a number of daily household tasks.

Privacy and security concerns set aside, delegation of mundane daily life activities to autonomous systems should therefore still not be done indiscriminately. By further unpacking this stance, a number of key underlying factors emerged.

> ***Mike:*** *I felt really uncomfortable with it actually. Paul himself wasn't really living. He was just being maintained by a system. I thought his quality of life was probably not very good.*
>
> *[...]*
>
> ***Fiona:*** *I love to come home and it relaxes me. And I love to make other people happy by making a nice dinner. I love the look on my partners face, when... "ah that's a really nice dinner". The fact that my 3-year old will actually eat a curry is amazing. I love that. It's just... it's gonna take that away as well. Oh it's not gonna be like... "Oh mum what have you been doing all day? Cleaning and cooking.", it's gonna be like "Oh yay, thank your computer, we love you! Can you tuck us in at night?"*

By making our lives increasingly easy, participants felt like technology would gradually erode all kinds of physical and mental requirements that the everyday life poses. Removing much of the pressure we feel to do anything even remotely difficult would in turn make us lazy and passive. The problem, as a number of participants argued, is that a healthy and happy life requires us to be active. We need to feel motivated by incentives and a sense of accomplishment in order to stay active. As explained by Fiona, when investing time and effort into cooking dinner, the fact that the whole family can then get together to enjoy the dish makes it all worth it. By having autonomous systems do the cooking and other tasks for us, the incentives to do anything difficult would be gone. In fact, as one participant pointed out, with

modern voice-controlled interfaces, such as Siri, there's not even any reason to get out of the chair as everything can be controlled remotely. It was even suggested that such a 'braindead' lifestyle would make us prone to a range of diseases such as Alzheimer's or dementia. As Mike poignantly summed it up, we would not really live anymore, we would simply be maintained by a system.

In the longer term, a widespread adoption of autonomous technology could make many of the skills we are currently nurturing redundant. Self-driving cars would for example eliminate the need to know how to drive manually, while household robots would make cooking skills obsolete. If at some point in the future an autonomous system would for whatever reason break down, users would be put at great risk, because they would likely no longer possess the skills needed to manage their daily life without the assistance of computing technology; self-driving cars that fail while driving would crash, cooking robots that break would make people starve and so on. Once again, it was argued that machine automation should therefore be treated with a degree of caution when designing future technical solutions.

What's worse, participants suggested that widespread adoption of the depicted technologies would severely weaken many of the existing social bonds. In a world where everything is done for us by machines, rather than people, they argued, we would stop feeling gratitude towards each other. As the following vignette illustrates, not only would the quantity of social interactions decline, but also their quality:

> ***Max:*** *Systems like the first one anyway [the system depicted in scenario 4], they make the world more like that anyway. So they like help Paul have that lifestyle, but by doing so they actually make him harder to socialize with, coz when is he gonna meet another person?*
>
> ***Lana:*** *I found that scary, because I think it is taking everything away. The emotion… I'm not doubting that it's gonna go that way, coz it will. I just hope it doesn't during my era because I don't want to loose connection with people, do you know what I mean?*

> ***Keith:*** *I think the loss of communication is one thing, I think the other thing is it tends to make people very passive and somewhat disengaged. Although you know we saw that to an extreme in the first film [scenario 4]. Not so much in the second [scenario 3]. But I got the impression that was because Paula was quite… I don't know… 'right on' sort of person maybe, she was quite, you know, quite engaged. Maybe she's a little eco warrior or something like that in her spare time. But I think... to the average uncommitted person, it will make people who are inclined to be passive... which is most people... it will just make them even more so.*

The majority of our participants argued that they value traditional inter-human interactions and that no level of human-computer interaction could substitute for an old fashioned human-human conversation. A widespread adoption of autonomous ubiquitous technologies was in this sense seen as a threat to traditional social interactions. As Lana pointed out, we are already in the process of losing connection with each other due to technology usage. A different focus group, for instance, debated the fact that digital natives are primarily communicating with each other through their smartphones and thus contributing to the decline of social interactions. Such issues were generally seen as long-term concerns, with participants not worried about their own social lives, but rather about the social lives of their children and grandchildren. Unless the current trends won't be reversed, they argued, future generations would lose the traditional sense of social connections.

Lastly, multiple participants also pointed to potential economic and social divisions that might inadvertently arise as a result of our pursuit of convenience. On the one hand we were met by a strong sense of scepticism whether a sophisticated household robot, such as the one postulated in scenario 4, would be affordable for the mainstream consumer. It was suggested that hiring a human butler would represent a cheaper alternative. Another participant speculated that the prohibitive price of a humanoid robot would only make it available to wealthy individuals, thus amplifying the gap between rich and poor in our society.

On the other hand, a second line of argument that was frequently emerging during the focus group discussions was concerned with the ethical side of delegation, rather than the social or financial aspects. Specifically, multiple participants felt that by purchasing a personal robot to do their household work, they would in fact be taking away jobs from humans. For example, every robot designed to cook and prepare food would result in one actual human chef becoming redundant. If forced to pay for a service, participants generally reasoned that it would therefore be better to hire an actual human rather than a robot.

5.7 Key Takeaways

Our second user study helped further corroborate the four key problem areas surfaced by our pilot study. By using provocative video scenarios – instead of illustrated storyboards – to stimulate and bring forth the mundane reasoning of our participants, we were able to unpack these problem areas in greater detail and in so doing expand on our initial findings. Unsurprisingly, our participants once again found the visions conveyed by our scenarios deeply problematic and difficult to accept for a number of important reasons. Below we provide a brief summary.

Coordination was the first major problem area elaborated in our analysis. Much like in our initial study, the manner and degree to which an autonomous system ought to gear-in with the existing routines in a household proved to be a contentious topic. Whereas our initial study findings emphasized the importance of contextually appropriate coordination mechanisms, this time our participants appeared to focus more on the social nature of domestic routines. Routines, they argued, play an important role in family bonding and contribute towards a sense of social cohesion in the home. Instead of maximizing the convenience of users by substituting for them whenever feasible, an acceptable autonomous system should only overtake human labour that carries little social significance. This in turn will help preserve the existing inter-human relations implicit in many household tasks.

Our participants were however also able to acknowledge that identifying the labour that carries social significance can often be difficult. Given the intricate nature of domestic routines, it was generally seen as safer for autonomous systems to rely on recommendations, rather than direct physical actuation. Recommendations, however, are not fully devoid of risks either. By confronting its users with a set of recommended actions, an autonomous system would narrow down the user's perspective and potentially bias them towards a certain course of action. Given the deluge of choice available to most consumers, some distilling is often desirable, however our participants saw it as imperative that the user is kept in the loop on the underlying criteria employed in this process. Similar to physical actuation, recommendations thus ought to be employed sparingly in a manner that would not deprive users of the freedom to think for themselves.

The next problem area covered in this chapter revolved around control. The data collected during our pilot study appeared to indicate that instead of exercising binary control, users of autonomous systems ought to have access to a more granular control mechanism. The participants of our second user study filled in and expanded this stance. Fear of cyberattacks along with a deluge of privacy-related problems, they argued, demands a safety mechanism in the form of an "off-switch" that would allow users to terminate the operation of their autonomous systems. We have argued that such a mechanism has less of a practical use and rather serves as a reassurance increasing trust and peace of mind of its users.

When dealing with more regular use-cases, on the other hand, the participants of our two user studies appeared to be in agreement, arguing that a fine-grained control mechanism ought to be employed instead. Whereas the participants of our pilot study tended to focus on mere practicalities to justify this stance (e.g. the need to fine tune different aspects of a smart car to maximize its efficiency), the participants of our second user study went a step further, arguing that different users have different needs. A balance between machine autonomy and user engagement that fits the needs of a user, they explained, can only be converged upon if the user is given control to configure individual system features and thus customize its operation.

Computational accountability constitutes the third major problem area that we elaborated. The participants of our two user studies generally reached the same conclusion, arguing that it is not desirable for a system to spell out its inner workings to the user. Rather, our participants emphasized the need to articulate and make legible a set of other factors surrounding the autonomous system operation. Here, the findings of the two studies slightly differed, with participants in each study coming up with different factors they felt ought to be made intelligible. Whereas the participants of our pilot study advocated the need to make the grounds upon which actions are taken visible, the participants of our second user study placed greater emphasis on aspects concerning ease of use, arguing that the existing features and control mechanisms of a system ought to be made visible in order to make the system more accessible.

Besides ease of use, another important factor driving the need for accountability turned out to be the need for privacy, with participants arguing the flow of user's personal data ought to be made intelligible. This, they explained, is a necessary precondition enabling users to adjust the level of visibility for their personal data, making sure it is only available to relevant parties. As our participants explained, different levels of privacy ought to be applicable when sharing data with different parties, e.g. household members, neighbours or service providers.

On a similar note, the participants in our second user study also frequently argued that the precise competencies of the system provider ought to be articulated and made fully intelligible to users prior to singing up to any particular service.

The final major problem area that surfaced through our enquiry came in the form of social accountability. While our pilot study allowed us to lay out the basic principles behind this theme (i.e. the need for system behaviour that is visibly rational relatively to a social context), our second user study allowed us to fill in additional detail. For instance, an increased awareness of potential privacy hazards made our participants question the viability of giving providers direct access into their homes, e.g. to restock their fridges. Instead, utilizing autonomous systems to, for example,

compile a shopping list which the user could then review and submit in order to have new food items delivered to the doorstep, was seen as a more acceptable approach. While sharing personal data locally might be a precondition for enabling a range of beneficial services, such as exchange of goods and other resources, our participants voiced their concerns that such a practice might give rise to tensions between individual users, making them self-conscious about their consumption and pushing them towards normative behaviour. The notion of embodied humanoid agents represented another aspect of our scenarios that stirred up tensions amongst our participants. It is inappropriate for nonhuman entities in a domestic setting, they explained, to contain human cues or seek to develop social bonds with other humans. Lastly, our participants argued that a system should never strive to carry out all inconvenient tasks on the user's behalf. Being able to complete an inconvenient task, they argued, can lead to a sense of accomplishment and enrich the lives of the users.

The study described in this chapter has allowed us to enhance our understanding of the four broad problem areas surrounding domestic autonomous systems that we previously identified. While the feedback that we collected from prospective users offered us a broad sample of the many relevant human expectancies, there is, at the moment, not a sufficient amount of depth in our data to support the articulation of well-defined human challenges. Nor are we at this stage able to extract the key design principles that apply. In order to consolidate and refine our findings, in the following chapter we thus present a final user study, designed to help us drill down on the most relevant human concerns that we came across so far.

Chapter 6 The Drill-down Study

Our second user study helped corroborate and expand the problem areas originally identified in our literature review and - in the case of "social accountability" - in our initial pilot study. While the data accumulated up to this point provide us with a broad understanding of the key human challenges surrounding future domestic autonomous systems, we have argued that deeper knowledge regarding these expectancies is needed in order to arrive at any form of actionable conclusions. The objective of the study presented in this chapter is thus to drill down on the problems encountered so far in order to help us better address the research questions of this thesis.

Our attempt to elicit background expectancies impacting the adoption of autonomous systems in the home once again merges breaching experiments with scenario-based design and contra-vision to create provocative views on hypothetical future autonomous technologies for the home. A number of tweaks to our methodological approach had to be implemented this time in order to better serve the above stated goal of the study. In the remainder of this chapter, we provide a succinct description of this revised approach, including the process and rationale behind the production of our new scenarios. We will then describe our engagement with prospective users. Finally, we present the findings of our study.

6.1 Implementation

Much like in our previous user study, we have produced a set of new animated video scenarios depicting a range of everyday situations in the autonomous home of the future. The scenarios follow a fictional character named Donald as he carries out mundane domestic activities while living in a home filled with hypothetical autonomous solutions. Once again, we sought to frame the depicted hypothetical autonomous systems around household activities that every participant could reasonably relate to, such as grocery shopping or food management.

Unlike in our previous two user studies, this time we refrained from conveying a symmetrical dualistic vision (e.g. utopia and dystopia). Instead, we produced three distinct scenarios, each postulating a different form of coexistence between domestic autonomous systems and their users. Neither of these scenarios was explicitly positive or negative. Rather, each vision presented its own set of pros and cons which were based on our findings from the two initial user studies. Instead of putting forward polar extremes, which the study participants would then simply reconcile by advocating for a middle ground, this approach allowed us to convey a more nuanced experience, prompting participants for deeper reflection and problematization of the issues at hand. Below, we offer a brief description of the three scenarios:

- **Scenario 5** - Donald lives in a fully automated smart home. Food items in the fridge and the bin are, for instance, constantly being monitored by cameras and other sensing technologies. As soon as an item gets depleted, the system orders a new one which is then delivered by a drone straight into the fridge. Donald's domestic routines are thus fully automated and processed in the background without requiring any of his attention. He never has to worry about any of the mundane tasks, such as cooking or going out to do some shopping. Thanks to his system, Donald lives a trouble-free life and has a vast amount of free time which he can spend as he likes. On the other hand, his

reliance on the system contributes to a rather monotonous lifestyle devoid of much variety, while also diminishing much of his privacy.

- **Scenario 6** - Donald lives in a semi-automatic smart home. His system keeps track of upcoming events by scanning his various social media channels and provides him with notifications as well as on-demand recommendations. Any particular recommendation, such as a meal suggestion, must however be reviewed and authorized by Donald himself manually before it gets implemented by the system. One morning Donald receives a notification reminding him that two of his friends - Paul and Paula - are coming for lunch. He orders his smart home system to scan Paul and Paula's social media accounts in order to produce recommendations of suitable meals. The system recommends purchasing hamburgers and fries for both of them. Donald, however, happens to know that Paula is a vegetarian. He therefore overrides the system recommendation and orders a carrot salad for her instead.
  Each recommendation is visibly linked to a source (e.g. a person's social media activity), making Donald explicitly aware of the grounds upon which it is being proposed. The lack of conventional monitoring and data collection technologies results in a situation where the autonomous system produces less precise assessments and sometimes erroneous recommendations (e.g. proposing an unsuitable meal). Consequently, reviewing recommendations requires time and effort on behalf of the user. Moreover, the retrieval and processing of data from social media channels means that users are frequently made explicitly aware of each-other's activity, such as eating habits, which might undermine their privacy.

- **Scenario 7** - Donald lives in a smart home with minimal proactive features. When in need of refilling his fridge, he has to compile a shopping list manually using a smart home app. Donald's previous shopping history is

> tracked and used by the system to provide him with recommendations of items he might want to add to his next shopping list. The recommendations are provided through a variety of mechanisms, such as a voice-based interface, smart-phone messages or visual notifications on his TV. Every recommendation is labelled with a "confidence level", providing Donald with an indication of how confident the system is about the accuracy of the given recommendation. Other factors affecting the recommended food items are variables such as price or the vendor. The depicted system is thus merely helping out by processing information that is already available to the user. Besides being time-consuming, and offering only limited assistance, the design of Donald's system means that he also has to cook all his food manually, which results in less spare time.

The scenarios were initially sketched out in the form of graphic storyboards (see figure 2 for examples), before undergoing several rounds of iterative refinements based on consultation with field experts from our laboratory who helped us make sure the depicted solutions did not drift too far from reality. Once consensus regarding the scenario content was reached, the storyboards were converted into animated cartoon movies[3]. This time we strived to create a more high-fidelity depiction of the future technologies and social settings, featuring richer visualizations. Unlike the 2D cartoonish illustrations used in our second user study, this time we utilized stylized 3D graphics to convey the visions. (see figure 3 for examples).

[3] Scenario 5 is available online under the following link: https://youtu.be/4tUWGJxkx-U
Scenario 6 is available online under the following link: https://youtu.be/6uo0Z-AMwD8
Scenario 7 is available online under the following link: https://youtu.be/Qu-TPGVBju0

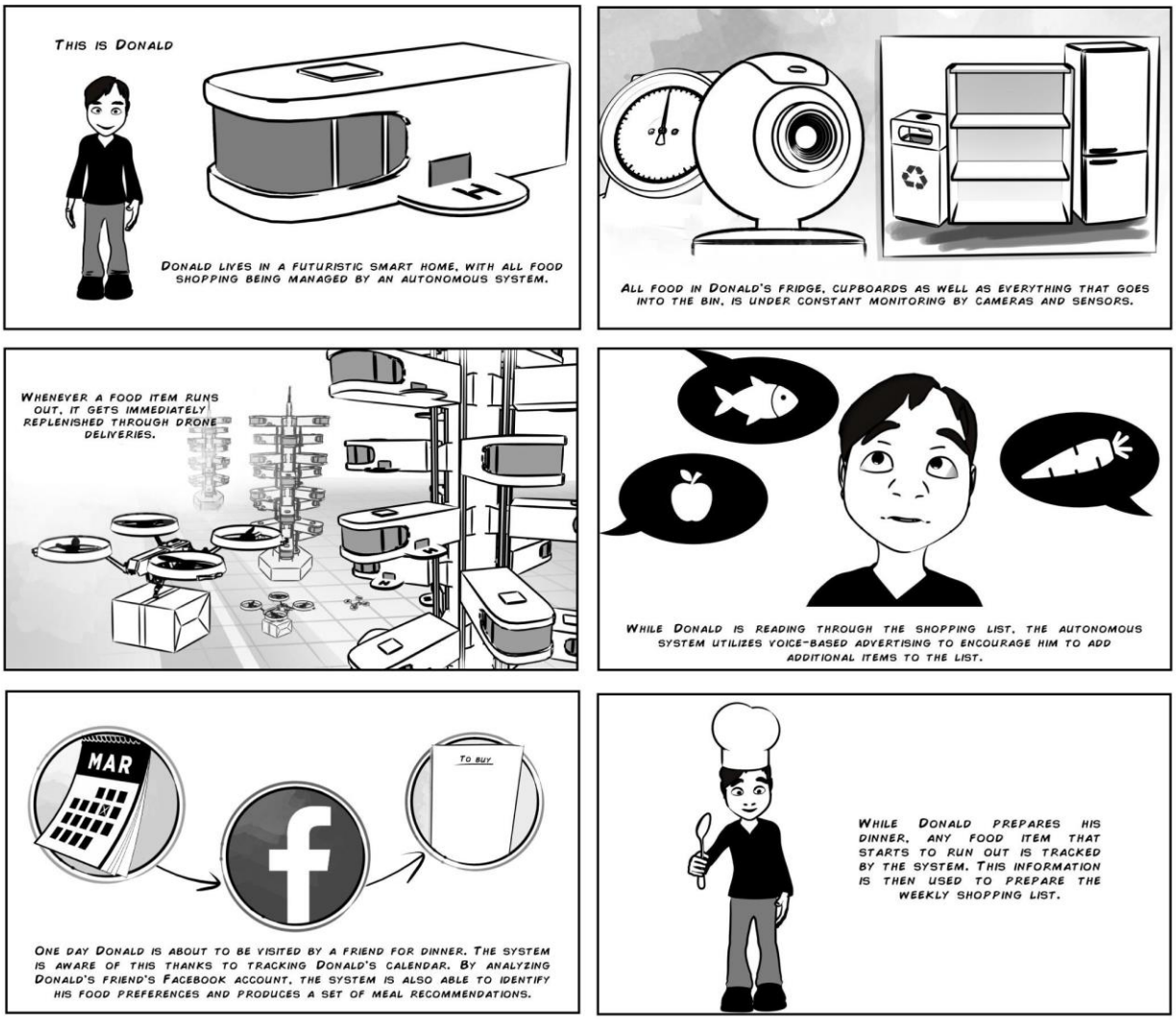


Figure 2 - Collage of graphic storyboards depicting the early version of our scenarios

Thus designed scenarios formed the focal point of our engagement with prospective users.

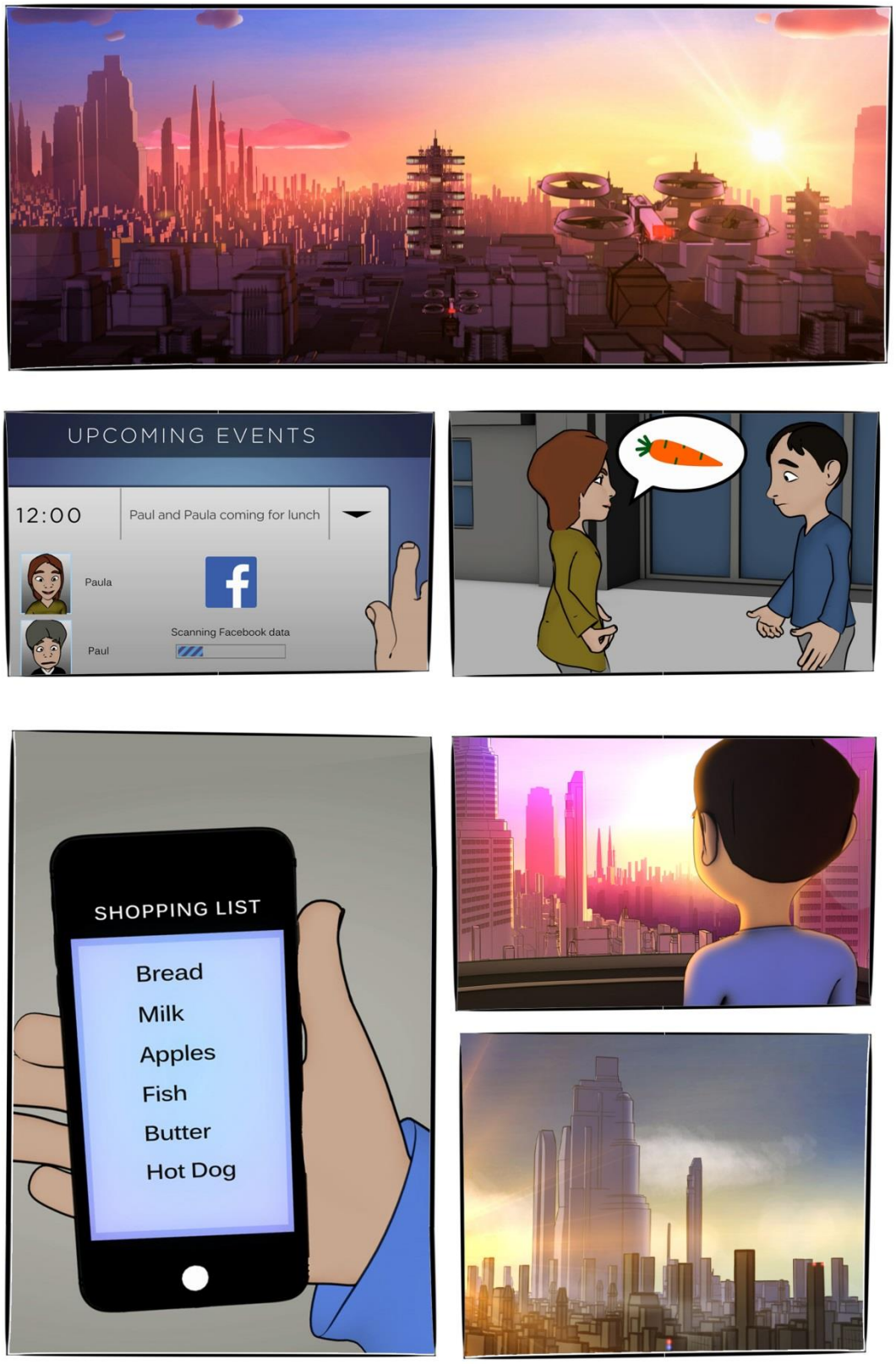


Figure 3 - Collage of screenshots from the video scenarios developed for our third user study

6.2 User Engagement

Unlike in our two initial user studies, this time we adopted a semi-structured interview approach to engage prospective users. Below we provide the rationale for this methodological position along with details concerning the participant recruitment procedure and the organization of the interview sessions.

6.2.1 Participants

Upon obtaining ethics approval from the university's ethics committee, a total number of 20 people were recruited to take part in the study across 16 interview sessions (see table 6). Out of practical reasons, three of the interview sessions (12, 13 and 15) featured multiple participants who were interviewed together. In these cases, great care was taken to provide each of the participants with balanced time to express themselves, making sure everyone would receive sufficient room to answer our key questions.

The participants were a mixed bag of students, retirees, academics as well as people from the industry, both male and female (see table 5). All participants were recruited through email and social media advertising. Before signing up to the study, all participants were informed about the study details and given assurance about ethical principles, such as anonymity and confidentiality. All participants were free to withdraw from the study at any time.

The majority of the interview sessions took place at the university in a quiet office space, devoid of any distractions. Three of the sessions (11, 14 and 15) were carried out in the homes of the given participants in order to better meet their needs (e.g. a tight time schedule). The length of each interview session varied, as we sought to provide each participant with as much (or as little) time for reflection and elaboration of their positions as they required. The typical duration of one-on-one interview sessions was approximately 40 minutes. The sessions involving multiple

participants typically lasted around 80 minutes. Upon concluding the study, each participant was debriefed and incentivized with a £15 Amazon gift voucher.

Group	Name	Age	Background
1	John	20's	Student
2	Emilia	40's	Psychologist
3	Max	50's	Psychologist
4	Hilda	40's	Manager
5	Bertha	50's	Healthcare professional
6	Anne	40's	Psychologist
7	Natasha	30's	Home-based
8	Miranda	40's	Manager
9	Veronica	30's	Biologist
10	Jill	70's	Retired
11	Francis	60's	Graphics designer
12	Markus	30's	Student
	Oliver	20's	Student
	Samantha	20's	Student
13	Liam	30's	Student, philosophy
	Nancy	20's	Student, law
14	Oscar	20's	Student
15	Matt	20's	Student
	Adam	20's	Student management
16	Paula	50's	Public-health professional

Table 6 - Study 3 participant overview

6.2.2 Interview sessions

Several factors contributed towards our decision to adopt an interview approach rather than centring our final study around focus groups. Above all, due to providing the interviewee with greater control over the direction of the enquiry, semi structured interviews have been described as an ideal approach in situations where significant 'depth' is required (Gill et al., 2008). Moreover, interviews have also been noted for helping address any peer pressure that might arise in a focus group session, thus allowing participants to answer questions more openly (Morgan, 1997). Similarly, Harrison et al. (2015) suggests that interviews constitute a preferable approach in situations where *strategic* group biases are anticipated (such as participants claiming an overly-sustainable or egalitarian position).

Each interview session consisted of sequential screening of the three scenarios, each being followed by a discussion. Whereas our two initial user studies adopted a relatively liberal approach, with participants being given considerable amount of freedom to steer the debates into whichever direction they preferred, here we sought to introduce a degree of standardization across the sessions. Specifically, the screening of each scenario was followed by a semi-structured interview centred around 5 verbally administered key questions. The opening question was made deliberately generic and easy for the participant to answer. Such an approach has been suggested to help put the respondents at ease, build up confidence and rapport and potentially even generate rich data that subsequently develop the interview further (Gill et al., 2008). This opening question was then followed by 4 other questions, each alluding to one of our key problem areas surfaced and elaborated through previous user studies. Each question was open-ended (i.e., it required more than a yes/no answer), neutral, sensitive and understandable in order to maximize its accessibility to the participants (Britten, 1999). Below, we outline the key interview questions associated with each of the three scenarios:

- **Scenario 5**
  - *What do you think of this future world? (generic opening question)*
  - *What do you think are the consequences of Donald's food being reordered in this manner? (Coordination)*
  - *Do you believe such a system would be responsive to your preferences? Why? (Control)*
  - *Is it clear to you how Donald gets his food in this future world? How do you think the system works? (Computational accountability)*
  - *Would you find this kind of technology acceptable in your own life? Why? (Social accountability)*

- **Scenario 6**
  - *Compared to the previous scenario, which one do you find more acceptable/useful? Why? (Generic opening question)*
  - *Would you find it inconvenient to have to change or override system recommendations? Why? (Coordination)*
  - *How important is it for you to be able to intervene and change system decisions? (Control)*
  - *How do you think the system came up with the meal plan? (Computational accountability)*
  - *How do you feel about systems like these using your own and your friend's personal data? (Social accountability)*

- **Scenario 7**
  - *How do you see the acceptability and usefulness of these mechanisms compared to those in the previous two scenarios? (generic opening question)*
  - *Do you agree that buying every item at the cheapest available price should be the key system priority? Why? (Coordination)*
  - *Should the user authorize every food order? Why? (Control)*

- *How do you think the shopping list was constructed? (Computational accountability)*
- *Would it be appropriate for a system to be learning about people's behaviour and then use this understanding to help them shop? Why? (Social accountability)*

The key questions outlined above were deliberately broad, designed to elicit answers of a sufficient scope to enable follow-up questions. Such follow-up questions were typically used to prompt the participant to pursue an idea or response in more detail.

A key strength of such a semi-structured interview approach, particularly when compared to structured interviews, is its flexibility, allowing the researcher to seek deeper knowledge regarding a preconceived theme, while also enabling discovery or elaboration of information that is important to participants but may not have previously been thought of as pertinent by the research team (Gill et al., 2008).

All interview sessions were recorded on audio and subsequently transcribed. The transcriptions were used to analyse the thematic concerns that populated our participant's mundane reasoning about the depicted domestic autonomous systems. Below we explicate the different orders of mundane reasoning found during our study, along with the thematic concerns that characterise them.

6.3 Coordination

Much like in our previous user studies, when debating matters concerning coordination, the reasoning of our participants revolved around the feasibility of the various mechanisms by which the solutions featured in our scenarios geared-in with everyday life, along with their potential impacts on the domestic routines.

Cameras monitoring the availability of food items in the user's fridge, recommending applications scanning people's social media accounts, automatic tracking of shopping habits and direct voice-based inquiries were but a few of the means through which the hypothetical systems postulated by our scenarios coordinated themselves with their users in an attempt to seamlessly mend in with the everyday life.

Each of these coordination approaches comes with its own set of affordances as well as limitations. While we came across a number of cases where our participants preferred solutions operating fully autonomously in the background, without requiring the user's direct attention, there were equally many cases where this was seen as undesirable for a range of different reasons. Our participants, in other words, made it clear that the ideal form of coordination would be heavily dependent on the context. As we shall see below, the personal values invoked by our participants as they argued their case hold high relevance for the design of future autonomous solutions.

6.3.1 Reasoning about machine initiative

One of the key appeals of autonomous technology resides in the notion of machines carrying out mundane and repetitive tasks on our behalf, thus allowing us to spend our valuable time on more meaningful activities. While various permutations of this basic argument were reiterated by the majority of our participants throughout the study, as the argument unfolded, it was also becoming increasingly clear that a desirable form of human-machine coordination cannot always be achieved solely by an autonomous system's own initiative.

> ***Researcher:*** *So you feel like a system autonomously reordering basic items for you would be fine?*
>
> ***Miranda:*** *Yes. It would be of help. It would be helpful to always have milk in my fridge without me finding out... when I go and pour myself some*

cereals and find out I've run out of milk... when I want it, that's when I find out I don't have it. That's annoying. It happened this morning here in the office.

Researcher: *Really?*

Miranda: *Yeah, we went to take coffee and there was nothing in the fridge. Monday morning! [...] When the scenario [5] started, before I saw the food consumption, I thought this was going in the direction of… do you remember the Jetsons? Did the Jetsons cartoon make it in England, or Europe? It was about a futuristic family... It's 30 years old… and it was just like this video, and it is 30 years old, imagine! And everything was like... you woke up, you pressed a button and a robot came and dressed you, opened up the blinds, let the dog out, walked the dog. Those things I do take pleasure in. I take pleasure in doing those things and it would affect my creative life because I do like doing those things. Buying food is not one of my creative outlets. I do take pleasure in opening up my blinds and seeing what the weather is, or doing my morning routine. I take pleasure in doing that. I would not want a machine doing that or that being done for me. The mundane things… the things I consider mundane... of course I would like that being done. Stocking the fridge is mundane.*

Researcher: *So you feel like some work… some level of inconvenience actually increases the quality of our lives, if I put it that way? Like walking the dog, doing some labour... you would not want a perfectly 100% convenient life where everything is taken care of for you?*

Miranda: *No. So in that example, I enjoy going outdoors and taking a walk. As a matter of fact my dog makes me do that. Without my dog I would not do that. I enjoy nature, I enjoy being surprised every day. And I enjoy the labour of some things. Ironing? Washing clothes? That labour I don't like doing. And my labour… what I enjoy for labours would conflict with other people.*

As Miranda explains, enjoying a considerable amount of domestic automation, as postulated by scenario 5, would indeed have the potential to help out with her daily routine. The myriad of everyday variables that we all need to keep track of, such as the quantity of various food ingredients in our fridge, inevitably results in occasional human errors, such as forgetting to reorder milk. An autonomous system keeping track of such everyday essentials and making sure depleted items are being restocked would in a sense offload some of our responsibilities, making our life more convenient whilst also eliminating some of the friction that we regularly have to deal with, including "annoying" surprises while preparing breakfast on a Monday morning.

Autonomous systems designed to calmly operate in the background by gathering relevant data and acting proactively to maximize their user's convenience were not just endorsed by our participants as viable solutions for reordering everyday essentials, such as milk, soap or toothpaste, but also for other routine tasks including handling the laundry or providing music recommendations. Part of this reasoning appeared to stem from the belief that automating mundane repetitive activities in a manner that would be coordinated with the user's routine is technically feasible. In other words, the simplistic nature of these tasks made our participants feel like a computer system could be trusted to carry them out reliably on a consistent basis.

Superficial reflections like these no doubt paint an idealized vision of a problem-free coexistence between people and their future autonomous systems. As the vignette above exemplifies, however, upon a closer thought it became evident to most of our participants that not all mundane activities can be treated the same and that the viability of proactive coordination mechanisms depends on factors stretching far beyond simple technical feasibility.

There is an often fine line separating the daily work tasks that represent genuinely tedious chore from those that might provide a source of enjoyment or implicate some form of creative outlet. Hence opening the blinds and carrying out the morning routine is something Miranda "takes pleasure in" doing herself, while reordering groceries, ironing or washing clothes perhaps not so much. A key

challenge for any fully autonomous system seeking to gear in with the everyday life would thus revolve around correctly assessing which work tasks its users prefer to carry out manually. This problem is further exacerbated by the disparity of people's preferences, for as Miranda notes, the particular work that she wants to remain in control of "would conflict with other people".

Indeed, several of our participants described grocery shopping, for instance, as something they value doing on their own. Discovering new special offers or experimenting by trying out new food items were mentioned amongst the factors turning this work into a rewarding and enjoyable experience while also contributing towards a feeling of accomplishment and self-fulfilment. Other participants stressed their appreciation of the social interactions and community-building feeling that comes with visiting the local supermarket.

This suggests then that achieving coordination between fully proactive systems and their users is at its heart not a problem of technical feasibility, but rather of human idiosyncrasies. To make matters worse, even if such idiosyncrasies were to be respected and only those mundane tasks seen by the user as unpleasant would be automated, once considering such a practice in longer-term, our participants were still able to find numerous points of friction.

> ***Max:*** *It obviously just ordered the same thing again. And the methods seemed to be a bit of an overkill… to determine that it [a bottle of milk] was empty. And obviously, these are reasons why you might want to replace your milk that has gone off. But it gets replaced by the same thing. Maybe you want an orange juice or apple juice next time. That's why I'm saying it might be a bit limiting. I've done online shopping, but then you have all the choices, things are not pushed at you... In the video [scenario 5] the person is more like a farm animal than a person.*
>
> ***Francis:*** *There is no planning in there at all. He's got what he's got, he uses it and then the new one just replaces it. So really, it takes away the need to plan, to organize etcetera. And I think that's a bad thing. I think as*

> *a species we're getting very short planning time and attention time, I think that [autonomous systems] is adding to it.*
>
> ***Liam:*** *Everyone might change preference and behaviours. There should be mechanisms for that in the system. It should have reminders like "do you really want to have more of that?" or "you might want to try something different?" that sort of stuff. Like what if you wake up one morning and you are a completely different person. It can happen, people can decide to make a significant change in their life.*

Automatic provision of services is not only problematic due to the often idiosyncratic preferences surrounding mundane domestic work, but likewise due to the changing nature of these preferences. Systems basing such provision on mere extrapolations from the user's previous consumption habits would result in a state of inertia with the same items being reordered repeatedly. Yet, as evident from the extracts above, many prospective users view a varying and diverse lifestyle as a crucial part of a positive experience.

A system that, for instance, correctly infers a bottle of milk being depleted and successfully replaces it with a new one will still end up breaching the expectancies of its users if they were intending to instead purchase a different beverage. In such a situation the autonomous system ceases to be a helpful domestic agent and instead becomes a form of limiting mechanism constraining our freedom of choice. As Max fittingly points out, a user who is merely passively accepting the goods and services that are being provided is in effect "more like a farm animal than a person".

A similar line of argument is followed by Francis, who claims that a missing element of human long-term planning represents another key factor. The distinction between short-term and long-term automation of tasks thus seems to be of significance here. Indeed, if persistently repeated over a long time, even a trivial action, such as the ordering of milk, can have a powerful impact on our lifestyle.

As suggested by Liam, even a fully autonomous system should therefore be prevented from extending a particular pattern of actions over a longer time frame without to some extent involving the user. Instead of basing actions on mere extrapolations from previous consumption habits, autonomous systems thus need to offer channels for human input, granting its users the means to alter the self-perpetuating cycle and align it with their own long term goals.

A system becoming a helpful agent in the intricate landscape of mundane household work, while at the same time also respecting the largely un-inferable aspects of user's personal preferences, is difficult to accomplish through machine initiative alone. As elaborated throughout this section, such a one-sided coordination would run the risk of inadvertently reducing user's capacity of creative expression, enjoyment, freedom of choice and ultimately - if extended over a longer term - erode a sense of human autonomy.

6.3.2 Reasoning about human autonomy

The varying degree of human agency represented one of the chief contravision points that set our three scenarios apart from each other. Whereas scenario 5 depicted a system reordering food items proactively and automatically without any direct user input, scenario 6 revolved around a system which automatically compiled a list of meals to be purchased but required explicit user authorization before committing to the purchase. Finally, scenario 7 had the user construct shopping lists manually while being provided assistance in the form of recommendations generated by the smart home system.

With the exception of the most basic household chores, most participants tended to reject the ideas put forth by our first scenario, frequently citing the need for greater control over one's life as a prime reason. Each time participants expressed such views, we generally sought to challenge them by theorizing that autonomous systems could produce better assessments of a range of factors (such as the user's

health) and consequently by allowing the system to be in charge, we could for instance achieve a healthier lifestyle. As demonstrated by the following vignette, arguments such as these did little to sway the views of our participants who maintained that it is the user who should always have the last say.

> ***Researcher:*** *So what if someone would be eating really unhealthy food, just buying pizza all the time, would it then be OK for the system to step in and say "I'm not going to allow you to buy another pizza, you will have vegetables instead" or something like that?*
>
> ***Jill:*** *Umm… maybe not so much make that decision, but maybe offer alternatives.*
>
> ***Researcher:*** *So the user should still have the option to choose to ignore those recommendations and keep ordering food that is unhealthy?*
>
> ***Jill:*** *Oh yes I think so. But it would be really good if the system said "oh you had pizza yesterday and the day before. Why don't you try this?" You know, and perhaps offer a more healthy option. That would be interesting, wouldn't it?*
>
> ***Researcher:*** *So would it be correct to say that you would like your system to inform you, rather than control you?*
>
> ***Jill:*** *Yes, I do. Yes.*

Jill makes it repeatedly clear that no autonomous system should exercise direct control over its users, regardless of how good its intentions are. If users choose to make decisions to their own detriment, then so be it. Ultimate responsibility over one's own life is still something that must not be delegated. That is not to say that computational automation is always bad and inherently in conflict with human autonomy. As Jill points out, there are other more acceptable ways in which its capabilities might be put to good use. Recommendations, such as suggestions of

healthy alternatives to items being purchased by the user, represent one specific example that "would be interesting".

Other examples of contextually appropriate information provision proposed by our participants included the use-by dates of food items or details of their provenance (one participant for instance expressed the preference for free roaming eggs), which in turn would help users make informed decisions when shopping. Such information does not necessarily have to be provided in the form of explicit recommendations. As Liam explains in the following vignette, a system that keeps its users informed, without urging them to take any particular course of action, might often be an ideal design goal to strive for.

> ***Liam:*** *I think it will come down to agency as well, in terms of like… yeah, you might want to create bad behaviour in people. But there's a point where things will kinda standard you, and you see that already with kids that got mobile phones and it takes their agency away, and agency is related to creativity and so on, and then you've got these things again like stratifications. Like if I only ever do what I'm told by systems and organizations and other people, I probably would not be here doing my PhD. Because there is a certain amount of like… you need to inspire agency and decision-making of people and allow them for freedom. Because you might not get them going to further things... it's like a deeper philosophical thing for me, in terms of like... if you are just getting suggestions. Which is why I kinda like the interface where it allowed you to change your mind and get carrots instead [scenario 6].*

Recommendations, according to Liam, could easily end up "standardizing" the thinking of their recipients, impairing their ability to "further things" and to seek solutions with an open mind. Rather than handhold the user through each decision making process in the name of greater convenience, autonomous systems ought to help "inspire agency and decision making of people and allow them for freedom". The vignette above indicates that our sixth scenario did strike a particular chord

with such a line of thought, given its focus on the user's ability to override system decisions by virtue of their own independent thinking.

The user expectancies described throughout this section make it apparent that the problem of human-machine coordination revolves to a great degree around converging on a point of balance between computational agency and human autonomy. Our participant's reflections suggest that one potential way of reaching such equilibrium is by designing systems that merely provide their users with grounds for informed and conscious decision making, while the decision itself is left to the user. Instead of rendering the user into a passive receiver of services, a well-designed mechanism for information provision can thus help stimulate independent thinking and inspire into action. Such a state of coordination is dependent on a more direct form of interaction, with the users engaging in dialogue with their systems, rather than having them operate in the background. In the final coordination subsection we will take a brief look at some of the difficulties that might arise from such interactions.

6.3.3 Reasoning about the human in the loop

Just as in our previous two user studies, a particular strand of coordination reflected upon by our participants had to do with the prospective techniques employed to align the operation of autonomous systems with the daily routines of their users. As demonstrated throughout this study, proactive interference does not always represent a viable approach, not just due to the idiosyncrasies surrounding our mundane behaviour, but likewise due to the risk of infringing on human autonomy. On the other hand, relying entirely on user's manual control would render much of the benefits brought about by AI mute. Unsurprisingly, when asked directly about their ideal coordination mechanism, our participants thus generally sought to find a suitable combination of these two approaches.

> ***Researcher:*** *But this information about you, do you have to input it manually into the system or can it obtain it in a different manner?*
>
> ***Marcus:*** *It could be a mixture between obtaining it through all the information it's already collecting. And then also through manual input of say… you could put a big long questionnaire when it's installed or whatever when it is installed and then from there it could set the initial values and parameters, and then it could go on from there.*
>
> ***Oliver:*** *Calibration is a process, isn't it? So you have to bear in mind… my GPS sensor on my phone required recalibrating, so I had to just walk around to calibrate the GPS. And it would be similar to this.*
>
> ***Samantha:*** *I think people are quite accustomed to providing preferences. Like if you buy a new phone for example, you spend half a day just setting it up to the way you want it to work or how you want information provided. So it's not like we don't already do that.*

In other words, going through an initial stage of 'calibration', requiring a considerable amount of manual input and customization from the user, would be acceptable. As Samantha notes, when it comes to technologies, we are already accustomed to "setting it up to the way we want it to work or how we want information provided". As the extract above illustrates, participants generally argued that such an incubation phase could then be followed up by greater machine autonomy. In line with Oliver's reasoning, the majority of participants also agreed that even a calibrated autonomous system would however also need to have a mechanism in place allowing for a more continuous 'recalibration' and coordination. Amongst such mechanisms postulated by our scenarios were voice enquiries, as well as text messages, prompting the user for various forms of input. Once again, the timeliness and contextual appropriateness of such proactive enquiries proved to be of vital importance:

> ***Researcher:*** *Some people tend to hate messages popping up proactively and stuff.*
>
> ***Hilda:*** *Yeah, I'm not sure. I think if I were to have it, I would want to choose when I… you know when I was at a time and place where… if I was travelling or something… and I thought "OK I'll look at these things now". I would not want them to pop up without me going into the app. I don't want them to intrude into my life by popping up any time of the day when I have not chosen it.*
>
> ***Researcher:*** *Although this kind of already happens when you're browsing the internet. There's a lot of advertisements.*
>
> ***Hilda:*** *Yeah and I don't like it.*
>
> ***Researcher:*** *So even if these messages popping up on your smartphone would actually be useful, you would still not like them?*
>
> ***Hilda:*** *No, I would want to choose when I saw that information. I would not like the app to tell me things like "Oh, do you want to add chicken to your order" or "oh, your delivery will come tomorrow at 5 pm". I would not like that.*

A poorly timed enquiry addressed to the user could easily end up feeling intrusive or socially awkward, with negative implications for the acceptability of the autonomous system as a whole. Hilda makes perspicuous that such a risk can only be alleviated by allowing the user to choose precisely when and where any interaction takes place. In other words, the process through which users customize and coordinate their autonomous system must itself be subject to user customization. As we shall see in the following section, this in turn often requires a more fine grained user control.

6.4 Control

The study findings presented throughout our user studies helped surface a number of red lines standing in the way of a frictionless functioning of autonomous systems. The risk of a system misinterpreting everyday contingencies, user's suspicion of ulterior motives or various infringements on personal freedom and the ensuing lack of trust are but a few of the potential problems reported by our participants. The ability to step in and assume direct control over the system was in this context seen as essential in order to allow users to safeguard their interests by correcting any potential missteps made by the system.

The expectancies surfaced during this study did in this sense not deviate meaningfully from those brought up through our previous enquiries, but rather went into greater detail regarding the granular aspects of control, with the various aspects of AI behaviour that need to be subject to human control in order to preserve a sense of human autonomy.

6.4.1 Reasoning about customization

The customizability of autonomous system behaviour emerged as a key theme pertaining to control in our initial user study. Here we dig deeper and look into its nature and significance.

> ***Researcher:*** *So imagine a spectrum where at one end you have a system that does everything for the user and the user is enjoying maximum convenience, and at the other end of the spectrum the user has to do everything manually and enjoys a bit less convenience. Where would you position yourself generally? Somewhere in the middle?*
>
> ***Marcus:*** *No both. I would want everything to be automated, but I would also like to be able to manually control everything if I needed. So I would go*

> *into the settings and change everything. But I would like the potential to have it do everything for me. So I think it's…*
>
> ***Researcher:*** *So you would like to have an on/off button to turn off all the automatic functionalities if you wanted to?*
>
> ***Marcus:*** *Ummm… yeah, but I would also like to have manual control over all settings, so I could specify the settings to be right for me. So for example if you've got settings, you can go to 'advanced settings', and I would like to be able to do the 'advanced settings' but that would not compromise how much automation it does, you would just be able to customize the automation. So full control over what you could automate, but a potential for everything to be automated.*

Much like in our previous user studies, participants made it clear that an acceptable autonomous system would have to allow its user to intervene with its behaviour at any time in order to customize any facet of its operation. Such a level of control would not necessarily have to take on the form of complex interaction interfaces. Rather, a simple on/off switch was brought up as a generally sufficient mechanism for customization, as long as it would be applicable on the level of each individual system feature.

A large number of examples for such features were provided in the course of our third user study. Multiple participants, for instance, argued that the user should be able to turn off proactive provision of recommendations due to fear of being swamped with advertisements. Being able to disable voice based interfaces in favour of other more discreet communication modalities was likewise a commonly mentioned requirement, mainly due to voice being seen as inappropriate during various domestic situations. Outward flows of personal data represented another aspect of system behaviour that participants generally agreed the user should be able to shut down in order to preserve the user's privacy.

It was not just the system itself that participants argued a user should have control over. Being able to disable provision of services from companies that the user does not trust was likewise brought up as being essential.

Given the breadth and complexity of the autonomous systems depicted in our scenarios, one might easily reach the conclusion that providing the user with control over all of the system's features could end up overwhelming the user with too much choice. The effort required to configure and assimilate the operation of such a system to its user's liking could turn out to be so substantial that it would outweigh any benefits resulting from delegating human labour to the system.

When prompted to elaborate how the user ought to cope with the hassle stemming from excessive amounts of control, participants typically responded by postulating various workaround strategies that would make such a level of control more manageable, such as providing the user with a "recommended settings" option or various pre-set profiles.

However, as evident from the vignette above, the majority of our participants eventually reached the conclusion that system automation and human control are not mutually exclusive. When Marcus reasons that "I would want everything to be automated, but I would also like to be able to manually control everything if I needed", he makes it clear that providing a user with means to control and customize system operation does not mean the user has to use these to their full extent. The presence of control mechanisms in this sense serves just as much as a tool for gaining the user's trust, as it is a tool for practical utility. Provided that such a mechanism would be in place, full system automation would consequently per se not undermine human autonomy.

6.4.2 Reasoning about human autonomy

As evident from the previous section, our participants saw it as imperative for the human user to always have the upper hand over its autonomous system. An

argument frequently discussed in relation to scenario 6 was that a system can, in spite of having the best intentions, make erroneous decisions (e.g. recommend the ordering of meat for a vegetarian). Human users, given their presumably better ability to assess mundane variables and situational contingencies, should thus be able to exercise enough control over the system to correct for such errors. As the vignette below exemplifies, however, this principle is not universally applicable.

> ***Researcher:*** *So what if like a 6-years old kid would use the system to order lots of alcohol and stuff like that?*
>
> ***Hilda:*** *I would not let my 6-year old child anywhere near it. It would theoretically not be allowed to… obviously it would have to be password protected in some way, so they would know that it is me using it and not some child. Because presumably a child should not be allowed to have that kind of system.*
>
> ***Researcher:*** *So in your household, only parents should have access to a system like this?*
>
> ***Hilda:*** *Yeah, I mean I would hopefully never have that system, but if I did, I would lock it down so that a child would not be able to use it... or use it unsupervised.*
>
> ***Researcher:*** *But adults should have the freedom to order whatever they want? Let's say you were eating like really unhealthily, like loads of pizza, would you be fine with the system intervening and telling you "I'm not going to allow you to buy another pizza, you should eat more vegetables!"*
>
> ***Hilda:*** *Absolutely not! [laughing] Because again it's about the autonomy. And how do they know the context, what if you have a party and you've ordered six pizza's because you have lots of guests coming over. The system might not know that, it might assume that you've eaten six pizzas all yourself. But it's all of the context, isn't it? Or what if you were told by your doctor that you are underweight and you needed to eat more fatty food*

and stuff to gain weight. What if… it's all this context that the algorithm would not necessarily know.

Researcher: *So what if it was your doctor on the other side pulling the strings and telling you you're not allowed to buy another pizza. Would that make it more acceptable?*

Hilda: *Oh that's a tricky one. Umm… no I don't think so, I don't think that should be allowed because… yeah that's a really interesting question. I don't think so because I think a doctor… shouldn't… like you still have free will, you still have autonomy and the doctor can advise "I think you're eating too much pizza, I think you need to cut down a bit", but I don't think they have any right to actually stop the order. Unless it's an actual threat to your life. Like if you eat another pizza, you're going to die. OK, that might be an extreme situation but if you're overweight and you're ordering another pizza, it's ideal to have that support to help not eat pizza and help to reconsider their ways and think of more healthier food… but I don't think they should have that right taken away. And this is a problem that I think could end up moving into something a lot more… I'm not sure what the right word is, not 'regulated', but… who ends up being in control. You know, this thing that you think is convenient, having your next bottle of milk ordered… but then who ends up being in control if this information gets sent to your doctor and then stays dark while your being monitored and that information about you gets shared, that's when it all becomes very problematic for me.*

Researcher: *So if someone wants to eat unhealthy food and ruin their health, they should be free to do so?*

Hilda: *Yeah.*

It would be too easy to simply conclude that making sure the user always has the last say, ought to be a guiding principle for design of domestic autonomous systems.

As Hilda makes perspicuous, in a real world domestic setting, control is scarcely evenly distributed between the individual household members. Parents, for example, are generally expected to enjoy greater control over the household than their children. Societal laws are likewise constraining what can and cannot be done. By ignoring such norms and granting each user unlimited control, an autonomous system could inadvertently upset the household order in a detrimental manner.

Whereas our participants were in agreement that an autonomous system should not determine on its own to withdraw control from any of its users, they were thus generally in favour of using the system to protect - and if necessary enforce - the existing hierarchies of the home, along with societal laws. Designing a mechanism that would grant users a suitable level of control over various features of an autonomous system is in this sense just as much an ethical and social question as it is a practical one.

Whereas overriding user's decision could in this sense be seen as legitimate, Hilda makes it clear that such situations are few and as a rule of thumb, human autonomy should always be upheld. For any constraint of human control to be acceptable, it is imperative that the grounds upon which such constraint is implemented is made clear to the user (e.g. "this action has been prohibited by your parent"). A system depriving its user of control in a manner that is acceptable is, in other words, preconditioned by a high level of computational accountability.

6.5 Computational Accountability

It has been established through the previous two user studies that a central problem area that future autonomous systems will have to tackle revolves around computational accountability - i.e. the need for legible and intelligible system behaviour. Below we revisit some of the relevant themes that surfaced through our previous two user studies and elaborate them further.

6.5.1 Reasoning about accountability

Although our video scenarios were set in a futuristic world, characterized by a widespread presence of autonomous systems, the work carried out by these hypothetical solutions was mundane and did not markedly deviate from contemporary everyday household activities. The depicted system behaviours were consequently seen by the interviewees as generally relatable and explainable.

> ***Researcher:*** *Did you feel like you understood how the system worked?*
>
> ***Emilia:*** *Yeah, it looks like an internet of things system. The camera had some kind of intelligence that links to the fridge and to the bin and is able to send a message and then a drone is dispatched to restock the fridge… so in terms of the tech, I felt I understood what was happening and I felt the system was quite straightforward.*
>
> ***Samantha:*** *Yeah, it was scanning his friend's Facebook account and preferences and placing online orders. It was quite simple, yeah.*

Participants showed limited interest in the technical details of the depicted system's inner workings, typically describing these using somewhat ambiguous terms such as "some kind of intelligence". What is significant is that this did not prevent them from perceiving the overall system behaviour as "quite simple" or even "straightforward". The particular system behaviours featured in our scenarios, such as the ordering of milk to restock a fridge, the compiling of a shopping list or the tracking of user's calendar for upcoming events all represent mundane and largely self-explanatory patterns of domestic conduct. The intimately familiar nature of such daily routines left our participants with little doubt as to what the depicted systems were doing.

While revealing the inner technical workings of a system is not a necessary or indeed even a desirable design goal to strive for, evidence quickly emerged that it is the

grounds upon which a system makes its decisions and its broader agenda, along with any personal data collection practice implicit in these, that are in need of further explanation.

6.5.2 Reasoning about algorithmic decision making

Whereas our participants were generally able to see and understand the purpose of the depicted system actions, the underlying reasons that caused the system to carry out these actions frequently came across as opaque and difficult to decipher. Such illegibility of the underlying reasons in turn appeared to push many participants towards questioning the accuracy of autonomous decision making. This became particularly apparent when discussing the notion of a system inferring information about users from social media and using those as grounds for meal recommendations.

> ***Marcus:*** *So if you've liked hamburger and fries on Facebook and you publicly say online that you like hamburger and fries, then it will get you hamburger and fries and it will say like "I found it on the social media that you like this".*
>
> ***Oliver:*** *People, like vegetarians, could also say on the social media that they like something ironically, or sarcastically, can't they?*
>
> ***Samantha:*** *Yeah it depends on the data, like how the system figures it out. Like nobody goes on to Facebook and likes a photo of hamburger and fries, it's more often things related to it.*

Human behaviour, in other words, is notoriously difficult to decipher - in part due to various human expressions, such as irony and sarcasm, as argued by Oliver. In the light of such complexity it does indeed seem questionable whether a machine could be trusted to correctly interpret a situation and produce appropriate

recommendations. The particular dataset and knowing *"how the system figures it out"* thus becomes necessary in order for the user to establish whether a recommendation has been based on correct grounds.

Transparency, as expressed here, thus revolves around the need to give users a better clue regarding how their system arrives at its decision, in turn equipping them better to identify and workaround wrong decisions and making them more comfortable accepting and acting on those decision deemed to be correct.

In an effort so somewhat lift the lid on this decision-making process, our third scenario postulated the use of a 'confidence level' indicator attached to each shopping list recommendation. By expressing the degree to which a recommendation can be trusted this mechanism thus served as a warning, making the user explicitly aware if the grounds upon which a recommendation is made appeared particularly fuzzy and ambiguous. While generally seen as a step into the right direction, as illustrated by the following vignette, this alone would not fully address the issue.

> ***Researcher:*** *Do you think that makes a system more acceptable, if it can express lack of confidence?*
>
> ***Max:*** *I think yes, because I think it tells you that it's not fully… it's not just giving you "yes" or "no", it gives you "potentially yes" or "potentially no", because it is something that it thinks about you, but then the question is... how does it derive the confidence estimate?*

Knowing that a system has limited confidence in the correctness of its own recommendation is useful in the sense that it tells the user that heightened caution is required. But what is there to say that the system has correctly assessed this confidence in the first place? Instead of addressing it, our proposed approach thus only appeared to have shifted away the problem slightly. As Max explains, the provision of an account would therefore not only need to consist of self-assessments generated by a system, but also reveal some of the raw input leading up to these

assessments. Explaining just "what" a system bases its decisions on would in other words still be required.

Further details regarding what exactly such an account should consist of surfaced when discussing the assisted shopping list construction depicted in scenario 3.

> ***Researcher:*** *So what about proactive recommendations with the system shouting out things like "Would you like to add tomatoes?" Would you be fine with that?*
>
> ***Veronica:*** *Umm… so what would that be based on? On your own activities? How would the algorithm be able to give that recommendation? So where does the data come from? There would have to be a bit more clarification on how you get to that point where I can say "yes" or "no" to the recommendations. I would need more information.*

Once again, it is apparent that a system's failure to account for the basis upon which a recommendation is produced leaves its users guessing. As exemplified by Veronica, most participants had a vague understanding of machines being able to learn from our activities and subsequently using this knowledge to produce suggestions. But how do we know exactly which of these activities have impacted a particular recommendation? The vignette above hints that revealing the origin of a recommendation, i.e. the collected user data that has been involved in shaping a recommendation, might go a long way towards making recommendations intelligible.

It would however be premature to conclude that making the data that is being fed into a system fully legible should be a design goal to strive for. As other participants elaborated, relatively simple explanations could often get the job done without overwhelming users with information.

> ***Researcher:*** *But if a system would be completely transparent, wouldn't you be bothered by like information overload?*

> ***Liam:*** *I suppose that depends on the level of information. I think if you get like an overview, like a general kinda quick thing of like… even like a sentence… like "you tend to enjoy chickens on Tuesdays", like that seems like a kinda good short summary.*

Explanations of reasons behind system behaviour must be presented as a support mechanism, helping users make a choice or take an action. It is not the algorithmic decision-making process then that needs to be spelled out to the user, but rather the rational informing the user how a particular recommendation was converged upon.

Being able to understand how an individual system decision was made is however only part of the story behind computationally accountable autonomous system behaviour. It is equally important to understand the bigger picture and the overarching motivation that underpins the overall direction of autonomous system behaviour. For that we have to turn towards motives.

6.5.3 Reasoning about motives

The fact that autonomous system actions, such as recommendations, have the power to alter our consumption habits was brought up as a potential liability making these systems susceptive to being exploited by their providers for various ulterior goals.

> ***Researcher:*** *Do you think the system should make its users aware whenever it's altering their behaviour in a certain direction?*
>
> ***Anne:*** *Yeah, I think it should all be very transparent. I would not mind it if I got to choose what I believe in or what I would like it to nudge me into. Like exercise... I'm terrible at going to the gym, so it'd be great if I could choose to make it nudge me towards going to the gym. Now, I can completely ignore it, but if it tries to find ways of reminding me that the gym exists, that would be great.*

As Anne makes perspicuous, being 'nudged' into a certain direction by ones autonomous system is per se not a problem. Indeed it might even be seen as desirable for a system to encourage certain types of behaviour, e.g. pursuing the user to go to the gym and doing more exercise. It is however imperative for any such motive to be "very transparent". The ability of AI to explain itself and account for the motives behind its algorithmic decision-making is in this sense a prerequisite for allowing the user to make a conscious decision whether to act on a recommendation or whether to "completely ignore it".

As the interviews progressed, it became evident that painting the motives of autonomous systems fully legible in the eyes of its users can frequently be a complex matter.

> ***Researcher:*** *Would you trust a system if it would recommend you new things every now and then?*
>
> ***Emilia:*** *This is again a question of transparency, who's behind the system, and intentions of the system. And the ethics behind the system. So if I can engage and trust the system, then I'd be more prone to accepting suggestions. But if obviously there is some obscure element that I don't understand then I would not like it.*

Here Emilia makes it apparent that accountable motives are not simply dependent on a system articulating its "intentions" but also on "who is behind the system" along with the ethical issues bound up in this. A system can thus articulate and explain its behaviour all it wants, but if the users feels like they cannot trust the people behind it, they will regard the true motives of the system unaccountable. The burden of making system behaviour accountable should in this sense not rest solely on the system. Some of it can - and should - be offloaded on to the humans responsible for the system operation. Non-computational factors, such as the

reputation of a system provider, can thus play a vital role in shaping the manner in which a user interprets the motive behind autonomous actions.

As the following vignette makes perspicuous, even trivial information, such as knowing the name of the system provider, can go a long way towards helping the user make an assessment of their system's motives.

> ***Researcher:*** *Let's say the system is assisting you with compiling a shopping list and it gives you some recommendations, would that make you suspect bias?*
>
> ***Miranda:*** *I think I'm pretty good at seeing that. I can see when things like that are biased recommendations.*
>
> ***Researcher:*** *So let's say a system is owned by Amazon and you see lots of Amazon products amongst the recommendations, does that make you feel suspicious?*
>
> ***Miranda:*** *Absolutely, it's almost obvious what they are doing.*

A system articulating the motive for its action (e.g. recommending an Amazon product in order to help the user construct a shopping list) in combination with the user's knowledge of the system provider (e.g. Amazon) are thus two distinct components allowing the user to converge on their own interpretation of the motive (e.g. service provider seeking to maximize its own profit, rather than pursuing the user's best interest), for as Miranda explains, it becomes "almost obvious what they are doing".

Our previous user studies established that motive constitutes a vital part of an account. The data surfaced throughout this study enables us to go a step further. What we see here is that a motive is made legible not just through algorithmic means, but likewise through human means. The former has to do with a system explicating the goals its actions are intended to accomplish, the latter can best be described as the level of trust established between the user and the party responsible

for the system operation. Users must in other words be permitted to question and assess not just their systems but also the organization behind it. This can only be accomplished if both are laid out for the user and made available for scrutiny. Once again, it is evident from the participant statements above that such accounts would need to be provided as a preface to and/or in vivo feature of autonomous behaviour, rather than a post hoc explanation.

6.5.4 Reasoning about personal data flows

A topic closely interrelated with motives is the practice of monitoring and personal data collection. While our participants were generally understanding of the important role these play in the functioning of autonomous systems, transparency once again emerged as a critical condition for such practices to be seen as acceptable.

> ***Researcher:*** *You said it's the collection of personal data that you're worried about, that you would not like too much information to be collected about you?*
>
> ***Veronica:*** *Or not knowing. I mean… if that's needed for a purpose that is in my interest, I would like to have more details, I would like the system to be transparent in a way so that I could understand it, transparent in a language so that the users can understand the things that concern them. About the data, what it's used for, who they are sharing it with and so on. All the things so that the user would feel comfortable with it.*

It is hardly surprising to see that the notion of being watched without knowing about it makes prospective users feel uncomfortable. As exemplified by Veronica, this did not prevent the majority of our participants from seeing that the collection of personal data could potentially serve as a means for accomplishing goals that are in the user's own best interest. Any flow of personal data would however need to be

made accountable to the end users in order to allow them to assess and make sure it is properly aligned with their own preferences. The vignette above illustrates the host of variables constituting an account in this case, including the ownership, use and whereabouts of personal data as well as the owners' sharing policy. Moreover, such an account needs to be provided in an accessible manner, using *"a language so that the users can understand the things that concern them"*.

Just as was the case with motives, general lack of transparency in personal data collection stirs up fears of ulterior motives. As postulated by the following vignette, a system monitoring and collecting data about users' food consumption in order replenish items that are running out could for example easily be exploited by the service providers for their own gain.

> ***Researcher:*** *So what if the system provider would for example sell your food consumption data to the insurance company which would then use this data to determine the price of your life insurance?*
>
> ***Miranda:*** *Well in that case hopefully I would realize this and argue that case, because you can't use... not only what I've done on Facebook or online - you can't even use what I've done in my personal life... against me today. I would fight it, but first I'd have to realize that it has been done.*

There is in other words little that can be done to combat such unethical activities unless the user is equipped with the means to detect their existence. Transparency in this sense becomes a tool enabling users to converge on a preferred balance between personal privacy and service utility by scrutinizing and - if deemed necessary - withdrawing from services whose data collection practice is seen as unacceptable.

So how exactly can such a transparency be achieved? While our scenarios were primarily not concerned with system implementation, a number of comments made by our participants during the study made it clear that an account does not have to be provided digitally by the system interface itself. Rather, simple affordances can go a long way towards communicating the manner in which the collection of personal

data is carried out. Monitoring equipment that is physically constrained to a specific location in the home, such as a static camera pointed at a specific shelf in the fridge, has been brought up as a viable way of making users aware of what exactly is being monitored. Likewise, a range of external factors, such as service agreements were argued to play an important role in achieving sufficient data collection transparency.

6.6 Social Accountability

Our third user study quickly confirmed the recurring pattern we found in the two preceding studies, namely computational accountability taking the backseat as soon as participants begun to elaborate how the depicted technologies would fit in with the broader context of their own daily lives. Instead of stressing the importance of making systems transparent and legible, participants were repeatedly voicing the expectation that an acceptable system solution would also need to operate in a manner that is accountable to the social circumstances it is embedded into.

6.6.1 Reasoning about acceptability

The importance of making autonomous systems socially accountable became evident from numerous participant comments made throughout the study. A frequent common denominator of these comments was the notion that autonomous system behaviour can be perfectly legible and understandable when assessed on its own, but equally come across as nonsensical the moment it is placed within a particular social situation, as made perspicuous through the following monologue by Liam:

> ***Liam:*** *I think it was horrendous! It's like immensely… like if you're getting individual items delivered by drones, I find that problematic somehow, it doesn't seem efficient. I don't like the surveillance, what's going on with his camera watching everything. And it also apparently has*

> *x-ray vision looking into the bin and identifying what's inside there. So what says it's not going to look at you as well?! There's stuff I've recently seen about in-car advertisements and a lot of the data enabling the advertisements is gained from this sort of stuff... businesses want to use it for analytics and I think it's awful! And I also dislike giant megacities and everybody living in tiny cubes. There's very much that… despite of being someone that has a lot of experience with technology, or perhaps because I've got all the experience with technology, I really don't like it and I can see multiple points of failure throughout it. So I'm very much like I don't think I would ever like to be involved with that. And to go on further about it… I don't like the idea of turning us into dumb consumers who just absorb whatever is given to us. So I'm just... looking at the whole thing, for me... I'm just really pessimistic about it. There's a lot of stuff in there that I disliked immensely.*
>
> ***Researcher:*** *So was there anything at all that you did like about the scenario?*
>
> ***Liam:*** *[long pause] … the animations were kinda nice.*

In the light of his previous experience with technology, Liam could see a number of glaring potential problems implicit in our scenarios, all adding up into a vision of the future which he "disliked immensely" and one that he does not think he "would ever like to be involved with". It seems clear that this contempt did not result from the failure of any depicted system to articulate its actions in an intelligible manner, but rather from its failure to act in a way that would be intelligibly appropriate when a host of other contextual variables got factored in. Besides the apparent issues with "surveillance" - which we shall examine more thoroughly later - it appeared questionable, for instance, whether a system centred around the principle of having each food item delivered by a drone immediately after being depleted could really be efficient in a real world setting. Furthermore, the notion of delegating our everyday tasks - such as shopping - to smart home systems was seen as

inappropriate, as argued by Liam, due to turning its users into "dumb consumers who just absorb whatever is given to them".

Similar views that appeared to question the appropriateness of the core idea behind automation were echoed repeatedly by our participants, leading many to adopt a somewhat sceptical stance when confronted with our scenarios. It is perhaps not surprising then that when asked whether they would potentially be willing to adopt any of the depicted systems, many participants were gravitating towards the solution that would have the least upsetting impact on their existing patterns of domestic conduct.

> ***Bertha:*** *It depends on the person. For people who have mobility issues it might be very useful to have something like that. But for someone who is more physically able, I think there's always a risk that it might increase... like laziness, or make people more sedentary at home... they would not be getting out or whatever, and just having this culture of deliverable… you know… instantaneous… not having to do anything. You get more dependant on other things, you know, you start relying more on other technology rather than thinking more independently, if that makes sense?*
>
> ***Researcher:*** *Yeah. So you enjoy having that independence and freedom to choose what you want to buy and when you want to buy it.*
>
> ***Bertha:*** *Yeah I do actually, very much so. [...] I probably would not like to have it [the system] in my home. I already have a webcam on my laptop and the iPad and whatever and the webcam is taped over, because a friend who works in the security told me to do that, because that's what he does. And yeah, I think having a camera in your home obviously is… I think there can be this idea of tech for good, but at the same time there is risk in everything.*

As explained by Bertha, the perceived appropriateness of autonomous systems depends on a variety of complex factors, including the physical shape and the

domestic routine of its user. Whether autonomous system behavior could remain appropriate in the face of such complexity while avoiding a plethora of pitfalls seems questionable. When confronting our participants with the dilemma of accepting or rejecting the proposed systems, status quo thus seemed to generally be the preferable and safer option. As Bertha makes perspicuous, if given a choice, she would personally prefer to keep the use of autonomous systems to a minimum, or even avoid it altogether.

A reoccurring theme in our study was thus the notion that if things in our lives work fine the way they are, there is no need to change them. While some potential benefits of autonomous systems might shine through, and technology could indeed be used "for good", the plethora of risks associated with adoption of autonomous systems – such as increased dependency or security issues - would render such a transition detrimental. For as Bertha explains, "there is a risk in everything".

6.6.2 Reasoning about risks

Monitoring and collection of personal data is the lifeblood of autonomous systems, granting them the means to learn from their surroundings and take appropriate proactive measures on behalf of their users. Yet it was precisely this aspect of autonomous systems that quickly emerged as a key obstacle hampering the acceptability of the depicted solutions, with many participants voicing concerns that their privacy would be put at risk as a result of such monitoring practices.

> ***Researcher:*** *So do you feel like that's an acceptable aspect of a system? To have cameras in someone's home?*
>
> ***Francis:*** *No, no I don't. I would not like to be constantly surveyed. Not that it would be very exciting for anyone to watch me, but be it as it may, it's… no I don't think it's… the camera needs…*
>
> ***Researcher:*** *Is that out of privacy concerns? Or more about security?*

> ***Francis:*** *I suppose privacy. I mean people could argue that if I'm not doing anything wrong, what does it matter? But I don't think it's about doing anything wrong, it's about being watched all the time. I would say that's the problem.*

Much of the monitoring activity depicted in our scenarios was deliberately centred around video-cameras due to their explicit and provocative nature. More subtle monitoring technologies, such as audio listening devices, were however generally seen as equally inappropriate in a domestic context. While a number of potential security threats were brought up as a reason for this (such as the risk of monitoring devices being hacked and exploited by malicious parties), the most common cause behind such reneging attitude towards personal data collection practices was the prospect of surveillance in one's own home. As exemplified by Francis, the principle of users "being watched all the time" is deeply problematic.

Such perceived risks associated with personal data collection naturally became the most prominent when our participants reflected on the first scenario, which featured a fully autonomous system constantly monitoring all activities of its user. Our second scenario featuring a smart home system inferring data about its users through social media platforms, such as Facebook, was on the other hand generally received more favourably, albeit not without reservations.

> ***Researcher:*** *So you're not worried by the system monitoring the information... the monitoring act itself?*
>
> ***Oliver:*** *No because it uses public information from Facebook… so it's not the mechanism that's the problem, it's the fact that the information has to be open for the mechanism to work, and in a public setting that's fine, but in a private setting it isn't.*

Clearly, the distinction between publicly available and private data played a critical role when our participants sought to assess the acceptability of the various depicted approaches to personal data collection. Whereas social media platforms grant their users control over what information they want to make public, monitoring technologies deployed directly in the intimately private environment of their homes are less discriminate. Yet, as Oliver hints, the ability to *close off* private information from an autonomous system would be essential for any form of socially accountable domestic monitoring.

Besides inferring information from public social media channels, another hypothetical approach to such *closing off* of private information was postulated by our final scenario in which all personal data collection took on the form of direct communication between the user and his system.

> ***Researcher:*** *So let's say there would be no automatic monitoring going on behind your back or anything like that. The system would purely infer your preferences by asking you questions directly, would that make it alright?*
>
> ***Max:*** *Right, so as a result of keep asking you, it would build up some internal database of your preferences that should not go outside?*
>
> ***Researcher:*** *Is that important? If the preferences would go outside, would that matter?*
>
> ***Max:*** *I'm just saying… I have no track over where the data is, whether it's on their server or who's using them. It's not completely closed.*

As Max elaborates, having a complete control over what data an autonomous system infers would still not fully eliminate the perceived risks of domestic monitoring. An equally important factor is the whereabouts of any collected data along with their utilization. A prevailing concern amongst many of our participants was the notion that service providers could exploit a user's personal data for their own profit, such

as by selling them to third parties who would in turn use it as a basis for widespread targeted advertising, as illustrated by the following two vignettes.

> ***Researcher:*** *So you're worried that whoever owns the system could maybe leak some confidential information?*
>
> ***Jill:*** *Yes, like if a system identified that for example a particular person had a favourite food, then I could see how that person could end up being bombarded from the outside with adverts, for food. You know, like "you can buy carrots from such and such place". And they would think "how do these people know that I like carrots?!".*
>
> ***Researcher:*** *So what if advertisements would be very useful and contextually appropriate, would that make them acceptable?*
>
> ***Samantha:*** *I'm sure there would be other interfaces where you could get advertising, but your smart home system should just be there for your home and nothing else.*
>
> ***Marcus:*** *I'm sure, yeah. I would not use a smart home system if it meant more advertising. It's not like there's not enough advertising in the world already. I don't need it in my home.*
>
> ***Samantha:*** *Exactly!*

Many of our participants were categorically rejecting any form of advertising. As Samantha explains, a socially accountable autonomous system should serve the best interests of its users and their homes, rather than those of the service providers. While granting the user greater control over the data that is being collected by their autonomous systems can help protect their privacy, another major factor contributing towards the mitigation of risks associated with domestic autonomous systems can thus be summed up as the perceived trustworthiness of autonomous system behaviour.

6.6.3 Reasoning about trustworthiness

The need for trust was persistently coming up throughout the user study as a key requirement of social accountability. As we shall see in this subsection, every stage of human-data interfacing, including the process of data collection, the subsequent storing and protection of personal data, its utilization by the service provider as well as the provider's broader agenda would all help shape the perceived overall trustworthiness of a domestic autonomous system. In the wake of numerous heavily medialized data scandals, maintaining a relationship of trust across all these stages can pose a considerable challenge.

> ***Researcher:*** *So if a company would sell you a system like this and tell you "our camera will only be monitoring your food consumption", would you trust them?*
>
> ***Miranda:*** *[a brief pause] ...these days? No. Not these days.*
>
> ***Researcher:*** *Because of all these incidents like Cambridge Analytica?*
>
> ***Miranda:*** *Because of all those incidents, yeah. But maybe six months ago I would have.*
>
> ***Researcher:*** *So you feel like your understanding has changed a lot over the last few months?*
>
> ***Miranda:*** *Oh absolutely, you just become more aware of how companies are acting and the things that have happened to other people. There's just more awareness.*

While the majority of our participants had little to no knowledge of the technical processes underpinning the operation of autonomous systems, a rising awareness of "how companies are acting" or of the "things that have happened to other people" were nonetheless frequently invoked as hindrances to the trustworthiness of autonomous systems, as exemplified by Miranda.

Against the backdrop of such a climate of suspicions, our participant's responses made it clear that building trust into autonomous systems would require a substantial effort on behalf of the service provider.

> ***Researcher:*** *OK, so you feel like you would need to trust whoever was receiving your data?*
>
> ***Jill:*** *Oh definitely. I would need regular assurances of what's happening to my data. And I also need somewhere to go to if I felt my trust has been breached. Someone would have to be available for me to talk to and sort out any problems or issues.*
>
> ***Researcher:*** *So if a system like this would be owned by a large corporation, such as Amazon, and they would guarantee you to keep your data safe, would you trust them?*
>
> ***Jill:*** *I think so. I mean we have been dealing with Amazon for quite some time now, I mean buying products from them, and so far we've never had any problem with data breach, or confidentiality breach. You know, I know there's a lot of people out there that do not like Amazon, but up till now we've been happy with them.*

Through the exchange above, Jill unpacks in greater detail her expectancies regarding what exactly constitutes a trustful relationship. While a provider's ability to successfully and continually communicate the whereabouts of a user's personal data serves as a helpful assurance that no malicious activities are going on, our participants were likewise conscious that some level of friction might be inevitable. For such purposes the user needs to be able to also communicate concerns back to the provider and receive adequate support. A user's peace of mind, stemming from the knowledge that the service provider can be held accountable, is thus a key component of social accountability.

As the vignette above shows, another important factor affecting the trustworthiness of autonomous systems is the user's own personal experience with their use. A relationship of trust does not evolve overnight, but rather, as Jill makes clear, through a longstanding problem-free use.

Some participants went a step further, arguing that autonomous systems and their providers should be subjected to regulations (e.g. by government agencies) to make sure no user exploitation takes place. In some situations trusting fully autonomous systems was seen as problematic regardless of the provider's policy, in which cases automation should take the backseat and allow for greater user participation. Imposing such constraints on autonomous systems was seen as particularly important when dealing with sensitive issues, such as safeguarding a user's health and wellbeing.

> ***Researcher:*** *Would you trust an autonomous system to secure the safety and wellbeing of its users?*
>
> ***Anne:*** *So my understanding would be that the system should not make decisions about securing, without the input of humans… so that's not necessarily moving it away from automation, but it is about involving humans. So I think again, it's more a middle point in that. The security doesn't happen independently of the individual. The security should be worked on together between the system and the individual.*

As Anne explains, having some control over one's system can go a long way towards improving trust. A collaborative human-computer relationship of mutual interdependence, where all users affected by a smart home system are kept in the loop, thus often represents a preferable alternative to fully autonomous solutions. A system should aspire to become another social agent gearing in with the social expectations that apply within the given domestic environment. Trustworthiness in this sense goes hand in hand with preservation of human agency.

6.6.4 Reasoning about agency and entitlement

While our participants made it explicit throughout the study that a range of tasks in the home could indeed benefit from automation, the perceived need for such assistance is clearly stratified. While mundane tasks, such as recommending the user a suitable movie or music performer, were brought up as being generally acceptable, autonomous actions with a more far reaching implications for the user's lifestyle were generally treated with caution.

> ***Researcher:*** *What if a company would offer you a system that's monitoring your health and recommends you items that are the most compatible with your situation... the healthiest choices for you?*
>
> ***Veronica:*** *Knowing the things that we're consuming or choosing for the recommendations that we've been having over the last year or so... just getting me recommendations related to the things that I need?*
>
> ***Researcher:*** *Yes, like "you've eaten too few calories today, you should eat a cheese now. Would you like me to buy one?" or something like that?*
>
> ***Veronica:*** *No, because I think, you know, still you can just make your own choices on, you know, one day you just want to have cake full of calories, you don't care. And I just want to do it the way I want to… I prefer to find my own balance and managing my own calories. I think that would be a bit intruding.*

As Veronica explains, a system actively directing a user's lifestyle would be perceived as inappropriate or even "intruding". There is a fine line to be walked for any solution that provides useful help for its users without diminishing their agency and a sense of personal freedom. Further details regarding such a point of equilibrium surfaced when reflecting on the notion of systems tracking variables pertaining to the personal health of household members.

> ***Liam:*** *I think we have a responsibility to look after ourselves and our health as well, do you know what I mean? It's about making people aware of stuff and then hoping they'll make the right choice. Rather than taking the agency away from people in some respect.*
>
> ***Nancy:*** *Or taking away their decision-making ability.*
>
> ***Liam:*** *Yeah, I think for me it's problematic. Again, it's about being informed, rather than controlled.*

The responsibility over one's own health and wellbeing represents an essential component of human autonomy and the notion of it being delegated to computational agents is consequently not seen as an acceptable prospect, as clearly exemplified by the exchange above.

While this does not rule out the use of autonomous systems in this context, it is essential that they are based around the principle of creating the best possible conditions for their users to take sound responsible decisions by keeping them informed, rather than by taking those decisions on their behalf. It is in other words crucial for any socially accountable autonomous system to be designed in a manner that would complement human autonomy, rather than compete with it.

6.7 Key Takeaways

The expectancies expressed by prospective users in our third user study were closely aligned with those that surfaced through our previous enquiries, while further elaborating them in a number of potentially meaningful ways. Below we summarize the key points of discussion brought up by our participants during their interviews.

The responses our interviewees provided when confronted with questions concerning coordination further corroborated one of the central themes from our

two initial studies, namely the notion that coordinating an autonomous system's actions with its users can scarcely be achieved through the system's own proactive inference alone. Rather than simply expressing the need for greater situational awareness (study 1) or proposing systems that would rely on recommendations instead of physical actuation (study 2), here, a number of our interviewees proposed more actionable solutions. For instance, it was argued that a calibration mechanism ought to be put in place, allowing the user to manually setup the system operation. Such a mechanism should ideally be complemented by an interface enabling continuous coordination in the form of direct dialogue with the user (e.g. a voice-based interface). Our interviewees made it clear that this interface would likewise need to be subject to user configuration, making sure it is only used at a time and place suitable to the user.

Similarly to our second study, the interviewees also emphasized the importance of systems respecting the social nature of routines when gearing in with the domestic environment. It is imperative, they argued, that designers of future autonomous solutions recognize that mundane household tasks are not merely a necessary process by which people satisfy their material needs. While much of this household work does indeed represent tedious chore, even the most basic of such tasks, including shopping, opening of blinds or checking the weather, can potentially carry personal significance for the user. Indiscriminate automation of household activities can increase the user's convenience, but this may also come at a price, such as the loss of social cohesion, limited creativity or generally reduced enjoyment. A vital requirement of a socially acceptable coordination between a domestic autonomous system and its users is thus revolving around achieving a division of labour wherein a system refrains from interfering with tasks that the user enjoys carrying out manually.

When prompted to elaborate this position further, our interviewees generally made it clear that the design of autonomous systems cannot be guided merely by the goal of maximizing human well-being, but rather first and foremost by preserving the human autonomy. Users choosing to act in a manner that is detrimental to

themselves (e.g. by ordering unhealthy food) must be free to do so. Instead of taking decisions on user's behalf, the prime role of a domestic autonomous system should be to make sure that user's decisions are informed and conscious. This does not necessarily need to happen through explicit recommendations issued by the system to the user, as such an approach might inadvertently result in a degree of bias and narrow down the user's views. Instead, they argued, information provided to the user should be presented in a manner that would encourage agency and critical thinking.

A novel aspect of coordination that the interviewees in our final study brought up was the need to distinguish between *short-term* and *long-term* autonomous system behaviour. A particular autonomous action (such as the reordering of milk) can be highly desirable in a short term context, but once extended and repeated over a longer time frame, it runs the risk of affecting the user in a detrimental manner, such as by causing lack of variety or by fostering a passive and monotonous lifestyle. It is thus vital that any long-term system behaviour by default involves the users, such as by periodically prompting them for renewed authorization or by recommending alternatives to stimulate diversity.

Control, and the need for its granularity, once again surfaced as a major point of discussion. Here, our interviewees argued that granting the user a sufficient degree of control would require the customizability of individual system features. While this can result in an overwhelming amount of settings and options, the responses provided by our interviewees suggested that such a level of complexity is acceptable as long as the user is free to choose whether to deal with it or not. Even if left unutilized, high customizability will in this sense still help build a relationship of trust between the users and their autonomous systems. While we did come across a few cases where the interviewees in our final study argued that depriving the user of control could be legitimate (e.g. to protect the domestic hierarchy), these cases were scarce and it was made clear to us that the acceptability of such a machine behaviour would require it to be sufficiently and explicitly justified by drawing on existing

man-made decrees and social policies (e.g. the legality of user's actions) rather than a machine's own initiative.

Enquiring about computational accountability allowed us to further corroborate that the actions carried out by autonomous systems are themselves generally legible without requiring further explanation, particularly so when dealing with mundane domestic activities (such as the restocking of milk). Instead, in line with our two initial user studies, our interviewees argued it is the grounds upon which actions are taken that are in need of explanation. In comparison to our previous studies, this time our interviewees provided greater clarity regarding what an account ought to look like, arguing it is not *what* a system is doing that needs to be made accountable, but rather *how* and *why* the system arrived at its decisions (e.g. *How did it determine milk should be restocked? Why this particular brand? etc.*). Moreover, much like in our initial pilot study, we found that an account explaining *why* a system arrived at a particular decision should ideally include the reason along with the motive leading up to a system decision. Such accounts can be reasonably simplistic, such as "It's Tuesday today. You have a history of eating chicken on Tuesdays. Therefore I have ordered you some chicken".

The interviewees in our third study made it clear that it is not just the individual system actions that are in need of such explanation, but likewise the overarching agenda pursued by an autonomous system. This became particularly perspicuous when debating the potential use of food recommendations as a tool for altering the user's lifestyle. As our interviewees explained, this practice represents a double edged sword. It is crucial that the overall direction of system actions is made legible to its users in order to allow them to decide whether it is aligned to their own preferences, rather than serving the goals of any third party. A deluge of factors are involved in making such system motives accountable, including the ownership of the system and the perceived reputation of this owner in the eyes of the user. Along with legibility of system actions, accountable motives are thus to some extent also dependent on trust established between the user and the service provider.

Multiple of our interviewees stressed that such accounts should be provided not just out of trust or privacy related concerns, but also due to the fact that autonomous systems could be prone to making errors. Sufficient computational accountability would in this sense help users verify the exactness of autonomous decision making and - if necessary - allow them to rectify erroneous system decisions (e.g. the recommending of an unsuitable meal).

Another contentious topic revolving around computational accountability had to do with the visibility of personal data collection. Much like in our second user study, our interviewees argued that while home monitoring is generally acceptable (as long as it is taking place with the end user's consent), making the resulting flow of personal data accountable to relevant users would be crucial in order to allow them to safeguard this process and make sure the monitoring practice is not overstepping the limits set out in any user agreement. The interviewees in our third study went to a greater depth, explaining that an account here consists of information revealing *how*, *where* and *when* a user's data is being collected, as well as *where* the data goes and *how* it is being used.

Finally, in line with our two previous users studies, we did come across numerous instances where domestic autonomous solutions appeared to introduce various elements of friction that could not be simply attributed to poor coordination, lack of control or illegible machine actions, but rather to their inappropriate nature relatively to the broader social context within which they were situated. Much of the concerns pertaining to *social accountability*, as we labelled this novel problem area, are closely intertwined with the other three problem areas. Yet, as we will seek to argue in the discussion section later in this thesis, treating it as a separate strand of acceptability is indeed justifiable.

A majority of the relevant concerns brought up during our third user study dealt with trust and the crucial role it plays in any appropriate social conduct. The intimate nature of a domestic environment, for instance, prompted multiple interviewees to argue that legibility of personal data-collection processes alone does

not make them appropriate. Instead, for any data collection practice to be perceived as trustworthy - and thus socially appropriate - it would have to only be able to collect data that the user intended for collection. One example of such a mechanism brought up during our third user study came in the form of social media channels that allow the user to explicitly specify which content is supposed to be made public.

When prompted for further elaboration, the interviewees in our third user study made it perspicuous that trustworthiness of autonomous systems stretches beyond the data collection practice to encompass each stage of the human-data interfacing cycle, including the secure storing of collected personal data, its responsible utilization and a healthy relation between the user and the service provider. Secure personal data storage is underpinned by the service provider's transparency, regular assurances and an adequate level of support provided for the end user. Responsible utilization of collected personal data is in turn underpinned by actively keeping the user in the loop. Finally, a healthy relationship with the service provider is dependent on a long-standing problem-free personal experience.

The expectancies concerning social acceptability of autonomous systems voiced by the participants of our final user study were noticeably overlapping with those that surfaced during our two initial user studies. This suggests that a reasonable level of data saturation has been reached and we now have sufficient empirical grounds for articulating conclusive and actionable takeaways. In the following chapter, we thus turn to reflecting on the significance and potential use of our findings.

Chapter 7 Discussion

In this chapter we summarize the work conducted throughout our study and reflect on our key findings in the context of relevant research literature.

A total number of 80 prospective users representing all walks of life were recruited for 3 separate user studies and asked to evaluate a combined number of 7 distinct scenarios, each pitching a different set of hypothetical future autonomous systems for the home. Our initial study was designed to probe prospective users in focus groups for their expectancies regarding the general notion of actuating domestic autonomous systems. The findings made through this enquiry confirmed the existence of major problem areas hampering the acceptability of emerging autonomous systems for the home. Much in line with existing literature, we were able to identify concerns pertaining to *coordination*, *control* and *computational accountability*. Additionally, in a bid to articulate our participants' concerns with social appropriateness of machine behaviour, we introduced a fourth problem area in the form of *social accountability*.

Our second user study was designed to corroborate and expand on these initial findings. For this purpose we developed a set of two contrasting video scenarios, each postulating autonomous systems pursuing different priorities, such as the convenience of their users or the well-being of the local community. The study allowed us to surface additional details pertaining to the values that users invoke when reasoning about the acceptability of prospective autonomous systems.

Our third and final user study allowed us to consolidate our findings by revisiting the key concerns brought up by participants during our first two studies and elaborate these in greater depth.

It might be countered that these are grand claims to make on the basis of 80 people's say so. However, while articulated by a relatively small number of people, the background expectancies they have expressed do not belong to them alone. As Garfinkel (1967) put it, "Almost alone among sociological theorists, the late Alfred Schutz, in a series of classical studies of the constitutive phenomenology of the world of everyday life, described many of these seen but unnoticed background expectancies. He called them the *attitude of daily life*. He referred to their scenic attributions as the 'world known in common and taken for granted'."

The attitude of daily life, the world known in common, or in other words what anyone i.e. (with Bittner's (1973) caveat) "any normally competent, wide awake adult" (Whitworth, 2005) knows about everyday life. There is nothing special about our findings then. They do not speak of and elaborate rarefied knowledge or insight possessed by a privileged few. Rather, the background expectancies articulated by our participants are known in common, shared, used, recognised and relied upon by a much larger cohort, and it is for this reason that they elaborate significant challenges for the design of autonomous systems for the home.

In the remainder of this chapter, we shall consider our research methodology and our findings in greater detail. In so doing, we shall seek to offer an answer to each of the three research questions introduced at the beginning of this thesis.

7.1 Methodology: Lessons Learned

Let us now return to consider the first stated research question of this work. We said that we wanted to find out:

- *R1: How can we assess the manner in which prospective users will respond and relate to future autonomous systems for the home?*

As a centrepiece of our engagement with prospective users, we developed a set of scenarios, each depicting various hypothetical autonomous systems for the home. While none of these systems represented an actual existing solution, we took great care to base the depicted ideas on extrapolations from current technologies in order to not drift too far away from reality. Each of the depicted systems was portrayed in multiple contrasting versions, loosely following a methodological framework known as *contravision* (Mancini et al., 2010). The contrasting scenarios were intentionally created and employed to enable *breaching experiments* that disturb common sense understandings of domestic life and thus provoke our participants into revealing and articulating their social expectancies.

It was this combination of breaching experiments, scenario-based design and contravision then, that provided us with the means to elicit, collect and analyse user data pertinent to future autonomous systems for the home. As we grappled with our research findings, and indeed the methodology itself, our understanding and application of these methodological components evolved. Below we look back and consider this evolution more closely, reflecting on relevant literature and elaborating our key takeaways.

7.1.1 Breaching Experiments

The central idea behind our application of breaching experiments was that by employing provocative scenarios to create 'reality disjunctures', we would be able to surface taken for granted background expectancies in participants' efforts to repair them. These background expectancies are ordinarily used by people in their mundane reasoning to understand and order action and interaction in what sociologist Harold Garfinkel (1967) called "an obstinately familiar world". That is, a world that members are "demonstrably responsive to" in mundane interaction but are ordinarily "at a loss" to tell us just what the expectancies that lend commonplace scenes their familiar, life-as-usual character consist of (Ibid.).

One might argue that the novelty of presenting breaching experiments for design is rather limited. However, the novelty here lies in the intentional configuration of contra-vision scenarios to drive a workshop-based approach to breaching experiments. This sharply contrasts to previous uses of breaching experiments in design to understand *in the wild* deployments of technology (Mann et al., 2002; Crabtree, 2004; Tolmie and Crabtree, 2008; Poole, 2012), not that those previous approaches are homogenous.

Breaching experiments were, to the best of our knowledge, first introduced into design by Steve Mann et al. (2002), who used wearable computing to create a set of "visible and explicit sousveillance" performances "that follow Harold Garfinkel's ethnomethodological approach to breaching norms". Mann's use of breaching experiments was copybook i.e. he sought to make trouble and thereby "expose hitherto discreet, implicit, and unquestioned acts of organisational surveillance".

The *make trouble* approach to breaching experiments surfaced again in design in 2009, when Erika Shehan Poole (2012) sought to exploit breaching experiments to investigate "existing technology […] related practices in domestic settings". Poole asked participants in her field trial to interact with domestic technology "in ways that potentially disrupted the social norms of the home". More specifically, "each

home received weekly *homework* … intentionally designed to breach household technology installation, usage, and maintenance practices. The assignments the first week served as warm-up to acclimate participants to being in the study … For the second week, the participants were instructed to have the less technically oriented adult in the home complete the assignments. This choice was made to disrupt the normal family dynamic …".

The results of Poole's (2012) intentional efforts at disruption "provided explanations of why problems with technical advice sharing and home technical maintenance persist". Poole subsequently recommends that breaching experiments should be considered as "an asset and an indispensable part of a researcher's toolbox for understanding existing social norms and practices surrounding technology". We do not disagree but would advise caution be exercised when deliberately disrupting the dynamics of any social setting in which the researcher is essentially a guest, and note an alternative approach to breaching experiments in design.

Following Mann, in 2004 Crabtree (2004) introduced the notion of breaching experiments not as things that necessarily make trouble and cause disruption, but as an analytic lens on in the wild deployments of novel technology. This approach focuses on explicating through ethnography the "contingent ways in which novel technology for which no use practice exists is made to work and the interactional practices providing for and organising that work". In a similar vein, Tolmie and Crabtree (2008) treated novel technological deployments in the home as breaching experiments that "make tacit and taken for granted expectations visible… enabling us to see how even a simple arrangement of technology can breach ordinary expectations of where technology resides in the home, who owns it, who maintains it, and how user experience of it is accounted for".

However, it is not that there are at least two different approaches towards breaching experiments in design or the contrast between them that matters. Despite their differences, previous breaching experiments are oriented, as Poole succinctly puts it,

to existing technology and related practices, whether it be a novel prototype or technology that has been appropriated at scale and is well established.

Our own application of breaching experiments, on the other hand, is not focused on existing technology, whether or not it makes trouble and disrupts or provokes practice by virtue of it having to be *made to work* in the world. There is no actual system, functionality, connectivity or interactivity in our breaching experiments, only contrasting envisionments of autonomous systems at work in everyday life. Our breaching experiments are oriented to future and emerging technology and the acceptability challenges that confront their "adoption in everyday life" (Lindley et al., 2017). Rather than explicate, and even explain existing practice, they instead seek to engage potential end-users in reasoning about the place of future and emerging technologies in their everyday lives and thereby inform us as to key challenges that need to be addressed to make those technologies an acceptable feature of everyday life at an early stage in the design life cycle, before we have built anything. Furthermore, our breaching experiments are done through focus groups and interviews rather than performances, field trials or ethnography.

It would appear then that there is some novelty to our approach after all: we are not using breaching experiments to make trouble or disrupt or to provide an analytic lens on existing technology-related practice and we are not conducting them in previously practiced ways. The novelty in our approach lies in repurposing tried and tested methods to create contrasting scenarios that are designed to disrupt background expectancies that organise everyday life in familiar settings (in this case, the home), and as our findings make perspicuous those expectancies have little to do with existing technology-related practice too. The disruption lies in what are essentially incongruous visions of the future depicted by contra-vision scenarios, which in being presented to potential end-users create reality disjunctures that motivate efforts at resolution and repair. It is in the attempt to "resolve incongruities" (Garfinkel, 1963) and repair the reality disjunctures they occasion that ordinarily tacit and unspoken background expectancies are surfaced; expectancies

about which people usually have "little or nothing to say" when asked (Garfinkel, 1967) but which this methodological innovation enables early on in a design process.

7.1.2 Scenarios

Our focus on informing the early stages of a design process made us predisposed to relying on scenarios as a vehicle for conveying our ideas to prospective users. As explained by Gabrielli and Zoels (2003), "scenarios of use (through storyboards, videos, or demonstrations) can model ideas into seemingly real situations so that audiences have much more to respond to, prodding them into a fuller reaction. Thus [...] planners can gather information, gain insights, and affect the direction of projects much earlier in the process". The nature of the ideas we were able to convey through the use of scenarios is of a particular note here. By portraying hypothetical autonomous systems, services and providers alongside fictional characters living their mundane lives, we sought to recreate the effects of a broader socio-technical ecosystem brought about by a wider adoption of autonomous systems, far beyond the scope of what could reasonably be simulated through the deployment of an actual working system prototype.

Such a line of reasoning is not without a precedent. Indeed, the practice of using scenarios to envision future everyday life has historically been intimately intertwined with the design of emerging computing solutions. There is a particularly notable connection to the origins of ubiquitous computing research at PARC in the late 1980s, which in turn emerged from the efforts to develop a new computer paradigm to surpass and replace the PC (Press, 1999). In a retrospective reflection on their design process, Weiser et al. (1999) recount the manner in which the scope of their design vision expanded over time: "The program was at first envisioned as a radical answer to what was wrong with the personal computer: too complex and hard to use; too demanding of attention; too isolating from other people and activities; and too dominating as it colonized our desktops and our lives. [...] By 1992, when our first experimental ubicomp system was being implemented, we

came to realize that we were, in fact, actually redefining the entire relationship of humans, work, and technology for the post-PC era".

One of the ways in which this redefining process materialized was Weiser's (1991) now famous Sal scenario. Its pervasiveness and the extent to which it opened up the ideas behind the newly emerging UbiComp field to discussion led Bleecker (2009) to argue that it is in science fiction stories where many of the key principles and examples of the future technologies that were pioneered at Xerox PARC find their most complete expression. Bleecker concludes that the "UbiComp science fiction" can bring to light consequences, conclusions and implications much better than what could be done through any "science fact" paper, or laboratory demonstration: "Any good scientist should also be particularly good at science fiction, maybe even better than they are at science fact".

While we largely share Bleecker's enthusiasm for the use of science fiction stories (or 'scenarios' in the case of this thesis) in evaluation of emerging technologies, we also feel compelled to point out some potential limitations associated with this practice. As explained by Stead et al. (2018), if fictional stories are to facilitate meaningful questions around the futures they portray, they need to be conceived with the same commitment to detail as if they were actually being designed and produced. This is a mixed blessing. Not being held back by practical constraints, such as the technological feasibility of the proposed ideas, allows designers to convey rich, fully realized visions without the need to compromise. On the other hand, in so doing, detailed envisionings of the future can also inadvertently influence prospective users towards interpreting aspects of the vision as self-evident, making them consider the depicted technologies and activities in a particular way, rather than focussing on what would otherwise matter to them. This in turn can lead to major challenges associated with the envisioned technology going by unnoticed.

It is worth recalling Reeves et al.'s (2016) position on envisioning here, who called for more consideration for the ways in which visions influence design processes and research agendas. The authors argue that "grand visions" and projections continue

to shape subsequent technology development, for example in the UbiComp community. However, these visions may lack social legitimacy in that they can be narrowly focussed, may not take into account the concerns of people outside of the Ubicomp community itself, such as the general public, or miss goings-on in other parts of the world (Bell and Dourish, 2007).

By way of example, Weiser's own vision came under attack by Yvonne Rogers (2006), who argued that it downplayed the risk of UbiComp leading to a passive and solitary lifestyle. In response, Rogers formulated an alternative vision of UbiComp, one that sought to re-prioritize from "proactive technologies" towards "proactive people". Instead of augmenting the environment and calmly serving the needs of the users, and thus reducing the need for humans to think for themselves about what to do, Rogers argued that we should consider how UbiComp technologies could be designed to "augment the human intellect so that people can perform ever greater feats, extending their ability to learn, make decisions, reason, create, solve complex problems and generate innovative ideas".

Against the backdrop of this debate, a key challenge we had to deal with while designing our scenarios was then to ensure that the contrasting (e.g. utopian and dystopian) futures depicted in them would not predefine any problematic aspects of future or emerging technology. Rather, as outlined above, the job of our scenarios was to disrupt and create a reality disjuncture, whose repair surfaces the taken for granted expectancies that impact future and emerging technologies in everyday life.

In retrospect, is there any way for us to tell if we succeeded, if our scenarios avoided limiting the range of issues brought up by participants by constraining them to the incongruous topics depicted in them? The vignettes presented in the course of our 3 study chapters would suggest we did, insofar as our participants' talk and reasoning can be seen to range across a great many matters not depicted in the scenarios. It would be more apposite, then, to see the contra-vision scenarios not as pre-defining design issues or topics but as provocative social objects that elicit multiple

background expectancies, which participants themselves come to shape and prioritise in their talk as they go about resolving the incongruities they create.

Consider, for instance, the manner in which our participants responded to the promise of greater convenience by virtue of automation, as postulated by many of our scenarios. In a sense our participant's reasoning shows that while people welcome values such as convenience and efficiency as drivers of technology design, this is preconditioned by the technology manifesting conformity insofar as people need to be able to integrate it into their practices of everyday life without causing too much disruption.

Even in cases where our scenarios envisioned a perfectly problem free and convenient lifestyle brought about by autonomous systems (e.g. scenario 4), our participants were frequently problematizing the conveyed ideas. Instead of criticizing convenience or efficiency, it was the cost at which they might be achieved that became the subject of scrutiny. Designing solutions that prioritise the convenience of users was suggested to, for instance, help free up time for activities that are highly valued, such as enabling a user to spend more time with loved ones. In a different context, however, prioritizing user's convenience was agreed to give rise to negative implications, such as reducing a users' sense of self-determination and freedom of choice.

Similarly, while our work confirms that people frequently value the notion of autonomous solutions stimulating a more active lifestyle and a sense of community-building (e.g. scenario 3), our participants made it clear that more fundamentally the technology would need to fit in with people's practical needs as a prerequisite, such as the need for privacy.

In spite of our contrasting scenarios, the outcome of our participants' reflections was generally such that values including convenience and efficiency were not seen as a polar opposite to those of a proactive and social lifestyle. By the same token, 'proactive technologies' and 'proactive people' were not seen as inherently constituting mutually exclusive design goals. Unlike the broad overarching visions

put forth by Weiser (1999) and Rogers (2006), the findings elicited through our scenarios thus suggest that the two can coexist, each with its own set of potential trade-offs emerging through their interplay with a rich palette of human values.

What our findings show more fundamentally then is that utilizing breaching experiments centered around scenario-based contravisions enables the surfacing of a rich contextual interplay of values that mediate whether or not a technology is seen as potentially acceptable. These are invoked in relation to locally accountable practices, for example in rejecting the domestic robot as a companion, while embracing the potential benefit to eat more healthily as a result of a robot's labour. We would suggest that one of the contributions stemming from our methodological approach is that it helps demonstrate those tradeoffs in terms of the complex ways in which values are drawn on in reasoning about domestic autonomous services, and that this can be instructive to sensitise designers.

7.1.3 Contravision

The notion of using "futuristic videos, or other narrative forms, that convey either negative or positive aspects of the proposed technology for the same scenarios", or 'Contravision' as Mancini et al. (2010) labeld this method originally, played a key role in enabling us to surface the contextual interplay of participant's values described above. Our own approach to contravision evolved over time from the conveying of simple polar opposites (e.g. utopian and dystopian visions) towards a more multifaceted envisioning of future autonomous systems and is thus perhaps best aligned with the work of Reeves (2016) in that he advocates to create "multiple scenarios based on different envisionings, rather than working under the implicit assumption of only one particular envisioning."

We have argued that the main benefit of this approach resides in its capacity to invite the participation of prospective users in shaping the conveyed visions of future technology. A goal which we share with the 'participatory design' practice.

Coulton et al. (2019) argues that design fiction itself was originally conceived as part of a participatory design process as it provides "simple and powerful means to include stakeholder groups directly with a proximate future and that future's potential implications".

In a sense, there is, however, a much older historical precedent. The ancient greek philosopher Socrates famously opposed the use of writing as a tool for advancing knowledge. Real knowledge, Socrates argued, can only be gathered via dialogue; a sequence of questions and answers where ideas are interrogated until the knowledge is truly understood. Once written down, on the other hand, ideas will inevitably lose their quality of flexibility and the power to adjust and correct in the light of new arguments (Williams, 2003).

Following a similar line of reasoning, we would thus argue that designers ought to resist fully embracing the narrative component traditionally associated with design fiction. A story featuring detailed claims about a user's interactions with a particular system, along with the consequences of these interactions, will run the risk of being perceived by prospective users as 'fait accompli' in the sense of offering little room for interpretation or negotiation.

In our efforts to prevent such an outcome, we reasoned that by diversifying the conveyed vision, by imbuing each of its instances with contrasting claims, the conveyed vision as a whole would give the impression of an ongoing dialogue, one that our participants were invited to join in on. The outcome was such that members of the general public partaking in our study ended up mixing, matching and critiquing the envisioned ideas, rather than accepting them as they were. Although members of the general public are not necessarily wiser about what the future might bring than we designers are, our research demonstrated that they have their own value judgements of how technology might improve their lives. Thus, increased participation is likely to generate a picture of the future that is more socially legitimate than what experts alone could create. This is significant, for as Reeves et al. (2016) explain, a broader base of legitimization creates fertile ground for the design of better future systems.

Our goals with the use of contravision are thus largely in agreement with those of Lindley et al. (2017), who advocate the 'World Building' approach to design fiction as a means for placing emerging technologies in potential futures: "Transcending the limitations that are inherited by associating Design Fiction with notions of story and narrative, 'World Building' aims to create plausible depictions of the near future within which today's emerging technologies may be prototyped as if they are domesticated technologies situated in everyday life/culture". World building thus puts the emphasis on conveying a plausible future context, rather than narration.

Some may object that we are downplaying the role our selected media platforms played in eliciting data during our user studies. Indeed, McLuhan's (1964) distinction between 'hot' and 'cool' media might seem of relevance here. Cool media, such as comic books or phone conversations, require high audience participation in the sense that the receiver must fill-in the missing information. On the other hand, hot media, such as movies or radio broadcasts, spoonfeed their receivers, in the sense of giving them little room for interpretation.

Using such a dichotomy, it might seem tempting to conclude that, in order to maximize the participation of users, it is cold, rather than hot media that designers should strive to utilize when conveying their visions. Yet, in our research, having started out with graphic storyboards and gradually moved up towards higher fidelity movie-like scenarios, this was not necessarily the case. We would argue, however, that the relative consistency of findings across our three studies demonstrated that it is not necessarily the medium itself that steers the degree of user participation. Rather, we reason that by conveying contrasting versions of future technologies, we invited user participation in reconciling the depicted inconsistencies and thus in a sense 'cooled down' the mediated experience. This in turn enabled us to harness the richness and detail afforded by some of the hot media, without excluding the user from participation.

That being said, we also acknowledge that we have here merely scratched the surface and more work is needed on elaborating the potential role of different media platforms in enabling and encouraging participation of prospective users in

envisioning practices. Similarly, in our work we `merely' sought to relay the general public's response to certain visions of future technology. Further participatory research is thus needed to address the lamented lack of `social legitimacy' (Reeves et al., 2016) to involve the general public in shaping truly diverse and inclusive visions beyond our provocative scenarios.

7.2 The Problem of Social Acceptability

It is evident that major challenges have yet to be solved if future autonomous systems are to be rendered acceptable and fully embraced by the general public in their homes. In this thesis we set out to shed more light on these challenges and in so doing offer an answer to our second research question:

- *R2: What are the key areas of friction impacting the acceptability of future domestic autonomous systems? What implications do these frictions have?*

In breaching taken for granted background expectancies that are usually left unspoken, our contravision scenarios have elaborated key human barriers to acceptability of future domestic autonomous solutions connected to coordination, control and computational accountability. A common denominator of all our user studies was, however, that these three preconceived problem areas frequently took the backseat in favour of social accountability, which ended up dominating many of the focus group discussions and interviews. Whereas the three former problem areas all represent long-standing and well documented concerns in Human-Computer Interaction, the notion of social accountability, as gradually formulated throughout this thesis, is novel and can thus be considered as one of our key contributions.

At the same time it is also important to acknowledge that this particular taxonomy was converged upon by synthesizing the framing and findings of this particular thesis. There are certainly different ways in which one could slice and dice the problems surrounding social acceptability of domestic autonomous systems. There

is, for instance, a sense in which all of the surfaced problem areas coalesce around *trust*. Trust that systems will do the right thing at the right time. Trust that systems will allow users to take command of their activity. Trust that systems will keep their users informed regarding their activity. Trust that systems will be responsive to the social circumstances of their actions.

The overhanging scepticism concerning the degree to which future systems could satisfy such expectations appears to be placing emerging autonomous systems on a trajectory towards a widespread crisis of trust. We are thus in agreement with researchers, such as Lahijanian and Kwiatkowska (2003), who argue that addressing the problem of trust represents arguably the greatest challenge faced by designers of future and emerging autonomous systems for the home. Finding a solution will be a difficult task indeed, one that will require a healthy dose of critical thinking while maintaining an open mind and drawing on a wide spectrum of academic disciplines stretching far beyond computer science and HCI to also include areas such as sociology, psychology, logic, politics and ethics. With this in mind then, looking back on our own findings, what have we learned?

7.2.1 Coordination and control

We treat coordination and control together here as they may be seen to directly complement one another and span a spectrum of expectations to do with the orchestration of autonomous and human behaviours. Our findings also make it visible that context awareness is seen as key to coordination by our participants to ensure that autonomous systems do not make inappropriate interventions. However, as Bellotti and Edwards (2001) point out, in order to become acceptable, systems cannot rely purely on context awareness to do things automatically on our behalf but should rather involve active input from users at least on some level. Similarly, Whitworth (2005) argues that computing systems often have a poor understanding of context, which makes it necessary to give users control in order to preserve their autonomy. An issue underscored by Yang and Newman (2012), who

have argued that optimal user experience should be achieved through balancing machine intelligence with direct user control.

Our research enquiry unveils a similar sense of scepticism regarding the ability of fully autonomous systems to correctly assess every given situation and act in line with our expectations on a consistent basis. A key expectation at work here concerns the timeliness of autonomous behaviours and the need to synchronise them with the user's situation. Our participants' expectations regarding timeliness are of particular note, including, but moving beyond, a "here and now" understanding of timeliness (e.g. raising potentially embarrassing matters while I am entertaining visitors). Of equal concern are the temporal horizon of action and the need for autonomous systems to be sensitive and responsive to temporally distributed patterns of human behaviour (e.g. long work schedules).

Our participants therefore expect that human input will be required, which they speak about in terms of *calibration*, *customisation* and user-driven *scheduling* of events. In this respect, it is expected that calibration and customisation would allow users to coordinate the behaviours of autonomous systems with occasional and established patterns of human conduct (e.g. long but not permanent work schedules and the reoccurring rhythms and routines of domestic life). This would allow users to configure appropriate actions and help address the thorny problem of how autonomous systems are to develop an awareness of context, including what is appropriate to do and when it is appropriate to do it. In effect, the problem of learning context is offloaded onto the user to some extent through the provision of coordination mechanisms that enable users to gear the behaviours of autonomous systems in with everyday life within the particular domestic environments they are deployed and used.

The qualification *to some extent* is important here, for no matter how much autonomous systems learn about everyday life, and how smart they become, they will likely never be able to anticipate and respond appropriately to all social circumstances (Suchman and Weber, 2016). As our enquiry demonstrated, the contingencies and social idiosyncrasies embedded into nearly every aspect of the

domestic conduct are unlikely to be inferable and assessable by machine autonomy alone. Whereas Horovitz (1999) argued that machine autonomy ought to be applied based on the level of *uncertainty*, our findings indicate that in a domestic environment, this is not necessarily the case. A deluge of domestic activities (e.g. shopping or opening of the blinds) can be predicted with a reasonably high level of certainty, and yet, for a range of personal reasons, the user might wish to carry them out manually.

Thus, it is expected that end-users will be able to exert direct control over autonomous systems. However, the expectation is not that users will simply be able to turn autonomous systems off—that may only be necessary in critical situations—but rather that direct control can be exercised granularly. In effect, it is expected that the intelligent features built into autonomous systems would be subject to individual control, enabling the user to increase or decrease the intelligence of the autonomous system according to circumstance.

The takeaway for design is that potential end-users do not expect autonomous systems to act independently of user input. End-users expect that autonomous systems will be responsive to context and act in timely fashion, where *timely* means they are responsive not only to what happens *here and now* but also what happens over time. It is thus expected that autonomous systems will gear their behaviours in with occasional and established patterns of human conduct, and that mechanisms be provided to enable users to configure autonomous behaviours around human schedules to enable effective orchestration and synchronisation. It is also expected that users will be able to exercise vary levels of control over the behaviour of autonomous systems in order to make them responsive to the inevitable contingencies of everyday life in the home.

7.2.2 Computational accountability

As noted above, computational accountability refers to the legibility or intelligibility of system behaviours. To borrow from our participants, it will not do for the behaviours of autonomous systems to appear "seemingly random", they must be "accounted for". The expectation that the behaviours of autonomous systems will be accountable has close parallels with Dourish and Button's (1998) notion of *translucency*, which emphasises the importance of a system's ability to give accounts of its own behaviour in order to render its actions legible and thus better support human–computer interaction.

For Dourish and Button, computational accountability is about making the inner workings of the "black box" accountable to users. They provide an example of file copying, replacing a general progress bar that glosses computational behaviour with data buckets and the articulation of flow strategies to elaborate the point. A number of studies (e.g. Hyland et al., 2018; Crabtree et al., 2018; Stumpf et al., 2012) have subsequently established that revealing more of what goes on under the hood of technological systems is often needed to avoid a range of problems, including trust-related issues, that could otherwise negatively impact the user's overall experience. However, when confronted with hypothetical systems operating more or less autonomously, our participants' expectations about computational accountability operate at a different level. Our participants were not so much interested in making the opaque inner workings of autonomous systems accountable, as they were in making their observable behaviours accountable.

Our participants thus expect transparency to be built into autonomous behaviours, where transparency means the grounds of behaviour are visible and available to account. The grounds of behaviour were spoken about in terms of *motive* and *reason*, where the former articulates what occasions behaviour and the latter articulates what is done by an autonomous system in response to motive. Previous enquiries into autonomous systems, such as a study by Lim et al. (2009), have stressed the importance of computer systems being able to articulate the *why* behind autonomous

actions. Our results expand on these findings and paint a broader picture. Of particular concern to our participants is the articulation of on whose behalf autonomous behaviour is occasioned, and the commensurate expectation that whatever is being done is being done to serve and meet end-user's need. Consequently our participants expect *simple* accounts of autonomous behaviour, i.e. accounts that articulate why is something being done and on whose behalf.

Although our findings shift the focus away from accounts articulating the inner workings of autonomous machines, we do find resonance in our participants' talk with Dourish and Button's notion of accountability (1998) in the way in which accountability should be expressed. As Dourish and Button (Ibid.) put it, "what is important … is not the account itself (the explanation of the system's behaviour) but rather accountability in the way this explanation arises. In particular, the account arises reflexively in the course of action, rather than as a commentary upon it".

Our participants speak of autonomous systems *telling* users the motives and reasons for their behaviour as a preface to and/or in vivo feature of that behaviour, rather than something that is bolted on after the fact to provide a post hoc explanation. Thus, and as Dourish and Button put it, computational accountability "becomes part and parcel of ordinary interaction with a computer system".

The takeaway for design is not simply that autonomous behaviours should be accountable to users. Rather, in expressing taken for granted background expectancies at work in domestic life, our participants have articulated what an account should consist of and look like. Thus, we find that the accountability of autonomous behaviours should not be concerned with the inner workings of autonomous machines but should instead articulate motives and reasons, detailing what are to be done, why and on whose behalf in particular. Furthermore, the articulation of motives and reasons should occur as preface to and/or in vivo feature of autonomous behaviour, rather than an after the fact explanation.

7.2.3 Social accountability

Social accountability moves beyond the expectation that autonomous systems will make their behaviours accountable to end-users to instead address expectations regarding the appropriateness of autonomous behaviour. Simply put, autonomous behaviour may be intelligible to end-users but that does not mean they will find it appropriate or acceptable. Acceptability is of longstanding concern in systems design. The technology acceptance model, or TAM, is often cited as a key approach to determining acceptability, being designed to measure a prototype's perceived *ease of use* and *usefulness* (Davis, 1985). While such an approach has been said to help maximise commercial success of novel systems (Davis et al., 1989), the TAM framework has also been a target of criticism for its overly generic nature and limited scope (Lindley et al., 2017).

Alternatively, user experience or UX models focus on *interface quality* and its impact on the "experiential component" of system use (Hassenzahl and Tractinsky, 2006; Hornbæk and Hertzum, 2017), which leads more generally to a concern with usability in HCI. However, as Kim (2015) points out, usability is only one aspect of user acceptance, a point underscored by Lindley et al. (2017), who also note that usability studies rarely look beyond prototypical implementations to consider broader challenges of adoption in everyday life. Indeed, as our user studies demonstrate, domestic autonomous systems introduce a range of social concerns that extend well beyond usability, interface quality, usefulness and ease of use to the *fit* of machine actions with social expectancies. The need for social accountability introduces a broader lens for considering acceptance, encompassing not just the *product* itself, but also the social circumstances within which it will be embedded and the implications of this for design.

At first glance, it may be thought that social accountability is an external prerogative, something that cannot be built into autonomous systems insofar as it turns on user perceptions of what constitutes appropriate behaviour and is therefore a subjective matter. However, this is not the case, (a) because appropriateness is an

intersubjective (social) matter as clearly articulated in the background expectancies our participants share, and (b) in articulating those background expectancies, it became evident that there is much for design to do in terms of supporting or enabling social accountability.

Thus, from a design perspective, it is clear that our participants expect autonomous systems to be responsive to the particular social circumstances in which their behaviours are embedded. There is a need then for autonomous systems to take what people are doing into account, not opening windows for example when human behaviour might be disrupted by such an action, and to tailor autonomous behaviours around the social context in which they operate. In other words, in addition to being autonomous, such systems also need to be context aware if their behaviours are to be seen and treated not only as intelligible but as intelligibly appropriate given the specific social conditions in which their actions are situated.

A key expectation in this regard is that autonomous systems effectively exhibit social competence. There is little to be obtained but trouble from automatically reordering foodstuffs, for example, if it is done without respect to the social consequences of doing so, and not in general but for the particular group of users that inhabits a particular home. As explained by Suchman (2007), although autonomous systems may never be capable of social reasoning the way we humans do, their actions do increasingly carry social consequences. Thus, autonomous systems need to act with respect to human agency and entitlement and tailor their behaviour around the differential rights and privileges at work in the home. Autonomous systems will in effect need to become social agents whose actions comply with the mundane expectations governing domestic life if they are to assume a trusted place in within it. This will not be easy, nor devoid of risks. While acknowledging that the boundaries separating machines from their human users are increasingly blurred, Suchman (2007) warns that this development does not simply mean that machine autonomy shifts the site of agency "from people to their machine progeny." Instead, she argues, the introduction of autonomous systems tilts the balance of power towards the human technologists that create and manage these systems.

There is a danger, in other words, that designers of autonomous systems become increasingly implicated in shaping inter-human relationships in the home. The point Suchman makes is that "the price of recognizing the agency of artefacts need not be the denial of our own". For, as she adds, agency resides "neither in us, nor in our artefacts, but in our intra-actions".

It is hardly surprising then that, as our studies demonstrated, trustworthiness of autonomous systems constitutes a critical expectation, which is closely interrelated with systems demonstrably acting in the end-user's best interests (Pitts, 2016). This is not only a matter of control mechanisms enabling users to manually gear their systems in with a social setting, nor computational accountability and making autonomous behaviours intelligible to people. It is also and effectively a matter of making it visible that actions done are done for *you* and not, for example, for the benefit of the system provider. Thus in addition to exhibiting social competence, the behaviours of autonomous systems must also be accountable to end-user's interests if they are to assume trusted status. Of particular note here is the accountability of personal data—the oil that lubricates the autonomous machine—and transparency of data flows, coupled with tools to enable end-users to limit them and even close them off if it is deemed that the machine is not acting in their interests.

The takeaway for design is that social accountability is distinct from computational accountability: the latter speaks to the intelligibility of autonomous behaviours, the former to their appropriateness and acceptability. Social accountability brings with it the need to build context awareness and trust into autonomous systems. Context awareness is needed to enable autonomous systems to respond appropriately to the particular social circumstances in which their behaviours are embedded, and trust is an essential condition of their uptake in everyday life. It requires that computational agency exhibit social competence and that autonomous behaviour complies with the differential rights and privileges at work in any particular home (i.e. not generalised rights and privileges but situationally specific rights and privileges). Trust also requires that autonomous behaviours are accountable to end-user's interests and turns on the transparency of data flows and ability to control them.

7.3 Design Challenges

The problem areas elaborated above provide us with a broad overview of the key human barriers to social acceptability of future domestic autonomous systems. One of the goals of this thesis was to go a step further and help generate more actionable insights that designers of future systems could draw on. Specifically, we said we wanted to find out:

- *R3: What are the key design challenges that ought to be tackled in order to improve the level of acceptability of domestic autonomous systems?*

The background expectancies articulated by our participants elaborate several distinct design challenges for the development of autonomous systems for domestic use. These can be formulated as follows:

- Building coordination mechanisms into autonomous systems to enable users to configure autonomous behaviours around occasional and established patterns of human conduct in the home. While home automation has made significant progress, multi-occupancy is a remaining challenge that has rendered for example a "learning thermostat" virtually unusable for families (Yang and Newman, 2012).

- Building control mechanisms into autonomous systems to enable users to exercise vary levels of control over the behaviour of autonomous systems in the home. While for example occupancy-sensitive home automation has been explored (Scott et al., 2011), our work seeks to draw attention to the numerous remaining challenges concerning how to best bring the human back into the loop (Fischer et al., 2017).

- Building accountability into the behaviours of autonomous systems by articulating motives and reasons for autonomous behaviours as a preface to and in vivo feature of those behaviours. What we mean here is different from the literature stressing the need to reveal inner workings of a system (e.g. Dourish and Button, 1998); instead we are concerned with the overarching motivations and agendas that drive an autonomous system's decision-making.

- Building in transparency into the relationship between users and the system provider. As our research demonstrated, a system acting in an ambiguous manner can easily give rise to suspicions regarding ulterior motives. A clearly articulated ownership of the system and the agenda of the system provider will go a long way towards dispersing (or confirming) any such trust issues and in the process help combat the rising influence technologists over our lives (Suchman, 2007).

- Building context awareness into autonomous behaviours to enable autonomous systems to respond appropriately to the particular social circumstances in which their behaviours are embedded. Drawing on a rich body of existing research in context-aware systems (e.g. Bellotti and Edwards, 2001), the unique challenge is the interactional nature of context articulated most prominently by Dourish (2004).

- Building social competence into computational agency to ensure that autonomous behaviour complies with the differential rights and privileges at work in the home in order to engender end-user's trust. This challenge is distinct from existing work on roles for example in multi-agent systems (Ramchurn et al., 2004) in that design solutions need to respond to the enacted and fluid ways in which rights and privileges are negotiated on an ongoing basis.

- Building transparency into the data flows that drive autonomous behaviours and data flow controls to further engender user trust. This design challenge can build on initial work in human–data interaction (Mortier et al., 2016), putting forward new models of personal data aligned with GDPR.

These are not requirements for autonomous systems in that they do not specify just what should be done or built to address them. Rather, they elaborate problem spaces and topics for design to explore. While we are aware of ongoing research that touches upon the above issues in various ways (e.g. Dourish and Bell, 2014; Crabtree et al., 2016; Mortier et al., 2016; Suchman and Weber, 2016), it is not clear, for example, what it would mean in practice to articulate the motives and reasons for autonomous behaviours in the in vivo course of their performance. Would an account have to be provided every time some behaviour occurred or only sometimes and only with respect to certain behaviours? It can be readily anticipated that constant articulation of motives and reasons would become an annoying nuisance having a negative impact on the acceptability of autonomous systems, particularly where relatively trivial behaviours are concerned.

The problem of course is that it is nigh impossible to say what constitutes *trivial* (or significant) behaviour in the absence of social context. Thus, a key research challenge here lies not only in building accountability into autonomous behaviours but also in working out how to best support the delivery of accounts to end-users. Ditto building coordination, control, context awareness, transparency and social competence into autonomous systems, which is to say that what any of these topics might look like and amount to in practice has yet to be determined and can only be determined through significant research effort.

Nonetheless, there is sufficient generality built into these design challenges for them to be widely applied in the design of autonomous systems. They might, in effect, be turned into a basic set of design guidelines or fundamental questions such that on any occasion of building an autonomous system, developers might ask themselves if

their designs respond to them. For example, in designing an autonomous grocery system its developers might ask the following:

- Does the system give an account of what motivates its behaviour to the user and the reasons for carrying out particular actions [e.g. that it is ordering XYZ grocery items because you are out of stock and they are the best deal available]?

- Does the system respond to the social circumstances in which it is situated [e.g. presenting accounts of a shopping order at situationally relevant times and places]?

- Does the system display social competence to users in executing its behaviours [e.g. not automatically reordering foodstuffs just because they have run out]?

- Does the system make the data it uses transparent and allow users to control its flow [e.g. not "sharing" grocery data with large supermarkets and thereby curtailing the flow of adverts and offers?]

- Does the system allow users to coordinate their patterns of behaviour with the system's behaviour [e.g. to 'share' their calendar with the systems as a resource for scheduling reordering]?

- Does the system provide users with granular choices over levels of intelligence and autonomy [e.g. allowing users to delegate certain aspects of grocery shopping to the systems and to retain others for themselves]?

- Does the system give an account of what motivates its behaviour to the user and the reasons for carrying out particular actions [e.g. that it is ordering XYZ grocery items because you are out of stock and they are the best deal available]?

- Does the system respond to the social circumstances in which it is situated [e.g. presenting accounts of a shopping order at situationally relevant times and places]?

- Does the system display social competence to users in executing its behaviours [e.g. not automatically reordering foodstuffs just because they have run out]?

- Does the system make the data it uses transparent and allow users to control its flow [e.g. not "sharing" grocery data with large supermarkets and thereby curtailing the flow of adverts and offers?]

The brackets [] may of course be removed or, perhaps more to the point, their content replaced with specifics concerning the autonomous system at hand. For as with triviality and significance, the question as to what constitutes "respond to social circumstances", "display social competence", "make data use transparent", "coordinate with patterns of behaviour" and "provide granular choices" are matters that will need to be worked out with reference to the specificities of an autonomous system and the particular social context into which it is to be situated and used. However, that does not mean the questions cannot be asked, nor answers sought and found.

By way of example, let us look at the Nest smart thermostat. The Nest was amongst the first thermostats targeting the mainstream consumer that based its operation around machine learning algorithms (Nest, 2020). The basic underlying idea was that by monitoring and inferring patterns in the behaviour of home occupants, the

Nest thermostat would be able to autonomously create and implement a personalized heating and cooling schedule. By subsequently regulating heating on behalf of its users, the thermostat would make the life of the occupants more convenient as well as helping to generate energy savings.

While the Nest was well-received overall, existing research has shown that its intelligent features were not perceived to be as useful or intuitive as expected, with users frequently struggling to understand what the Nest was attempting to learn about them and how exactly it was using its acquired knowledge to control their home's temperature (Yang and Newman, 2013).

Although the Nest thermostat was not directly dealt with in the course of our research, this does not hinder us from speculating regarding the manner in which the above outlined design challenges could be utilized to enhance its overall social acceptability.

- It seems reasonable to assume that dealing with multi-occupancy represents a key issue here. Inconsistent or irregular patterns in behaviour can in such cases easily lead to erroneous decisions by the thermostat. While context-aware technology may not be sufficiently competent to deal with such fluctuations, it is imperative that a sense of coordination is maintained by the system alerting its users when facing high levels of uncertainty.

- Correspondingly, any control mechanism must be sufficiently robust to enable fine-grained control over the system's behaviour and, if required, the correction of its actions. Rather than simply turning heating up or down, the Nest needs to allow its users to manually adjust any of the inferred patterns along with the resulting schedules.

- Much of the confusion regarding ambiguous system behaviour could be alleviated by the Nest articulating the motives as well as reasons behind its actions. Thus a drop in temperature should be accompanied by an account for what occasions the action (e.g. reducing energy consumption) as well as the

particular actions done with respect to motive (e.g. lowering temperature in the living room and the kitchen). It is crucial that such accounts are provided as a preface or in vivo, rather than notifying the users retrospectively.

- While the expectancies manifested by our participants suggested that a sense of transparency would also require the grounds upon which autonomous actions are taken to be made visible, we have argued that systems should strive to achieve a 'simple' level of transparency. Rather than revealing the system's inner working, it is the overarching motivation that ought to be communicated. Thus the motives and reasons concerning the lowering of temperature, for example, ought to be explained by being accompanied by a simple explanation (e.g. low occupancy detected).

- Building context awareness along with social competence into the system represents another key requirement underpinning social acceptability. It seems reasonable, for example, that all family members might not possess the technical knowledge required to configure the thermostat settings. Toddlers, for instance, should arguably be excluded from interacting with the thermostat. While computer vision and other context-aware technologies might be applicable, this particular case could be solved using a simple password protection mechanism.

- To establish a sense of trust, the user ought to be able to see all flows of personal data. What data exactly is being collected? Where is it stored and how is it used? Are there any ulterior motives involved (e.g. Nest service provider selling occupancy data to third parties who in turn utilize them for targeted advertising). Clearing out any doubts regarding such questions is imperative for the sake of social accountability. A simple service agreement might go a long way here.

The guidelines above are not meant to be taken literally, as we have not carried out any actual design and evaluation work featuring the Nest. Rather, they are intended to serve as a simple case study illustrating the ways through which our design challenges could help shape real world systems. It seems evident that each of our four key factors: coordination, control, computational accountability and social accountability would play a crucial role, which goes some way to validate and show the applicability of the findings of this thesis.

Chapter 8 Conclusions

People's expectations of domestic autonomous systems are poorly understood at this point in time. We have therefore sought to engage prospective users from the early stages of design, not least because it is widely acknowledged that social expectations concerning the intelligibility and trustworthiness of autonomous systems are as important to their adoption as proposed technological benefits. Our engagement with prospective users exploits a novel design methodology that combines traditional scenario-based design with a contra-vision approach to develop contrasting visions of the autonomous home that are designed to breach common sense reasoning and surface taken for granted background expectancies that impact the adoption of future and emerging technologies.

This methodological contribution may be exploited more generally but in the case reported here, the forward-looking visions built into our breaching experiments are based on application areas where autonomous systems are expected to make a substantial impact on domestic life. The utopian and dystopian characters of these visions disturb common sense reasoning and create a reality disjuncture that warrants repair. It is in repairing or resolving the incongruity between utopian and dystopian visions that the usually unspoken background expectancies that people use to make sense of and order familiar settings and scenes in everyday life,

including the uses of technology, are articulated and expressed or made visible and thus become available to our consideration.

The substantive contribution of this paper thus reveals that people expect coordination and control mechanisms along with computational and social accountability to be built into domestic autonomous systems. Coordination means it is expected that autonomous systems will gear in with discrete patterns of conduct in the home. Control means that it is expected that the intelligence built into autonomous systems will be subject to granular choice. Computational accountability means it is expected that the behaviours of autonomous systems will be made accountable to end-users. Social accountability means it is expected that the behaviours of autonomous systems will not only be legible but situationally appropriate and responsive to social context.

The results of our breaching experiments do not provide requirements for autonomous systems but acceptability challenges elaborating a number of discrete but interrelated problem spaces and topics that concern the nature of computational agency, data transparency, accountability, context awareness and the role of intelligence in the future smart home. What these challenges amount to in practice has yet to be determined. Nonetheless, the expectations are significant and enable fundamental questions to be asked of autonomous systems during their design. Answering them is central to the widespread adoption of autonomous systems in domestic life.

References

Abdul, Ashraf, Jo Vermeulen, Danding Wang, Brian Y. Lim, and Mohan Kankanhalli. "Trends and trajectories for explainable, accountable and intelligible systems: An hci research agenda." In Proceedings of the 2018 CHI conference on human factors in computing systems, p. 582. ACM, (2018). http://dx.doi.org/10.1145/3173574.3174156

Absurdism. Philosophy. http://www.philosophy-index.com/existentialism/absurd.php. Accessed 6 February 2019.

Adams, Anne, and Anna L. Cox. "Questionnaires, in-depth interviews and focus groups." (2008): 17-34. http://dx.doi.org/10.1017/CBO9780511814570.003

Ahmad, M. W., M. Mourshed, and B. Yuce. "Computational intelligence techniques for HVAC systems: A review. Build. Simul. Vol. 9: 359." (2016). http://dx.doi.org/10.1007/s12273-016-0285-4

Ahmeda, Rana Alaa El-Deen, M. Elemam Shehaba, Shereen Morsya, and Nermeen Mekawiea. "Performance study of classification algorithms for consumer online shopping attitudes and behavior using data mining." In 2015 Fifth International Conference on Communication Systems and Network Technologies, pp. 1344-1349. IEEE, 2015. http://dx.doi.org/10.1109/CSNT.2015.50

Alan, Alper T., Enrico Costanza, Sarvapali D. Ramchurn, Joel Fischer, Tom Rodden, and Nicholas R. Jennings. "Tariff agent: interacting with a future smart energy system at home." ACM Transactions on Computer-Human Interaction (TOCHI) 23, no. 4 (2016): 25. http://dx.doi.org/10.1145/2943770

Albus, J., and P. J. Antsaklis. "Panel discussion: Autonomy in engineering systems: What is it and why is it important? Setting the stage: Some autonomous thoughts on autonomy." In Proceedings of the 1998 IEEE International Symposium on Intelligent Control (ISIC) held jointly with IEEE International Symposium on Computational Intelligence in Robotics and Automation (CIRA) Intell, pp. 520-521. IEEE, (1998).

Aldrich, Frances K. "Smart homes: past, present and future." In Inside the smart home, pp. 17-39. Springer, London, (2003). http://dx.doi.org/10.1007/1-85233-854-7_2

Anjomshoae, Sule, Amro Najjar, Davide Calvaresi, and Kary Främling. "Explainable agents and robots: Results from a systematic literature review." In Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems, pp. 1078-1088. International Foundation for Autonomous Agents and Multiagent Systems, (2019).

Antifakos, Stavros, Nicky Kern, Bernt Schiele, and Adrian Schwaninger. "Towards improving trust in context-aware systems by displaying system confidence." In Proceedings of the 7th international conference on Human computer interaction with mobile devices & services, pp. 9-14. ACM, 2005. http://dx.doi.org/10.1145/1085777.1085780

Baudrillard, Jean. "Marshall MacLuhan, Understanding Media: the Extensions of Man, Mc Graw-Hill Book company, cop. 1964." L'Homme et la Société 5, no. 1 (1967): 227-230.

Barad, Karen. "Getting real: Technoscientific practices and the materialization of reality." Differences: a journal of feminist cultural studies 10, no. 2 (1998): 87-91. http://dx.doi.org/10.1215/9780822388128-006

Bauman, Zygmunt, Didier Bigo, Paulo Esteves, Elspeth Guild, Vivienne Jabri, David Lyon, and Rob BJ Walker. "After Snowden: Rethinking the impact of surveillance." International political sociology 8, no. 2 (2014): 121-144. http://dx.doi.org/10.1111/ips.12048

Bellotti, Victoria, Alexander Ambard, Daniel Turner, Christina Gossmann, Kamila Demkova, and John M. Carroll. "A muddle of models of motivation for using peer-to-peer economy systems." In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, pp. 1085-1094. ACM, 2015. http://dx.doi.org/10.1145/2702123.2702272

Bellotti, Victoria, Maribeth Back, W. Keith Edwards, Rebecca E. Grinter, Austin Henderson, and Cristina Lopes. "Making sense of sensing systems: five questions for designers and researchers." In Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 415-422. ACM, 2002. http://dx.doi.org/10.1145/503376.503450

Bellotti, Victoria, and Keith Edwards. "Intelligibility and accountability: human considerations in context-aware systems." Human–Computer Interaction 16, no. 2-4 (2001): 193-212. http://dx.doi.org/10.1207/S15327051HCI16234_05

Biran, Or, and Courtenay Cotton. "Explanation and justification in machine learning: A survey." In IJCAI-17 workshop on explainable AI (XAI), vol. 8, p. 1. (2017).

Bittner, Egon. "Objectivity and realism in sociology." Phenomenological sociology 110 (1973).

Blythe, Mark. "The hitchhiker's guide to ubicomp: using techniques from literary and critical theory to reframe scientific agendas." Personal and Ubiquitous Computing 18, no. 4 (2014): 795-808. http://dx.doi.org/10.1007/s00779-013-0679-6

Bleecker, J., (2009). Design Fiction: A short essay on design, science, fact and fiction. Near Future Laboratory.

Boden, Margaret A. "The Philosophy of Artificial Intelligence." Oxford University Press (1990).

Bowers, John, and Tom Rodden. "Exploding the interface: Experiences of a CSCW network." In Proceedings of the INTERACT'93 and CHI'93 Conference on Human Factors in Computing Systems, pp. 255-262. ACM, (1993). http://dx.doi.org/10.1145/169059.169205

Braun, Virginia, and Victoria Clarke. "Using thematic analysis in psychology." Qualitative research in psychology 3, no. 2 (2006): 77-101. https://doi.org/10.1191/1478088706qp063oa

Britten, Nicky. "Qualitative interviews in health care research." Qualitative research in health care 2 (1999): 11-19.

Brown, Barry, and Eric Laurier. "The trouble with autopilots: Assisted and autonomous driving on the social road." In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pp. 416-429. 2017. http://dx.doi.org/10.1145/3025453.3025462

Brown, Anna, Alexandra Chouldechova, Emily Putnam-Hornstein, Andrew Tobin, and Rhema Vaithianathan. "Toward Algorithmic Accountability in Public Services: A Qualitative Study of Affected Community Perspectives on Algorithmic Decision-making in Child Welfare Services." In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, p. 41. ACM, (2019). http://dx.doi.org/10.1145/3290605.3300271

Brézillon, P. "Context needs in cooperative building of explanations." In First European Conference on Cognitive Science in Industry, pp. 443-450. 1994.

CNN. 5G review. - https://edition.cnn.com/2019/08/09/tech/5g-review/index.html Accessed: November 29, 2019

Chi, Pei-yu Peggy, Jen-hao Chen, Shih-yen Liu, and Hao-hua Chu. "Designing smart living objects–enhancing vs. distracting traditional human–object interaction." In International Conference on Human-Computer Interaction, pp. 788-797. Springer, Berlin, Heidelberg, (2007). http://dx.doi.org/10.1007/978-3-540-73107-8_87

Costanza, Enrico, Joel E. Fischer, James A. Colley, Tom Rodden, Sarvapali D. Ramchurn, and Nicholas R. Jennings. "Doing the laundry with agents: a field trial of a future smart energy system in the home." In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 813-822. ACM, (2014). http://dx.doi.org/10.1145/2556288.2557167

Coulton, Paul, Joseph Lindley, and Haider Ali Akmal. "Design Fiction: Does the search for plausibility lead to deception?." (2016): 369-384. http://dx.doi.org/10.21606/drs.2016.148

Coulton, Paul, Joseph Galen Lindley, Miriam Sturdee, and Michael Stead. "Design fiction as world building." (2017).

Coulton, Paul, Joseph Galen Lindley, Adrian Ioan Gradinar, James Colley, Neelima Sailaja, Andy Crabtree, Ian Forrester, and Lianne Kerlin. "Experiencing the Future Mundane." (2019).

Crabtree, Andy, Alan Chamberlain, Stela Valchovska, Mark Davies, Kevin Glover, and Chris Greenhalgh. "I've got a sheep with three legs if anybody wants it?: re-visioning the rural

economy." Personal and Ubiquitous Computing 19, no. 8 (2015): 1247-1258. http://dx.doi.org/10.1007/s00779-015-0890-8

Crabtree, Andy, Tom Lodge, James Colley, Chris Greenhalgh, Richard Mortier, and Hamed Haddadi. "Enabling the new economic actor: data protection, the digital economy, and the Databox." Personal and Ubiquitous Computing 20, no. 6 (2016): 947-957. http://dx.doi.org/10.1007/s00779-016-0939-3

Crabtree, Andy, Tom Lodge, James Colley, Chris Greenhalgh, and Richard Mortier. "Building accountability into the Internet of Things." Available at SSRN 2881876 (2016). https://doi.org/10.13140/RG.2.2.27512.44803

Crabtree, Andy, Tom Lodge, James Colley, Chris Greenhalgh, Kevin Glover, Hamed Haddadi, Yousef Amar et al. "Building accountability into the Internet of Things: the IoT Databox model." Journal of Reliable Intelligent Environments 4, no. 1 (2018): 39-55. http://dx.doi.org/10.1007/s40860-018-0054-5

Crabtree, Andy, Tom Rodden, Terry Hemmings, and Steve Benford. "Finding a Place for UbiComp in the Home." In International Conference on Ubiquitous Computing, pp. 208-226. Springer, Berlin, Heidelberg, (2003). http://dx.doi.org/10.1007/978-3-540-39653-6_17

Crabtree, Andy, and Tom Rodden. "Domestic routines and design for the home." Computer Supported Cooperative Work 13, no. 2 (2004): 191-220. http://dx.doi.org/10.1023/B:COSU.0000045712.26840.a4

Crandall, Jacob W., Michael A. Goodrich, Curtis W. Nielsen, and Dan R. Olsen Jr. "Validating human–robot interaction schemes in multitasking environments." (2005). http://dx.doi.org/10.1109/TSMCA.2005.850587

Datta, A., Sen, S., & Zick, Y. (2016, May). Algorithmic transparency via quantitative input influence: Theory and experiments with learning systems. In Security and Privacy (SP), IEEE Symposium on (pp. 598-617). IEEE. (2016). http://dx.doi.org/10.1109/SP.2016.42

Davis Fred D (1985) A technology acceptance model for empirically testing new end-user information systems: theory and results. Dissertation, Massachusetts Institute of Technology. https://dspace.mit.edu/handle/1721.1/15192. Accessed 6 February 2019

Davis, Fred D., Richard P. Bagozzi, and Paul R. Warshaw. "User acceptance of computer technology: a comparison of two theoretical models." Management science 35, no. 8 (1989): 982-1003. http://dx.doi.org/10.1287/mnsc.35.8.982

De Russis, Luigi, and Fulvio Corno. "Homerules: A tangible end-user programming interface for smart homes." In Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems, pp. 2109-2114. ACM, (2015). http://dx.doi.org/10.1145/2702613.2732795

Dey, Anind K., Timothy Sohn, Sara Streng, and Justin Kodama. "iCAP: Interactive prototyping of context-aware applications." In International Conference on Pervasive Computing, pp. 254-271. Springer, Berlin, Heidelberg, 2006. http://dx.doi.org/10.1007/11748625_16

Diakopoulos, Nicholas. "Algorithmic accountability reporting: On the investigation of black boxes." (2014).

Dictionary, Oxford English. "Oxford english dictionary." Simpson, JA & Weiner, ESC (1989).

Digital Transformation Initiative. "In collaboration with Accenture." Unlocking $100 Trillion for Business and Society from Digital Transformation. WEF JANUARY (2017).

Doran, Derek, Sarah Schulz, and Tarek R. Besold. "What does explainable AI really mean? A new conceptualization of perspectives." arXiv preprint arXiv:1710.00794 (2017).

Dourish, Paul, and Genevieve Bell. "Resistance is futile: reading science fiction alongside ubiquitous computing." Personal and Ubiquitous Computing 18, no. 4 (2014): 769-778. http://dx.doi.org/10.1007/s00779-013-0678-7

Dourish, Paul, and Graham Button. "On "technomethodology": Foundational relationships between ethnomethodology and system design." Human-computer interaction 13, no. 4 (1998): 395-432. http://dx.doi.org/10.1207/s15327051hci1304_2

Dourish, Paul. "Accounting for system behaviour: Representation, reflection and resourceful action." Computers and Design in Context, MIT Press, Cambridge, MA, USA (1997): 145-170.

Dourish, Paul. "What we talk about when we talk about context." Personal and ubiquitous computing 8, no. 1 (2004): 19-30. http://dx.doi.org/10.1007/s00779-003-0253-8

Dunlop, Mark. 'Interactions between Automatic and Autonomous Systems'. Presented at the Systems Engineering for Autonomous Systems Defence Technology Conference, July 2009, Edinburgh.

EPSRC, Artificial Intelligence Technologies, https://epsrc.ukri.org/research/ourportfolio/researchareas/ait/ Accessed: Novembeer 29, 2019

Ecuity. Smart grids, microgeneration and storage: commercialising the benefits. https://www.ecuity.com/wp-content/uploads/2013/10/SGrid-Report-v_final3.0.pdf. Accessed 6 February 2019

Edmunds, Angela, and Anne Morris. "The problem of information overload in business organisations: a review of the literature." International journal of information management20, no. 1 (2000): 17-28. http://dx.doi.org/10.1016/S0268-4012(99)00051-1

Edwards, W. Keith, and Rebecca E. Grinter. "At home with ubiquitous computing: Seven challenges." In International conference on ubiquitous computing, pp. 256-272. Springer, Berlin, Heidelberg, (2001). http://dx.doi.org/10.1007/3-540-45427-6_22

El Naqa, Issam, and Martin J. Murphy. "What is machine learning?." In Machine Learning in Radiation Oncology, pp. 3-11. Springer, Cham, (2015). http://dx.doi.org/10.1007/978-3-319-18305-3_1

Felfernig, Alexander, Michael Jeran, Gerald Ninaus, Florian Reinfrank, and Stefan Reiterer. "Toward the next generation of recommender systems: applications and research challenges." In Multimedia services in intelligent environments, pp. 81-98. Springer, Heidelberg, (2013). http://dx.doi.org/10.1007/978-3-319-00372-6_5

Fischer, Joel E., Chris Greenhalgh, Wenchao Jiang, Sarvapali D. Ramchurn, Feng Wu, and Tom Rodden. "In-the-loop or on-the-loop? Interactional arrangements to support team coordination with a planning agent." Concurrency and Computation: Practice and Experience (2017): e4082. http://dx.doi.org/10.1007/978-3-319-00372-6_5

Friedman, Batya, ed. Human values and the design of computer technology. No. 72. Cambridge University Press, 1997.

Fuentes, Carolina, Martin Porcheron, Joel E. Fischer, Enrico Costanza, Obaid Malilk, and Sarvapali D. Ramchurn. "Tracking the consumption of home essentials." In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, pp. 1-13. 2019. http://dx.doi.org/10.1145/3290605.3300869

Gabrielli, Silvia, and Jan-Christoph Zoels. "Creating imaginable futures: using human-centered design strategies as a foresight tool." In Proceedings of the 2003 conference on Designing for user experiences, pp. 1-14. 2003. http://dx.doi.org/10.1145/997078.997114

Gao, Lingling, and Xuesong Bai. "A unified perspective on the factors influencing consumer acceptance of internet of things technology." Asia Pacific Journal of Marketing and Logistics 26, no. 2 (2014): 211-231. http://dx.doi.org/10.1108/APJML-06-2013-0061

Garfinkel, Harold. Studies in ethnomethodology. Englewood Cliffs, NJ: Prentice-Hall, (1967).

Gebhardt, Jan, Michael Massoth, Stefan Weber, and Torsten Wiens. "Ubiquitous Smart Home Controlling Raspberry Embedded System." In UBICOMM: The Eighth International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies. (2014).

Gill, Paul, Kate Stewart, Elizabeth Treasure, and Barbara Chadwick. "Methods of data collection in qualitative research: interviews and focus groups." British dental journal 204, no. 6 (2008): 291. http://dx.doi.org/10.1038/bdj.2008.192

Goetz, Ernest T., and Mark Sadoski. "Commentary: The perils of seduction: Distracting details or incomprehensible abstractions?." Reading Research Quarterly (1995): 500-511. http://dx.doi.org/10.2307/747628

Gregor, Shirley, and Izak Benbasat. "Explanations from intelligent systems: Theoretical foundations and implications for practice." MIS quarterly (1999): 497-530. http://dx.doi.org/10.2307/249487

Grinter, Rebecca E., W. Keith Edwards, Mark W. Newman, and Nicolas Ducheneaut. "The work to make a home network work." In ECSCW 2005, pp. 469-488. Springer, Dordrecht, (2005). http://dx.doi.org/10.1007/1-4020-4023-7_24

Grudin, Jonathan. "The computer reaches out: the historical continuity of interface design." In Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 261-268. ACM, (1990). http://dx.doi.org/10.1145/97243.97284

Gunge, Vaishnavi S., and Pratibha S. Yalagi. "Smart home automation: a literature review." International Journal of Computer Applications 975 (2016): 8887.

Gunning, David. "Explainable artificial intelligence (xai)." Defense Advanced Research Projects Agency (DARPA), nd Web 2 (2017). http://dx.doi.org/10.1126/scirobotics.aay7120

Happ, Alan J. "Usability foresight: strategic usability planning: a special interest group meeting report." ACM SIGCHI Bulletin 26, no. 1 (1994): 17-21. http://dx.doi.org/10.1145/181526.181527

Harper, Richard. "Inside the smart home: Ideas, possibilities and methods." In Inside the smart home, pp. 1-13. Springer, London, (2003). http://dx.doi.org/10.1007/1-85233-854-7_1

Harrison, Mariel, Julia Baker, Medard Twinamatsiko, and E. J. Milner-Gulland. "Profiling unauthorized natural resource users for better targeting of conservation interventions." Conservation Biology 29, no. 6 (2015): 1636-1646. http://dx.doi.org/10.1111/cobi.12575

Hassenzahl, Marc, and Noam Tractinsky. "User experience-a research agenda." Behaviour & information technology 25, no. 2 (2006): 91-97. http://dx.doi.org/10.1080/01449290500330331

Herlocker, J., Konstan, J. & Riedl, J. Explaining collaborative filtering recommendations. CSCW 2000, 241- 250. (2000). http://dx.doi.org/10.1145/358916.358995

Hindus, Debby. "The importance of homes in technology research." In International Workshop on Cooperative Buildings, pp. 199-207. Springer, Berlin, Heidelberg, 1999. http://dx.doi.org/10.1007/10705432_18

Hornbæk, Kasper, and Morten Hertzum. "Technology acceptance and user experience: A review of the experiential component in HCI." ACM Transactions on Computer-Human Interaction (TOCHI) 24, no. 5 (2017): 33. https://doi.org/10.1145/3127358

Horvitz, Eric. "Principles of mixed-initiative user interfaces." In Proceedings of the SIGCHI conference on Human Factors in Computing Systems, pp. 159-166. ACM, (1999). http://dx.doi.org/10.1145/302979.303030

House Of Lords Select Committee. "Ai in the uk: ready, willing and able." House of Lords 36 (2018).

Hung, Mark. "Leading the iot." Gartner Insights on How to Lead in a Connected World https://www. gartner. com/imagesrv/books/iot/iotEbook_digital. pdf

Hyland, Lewis, Andy Crabtree, Joel Fischer, James Colley, and Carolina Fuentes. "What do you want for dinner?"–need anticipation and the design of proactive technologies for the home." Computer Supported Cooperative Work (CSCW) 27, no. 3-6 (2018): 917-946. http://dx.doi.org/10.1007/s10606-018-9314-4

IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. "Ethically aligned design: A vision for prioritizing human Well-being with autonomous and intelligent systems (A/IS)." (2017).

Ian F. Akyildiz and Ismail H. Kasimoglu. “Wireless sensor and actor networks: research challenges”. In: Ad Hoc Networks 2.4 (2004), pages 351–367. ISSN: 15708705 http://dx.doi.org/10.1016/j.adhoc.2004.04.003

Jensen, Rikke Hagensby, Yolande Strengers, Jesper Kjeldskov, Larissa Nicholls, and Mikael B. Skov. "Designing the desirable smart home: A study of household experiences and energy consumption impacts." In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, p. 4. ACM, (2018). http://dx.doi.org/10.1145/3173574.3173578

Jivani, Mahesh N. "GSM based home automation system using app-inventor for android mobile phone." International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering 3, no. 9 (2014).

Johnson, Matthew, Jeffrey M. Bradshaw, Robert R. Hoffman, Paul J. Feltovich, and David D. Woods. "Seven cardinal virtues of human-machine teamwork: examples from the DARPA robotic challenge." IEEE Intelligent Systems 29, no. 6 (2014): 74-80. http://dx.doi.org/10.1109/MIS.2014.100

Kidd, Cory D., Robert Orr, Gregory D. Abowd, Christopher G. Atkeson, Irfan A. Essa, Blair MacIntyre, Elizabeth Mynatt, Thad E. Starner, and Wendy Newstetter. "The aware home: A living laboratory for ubiquitous computing research." In International Workshop on Cooperative Buildings, pp. 191-198. Springer, Berlin, Heidelberg, 1999. http://dx.doi.org/10.1007/10705432_17

Kim, Hee-Cheol. "Acceptability engineering: the study of user acceptance of innovative technologies." Journal of applied research and technology 13, no. 2 (2015): 230-237. https://doi.org/10.1016/j.jart.2015.06.001

Kiss, Gábor, and Éva Csilla Berecz. "Questions of Security in the World of Autonomous Vehicles." In Proceedings of the 2019 the 5th International Conference on e-Society, e-Learning and e-Technologies, pp. 109-115. ACM, (2019). http://dx.doi.org/10.1145/3312714.3312718

Kitzinger, Jenny. "Qualitative research: introducing focus groups." Bmj 311, no. 7000 (1995): 299-302. https://doi.org/10.1136/bmj.311.7000.299

Kleissl, Jan, and Yuvraj Agarwal. "Cyber-physical energy systems: Focus on smart buildings." In Design Automation Conference, pp. 749-754. IEEE, 2010. http://dx.doi.org/10.1145/1837274.1837464

Kuniavsky, M. Smart things: Ubiquitous Computing User Experience Design. Morgan Kaufmann (2010)

Kurzweil, Ray. The Age of Spiritual Machines. Penguin Books. ISBN 978-0-670-88217-5. (1999)

Kurzweil, Ray. The Singularity is Near. Penguin Books, (2005). http://dx.doi.org/10.1057/9781137349088_26

Kwon, Hyosun, Joel E. Fischer, Martin Flintham, and James Colley. "The connected shower: Studying intimate data in everyday life." Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 2, no. 4 (2018): 1-22. http://dx.doi.org/10.1145/3287054

Lahijanian, Morteza, and Marta Kwiatkowska. "Social trust: a major challenge for the future of autonomous systems." In 2016 AAAI Fall Symposium Series. (2016).

Lee, Jisoo, Luis Garduño, Erin Walker, and Winslow Burleson. "A tangible programming tool for creation of context-aware applications." In Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing, pp. 391-400. ACM, (2013). http://dx.doi.org/10.1145/2493432.2493483

Lehmacher, Wolfgang. The global supply chain: how technology and circular thinking transform our future. Springer International Publishing, Cham, Switzerland. (2017). https://doi.org/10.1007/978-3-319-51115-3

Lessig, Lawrence. Code: And other laws of cyberspace. ReadHowYouWant. com, 2009.

Lim, Brian Y. "Improving understanding and trust with intelligibility in context-aware applications." PhD diss., figshare, (2012).

Lim, Brian Y., Anind K. Dey, and Daniel Avrahami. "Why and why not explanations improve the intelligibility of context-aware intelligent systems." In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 2119-2128. ACM, (2009). https://doi.org/10.1145/1518701.1519023

Lipton, Zachary C. "The mythos of model interpretability." arXiv preprint arXiv:1606.03490 (2016). http://dx.doi.org/10.1145/3233231

Lowe, David. "Concerns Over the Surveillance Society: The Snowden Revelations." Routledge Handbook of Social, Economic, and Criminal Justice (2018).

Luger, George F. Artificial intelligence: structures and strategies for complex problem solving. Pearson education, 2005.

Luo, Suhuai, Jesse Jin, and Jiaming Li. "A smart fridge with an ability to enhance health and enable better nutrition." International Journal of Multimedia and Ubiquitous Engineering 4, no. 2 (2009): 69-80.

Ma, Xiao, Ganna Pogrebna, and Irene Ng. "Smart home, smart things and smart me in the smart city: the hub-of-all-things resource integration and enabling tool (HARRIET)." (2014): 9-6.

Maes, Pattie. "Agents that reduce work and information overload." In Readings in human–computer interaction, pp. 811-821. Morgan Kaufmann, (1995). http://dx.doi.org/10.1016/B978-0-08-051574-8.50084-4

Maier, Juergen. "Made smarter—review 2017." Made Smarter Review (2017).

Malekian, Reza, Kui Wu, Gianluca Reali, Ning Ye, and Kevin Curran. "Cyber-physical systems and context-aware sensing and computing." Computer Networks 117 (2017): 1-4. http://dx.doi.org/10.1016/j.comnet.2017.02.012

Malone, Thomas W. "Intelligent information sharing systems." (1986). http://dx.doi.org/10.1145/22899.22903

Mathiesen, Kay. "The human right to Internet access: A philosophical defense." International Review of Information Ethics 18, no. 12 (2012): 9-22.

Mayer, R. C.; Davis, J. H.; and Schoorman, F. D. An integrative model of organizational trust. Academy of management review 20(3):709–734. (1995). http://dx.doi.org/10.5465/amr.1995.9508080335

Meek, Taylor, Husam Barham, Nader Beltaif, Amani Kaadoor, and Tanzila Akhter. "Managing the ethical and risk implications of rapid advances in artificial intelligence: a literature review." In 2016 Portland International Conference on Management of Engineering and Technology (PICMET), pp. 682-693. IEEE, (2016). http://dx.doi.org/10.1109/PICMET.2016.7806752

Min, Kyeong T., Philip Lundrigan, and Neal Patwari. "IASA-indoor air quality sensing and automation: demo abstract." In Proceedings of the 16th ACM/IEEE International Conference on Information Processing in Sensor Networks, pp. 277-278. ACM, (2017). http://dx.doi.org/10.1145/3055031.3055043

Monostori, László. "Cyber-physical production systems: Roots, expectations and R&D challenges." Procedia Cirp 17 (2014): 9-13. http://dx.doi.org/10.1016/j.procir.2014.03.115

Morgan, David L. The focus group guidebook. Vol. 1. Sage publications, 1997. http://dx.doi.org/10.4135/9781483328164

Morollo KH. Smart homes provide convenience and sustainable living. Style. https://www.scmp.com/magazines/style/article/1780091/smart-homes-provide-convenience-and-sustainable-living. Accessed 6 February 2019

Mortier, Richard, Hamed Haddadi, Tristan Henderson, Derek McAuley, Jon Crowcroft, and Andy Crabtree. "Human-data interaction." (2016).

Muir, Bonnie M. "Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems." Ergonomics 37, no. 11 (1994): 1905-1922. http://dx.doi.org/10.1080/00140139408964957

Muller, Michael J. "Participatory design: the third space in HCI." In Human-computer interaction, pp. 181-202. CRC press, 2009. http://dx.doi.org/10.1201/9781420088892.ch9

National Institute of Standards and Technology (NIST). Autonomy Levels for Unmanned Systems (ALFUS) Framework: Volume 1—Terminology, Version 2.0. NIST Special Publication 1011-I-2.0, (2012).

Nest | Home. http://www.nest.com/

Neumann, Peter G. "The foresight saga, redux." Communications of the ACM 55, no. 10 (2012): 26-29. http://dx.doi.org/10.1145/2347736.2347746

Norman, Don. The design of everyday things: Revised and expanded edition. Basic books, (2013).

Obermeyer, Ziad, and Ezekiel J. Emanuel. "Predicting the future—big data, machine learning, and clinical medicine." The New England journal of medicine 375, no. 13 (2016): 1216. http://dx.doi.org/10.1056/NEJMp1606181

O'Brien, Jon, Tom Rodden, Mark Rouncefield, and John Hughes. "At home with the technology: an ethnographic study of a set-top-box trial." ACM Transactions on Computer-Human Interaction (TOCHI) 6, no. 3 (1999): 282-308. http://dx.doi.org/10.1145/329693.329698

Parasuraman, Raja, Thomas B. Sheridan, and Christopher D. Wickens. "A model for types and levels of human interaction with automation." IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans 30, no. 3 (2000): 286-297. http://dx.doi.org/10.1109/3468.844354

Pasquale, Frank. The black box society. Harvard University Press, (2015).

Peek, Sebastiaan TM, Sil Aarts, and Eveline JM Wouters. "Can smart home technology deliver on the promise of independent living? A critical reflection based on the perspectives of older adults." Handbook of smart homes, health care and well-being (2017): 203-214.

Pilling, Franziska, and Paul Coulton. "Forget the Singularity, its mundane artificial intelligence that should be our immediate concern." The Design Journal 22, no. sup1 (2019): 1135-1146.

Pitts M (2016) We need trust to innovate. Gov.UK: Innovate UK. https://innovateuk.blog.gov.uk/2016/08/17/we-need-trust-to-innovate. Accessed 6 February 2019

Pollner, Melvin. "'The very coinage of your brain': The anatomy of reality disjunctures." Philosophy of the Social Sciences 5, no. 3 (1975): 411-430. http://dx.doi.org/10.1177/004839317500500304

Pollner, Melvin. Mundane reason: Reality in everyday and sociological discourse. Cambridge University Press, 2010.

Pope, Catherine, Sue Ziebland, and Nicholas Mays. "Qualitative research in health care: analysing qualitative data." BMJ: British Medical Journal 320, no. 7227 (2000): 114. http://dx.doi.org/10.1136/bmj.320.7227.114

Press, Larry. "Personal computing: the post-PC era." Communications of the ACM 42, no. 10 (1999): 21-24. http://dx.doi.org/10.1145/317665.317670

Rajkumar, Ragunathan, Insup Lee, Lui Sha, and John Stankovic. "Cyber-physical systems: the next computing revolution." In Design Automation Conference, pp. 731-736. IEEE, (2010). http://dx.doi.org/10.1145/1837274.1837461

Ramchurn, Sarvapali D., Dong Huynh, and Nicholas R. Jennings. "Trust in multi-agent systems." The Knowledge Engineering Review 19, no. 1 (2004): 1-25. http://dx.doi.org/10.1017/S0269888904000116

Risteska Stojkoska, Biljana, Andrijana Popovska Avramova, and Periklis Chatzimisios. "Application of wireless sensor networks for indoor temperature regulation." International Journal of Distributed Sensor Networks 10, no. 5 (2014): 502419. http://dx.doi.org/10.1155/2014/502419

Rodden, Tom A., Joel E. Fischer, Nadia Pantidi, Khaled Bachour, and Stuart Moran. "At home with agents: exploring attitudes towards future smart energy infrastructures." In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1173-1182. ACM, (2013). http://dx.doi.org/10.1145/2470654.2466152

Rzepka, Christine, and Benedikt Berger. "User interaction with AI-enabled systems: a systematic review of IS research." (2018).

Salmerón-García, Javier, Pablo Iñigo-Blasco, Fernando Díaz-del-Río, and Daniel Cagigas-Muñiz. "Study of Communication Issues in Dynamically Scalable Cloud-Based Vision Systems for Mobile Robots." In Robots and Sensor Clouds, pp. 33-52. Springer, Cham, (2016). http://dx.doi.org/10.1007/978-3-319-22168-7_2

Sardar, Z. (2010) The Namesake: Futures; futures studies; futurology; futuristic; Foresight -- What's in a name? Futures, 42 (3), pp. 177–184 http://dx.doi.org/10.1016/j.futures.2009.11.001

Schaeffer, Eric. Industry X. 0: Realizing digital value in industrial sectors. Kogan Page Publishers, (2017).

Schaller, Robert R. "Moore's law: past, present and future." IEEE spectrum 34, no. 6 (1997): 52-59. http://dx.doi.org/10.1109/6.591665

Schneider, J. (2003). Commonalities and Differences in Design and Foresight. Design for Future Needs –European Union Project Report, http://www.dffn.org.26

Schwab, Klaus. The fourth industrial revolution. Currency, (2017).

Schwab, Klaus. The fourth industrial revolution: what it means, how to respond, World Economic Forum, Davos, January, 14; (2016).

Scott, James, A. J. Bernheim Brush, John Krumm, Brian Meyers, Michael Hazas, Stephen Hodges, and Nicolas Villar. "PreHeat: controlling home heating using occupancy prediction." In Proceedings of the 13th international conference on Ubiquitous computing, pp. 281-290. ACM, (2011). http://dx.doi.org/10.1145/2030112.2030151

Seabra, Jorge C., Mario A. Costa, and Mateus M. Lucena. "IoT based intelligent system for fault detection and diagnosis in domestic appliances." In 2016 IEEE 6th International Conference on Consumer Electronics-Berlin (ICCE-Berlin), pp. 205-208. IEEE, (2016). http://dx.doi.org/10.1109/ICCE-Berlin.2016.7684756

Shafiq, Zubair Ahmed, and Asif Muhammad. "A survey on Cloud Computing Security Challenges and Issues." (2015).

Shalev-Shwartz, Shai, and Shai Ben-David. “Understanding machine learning: From theory to algorithms”. Cambridge university press, (2014). http://dx.doi.org/10.1017/CBO9781107298019

Sheridan, Thomas B., and William L. Verplank. Human and computer control of undersea teleoperators. Massachusetts Inst of Tech Cambridge Man-Machine Systems Lab, (1978).

Sillitto, Hillary, Dov Dori, Regina M. Griego, Scott Jackson, Daniel Krob, Patrick Godfrey, Eileen Arnold, James Martin, and Dorothy McKinney. "Defining “system”: A comprehensive approach." In INCOSE International Symposium, vol. 27, no. 1, pp. 170-186. (2017). http://dx.doi.org/10.1002/j.2334-5837.2017.00352.x

Smart home market by product: global forecast to 2023. Market and Markets. https://www.marketsandmarkets.com/Market-Reports/smart-homes-and-assisted-living-advanced-technologie-and-global-market-121.html. Accessed 6 February 2019

Smith, Brian Cantwell. "Reflection and semantics in Lisp." In Proceedings of the 11th ACM SIGACT-SIGPLAN symposium on Principles of programming languages, pp. 23-35. ACM, (1984). http://dx.doi.org/10.1145/800017.800513

Stead, Michael Robert, Paul Coulton, and Joseph Galen Lindley. "Do-It-Yourself Medical Devices: Exploring Their Potential Futures Through Design Fiction." (2018): 2511-2523. http://dx.doi.org/10.21606/drs.2018.475

Steering Committee for Foundations in Innovation for Cyber-Physical Systems. Foundations for innovation: strategic opportunities for the 21st century cyber-physical systems-connecting computer and information systems with the physical world. Technical report, National Institute of Standards and Technology (NIST), (2013).

Stumpf, Simone, Margaret Burnett, Volkmar Pipek, and Weng-Keen Wong. "End-user interactions with intelligent and autonomous systems." In CHI'12 Extended Abstracts on Human Factors in Computing Systems, pp. 2755-2758. ACM, 2012. http://dx.doi.org/10.1145/2212776.2212713

Subashini, Subashini, and Veeraruna Kavitha. "A survey on security issues in service delivery models of cloud computing." Journal of network and computer applications 34, no. 1 (2011): 1-11. http://dx.doi.org/10.1016/j.jnca.2010.07.006

Suchman, Lucy, and Jutta Weber. "Human-machine autonomies." Autonomous weapons systems: Law, ethics, policy (2016): 75-102. https://doi.org/10.1017/CBO9781316597873.004

Suchman, Lucy. "Human/machine reconsidered." Cognitive Studies 5, no. 1 (1998): 1_5-1_13.

Suchman, Lucy. "Supporting articulation work." Computerization and controversy: Value conflicts and social choices 2 (1996): 407-423. http://dx.doi.org/10.1016/B978-0-12-415040-9.50118-4

Suchman, Lucy. "Human-machine reconfigurations: Plans and situated actions". Cambridge University Press, (2007). http://dx.doi.org/10.1017/CBO9780511808418

Suermondt, Henri J., and Gregory F. Cooper. "An evaluation of explanations of probabilistic inference." Computers and Biomedical Research 26, no. 3 (1993): 242-254. http://dx.doi.org/10.1006/cbmr.1993.1017

Sun, Daniel, Chun Zhang, Lihui Zhang, Fangxi Chen, and Zhong-Ren Peng. "Urban travel behavior analyses and route prediction based on floating car data." Transportation Letters 6, no. 3 (2014): 118-125. http://dx.doi.org/10.1179/1942787514Y.0000000017

Sundar, S. Shyam, Xue Dou, and Sangmee Lee. "Communicating in a ubicomp world: interaction rules for guiding design of mobile interfaces." In IFIP Conference on Human-Computer Interaction, pp. 730-747. Springer, Berlin, Heidelberg, 2013. http://dx.doi.org/10.1007/978-3-642-40480-1_51

Tanti, Adrian, and Dimitrios Buhalis. "Connectivity and the consequences of being (dis) connected." In Information and communication technologies in tourism 2016, pp. 31-44. Springer, Cham, (2016). http://dx.doi.org/10.1007/978-3-319-28231-2_3

Techpedia, 2019 https://www.techopedia.com/definition/7969/knowledge-based-system-kbs Accessed: November 29, 2019

Toffler, Alvin. "Future shock, 1970." Sydney. Pan (1970).

Tolmie, Peter, James Pycock, Tim Diggins, Allan MacLean, and Alain Karsenty. "Unremarkable computing." In Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 399-406. ACM, 2002. http://dx.doi.org/10.1145/503376.503448

Totman, Scott. "Autonomous versus automated: What each means and why it matters", 2019 https://www.techrepublic.com Accessed: November 29, 2019

Trendmicro, Cybercrime and Digital Threats. https://www.trendmicro.com/vinfo/us/security/news/cybercrime-and-digital-threats/national-cyber-security-awareness-month-the-enterprise-s-safety-online-is-everyone-s-business Accessed: November 29, 2019

Turner, Sam. CTO at HVM Catapult http://hvm.catapult.org.uk/news/ati-guest-blog-digital-engineering-and-manufacturing-leadership/

Twidale, Michael, David Randall, and Richard Bentley. "Situated evaluation for cooperative systems." In Proceedings of the 1994 ACM conference on Computer supported cooperative work, pp. 441-452. ACM, 1994. https://doi.org/10.1145/192844.193066

UK Ministry of Defence (MOD). The UK Approach to Unmanned Aircraft System: Joint Doctrine Note 2/11, (2012).

Ur, Blase, Elyse McManus, Melwyn Pak Yong Ho, and Michael L. Littman. "Practical trigger-action programming in the smart home." In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 803-812. ACM, (2014). http://dx.doi.org/10.1145/2556288.2557420

Ur, Blase, Melwyn Pak Yong Ho, Stephen Brawner, Jiyun Lee, Sarah Mennicken, Noah Picard, Diane Schulze, and Michael L. Littman. "Trigger-action programming in the wild: An analysis of 200,000 ifttt recipes." In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pp. 3227-3231. 2016. http://dx.doi.org/10.1145/2858036.2858556

Vagia, Marialena, Aksel A. Transeth, and Sigurd A. Fjerdingen. "A literature review on the levels of automation during the years. What are the different taxonomies that have been proposed?." Applied ergonomics 53 (2016): 190-202. http://dx.doi.org/10.1016/j.apergo.2015.09.013

Vakkuri, Ville, and Pekka Abrahamsson. "The Key Concepts of Ethics of Artificial Intelligence." In 2018 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC), pp. 1-6. IEEE, (2018). http://dx.doi.org/10.1109/ICE.2018.8436265

Voosen, Paul. "The AI detectives." (2017): 22-27. http://dx.doi.org/10.1126/science.357.6346.22

Weitzner, Daniel J., Harold Abelson, Tim Berners-Lee, Chris Hanson, James Hendler, Lalana Kagal, Deborah L. McGuinness, Gerald Jay Sussman, and K. Krasnow Waterman. "Transparent accountable data mining: New strategies for privacy protection." (2006).

Whitworth, Brian. "Polite computing." Behaviour & Information Technology 24, no. 5 (2005): 353-363. http://dx.doi.org/10.1080/01449290512331333700

Weiser, Mark. "The Computer for the 21 st Century." Scientific american 265, no. 3 (1991): 94-105. http://dx.doi.org/10.1145/329124.329126

Weiser, Mark, Rich Gold, and John Seely Brown. "The origins of ubiquitous computing research at PARC in the late 1980s." IBM systems journal 38, no. 4 (1999): 693-696. http://dx.doi.org/10.1147/sj.384.0693

Whitworth, Brian. "Polite computing." Behaviour & Information Technology 24, no. 5 (2005): 353-363. https://doi.org/10.1080/01449290512331333700

Wigmore, Ivy. "Internet of Things", Techtarget, https://internetofthingsagenda.techtarget.com/definition/Internet-of-Things-IoT Accessed: Novembeer 29, (2019)

Williams, Andrew P., and Paul D. Scharre, eds. "Autonomous systems: Issues for defence policymakers". HQ SACT, (2015).

Williams, Andrew. "Defining autonomy in systems: Challenges and solutions." Issues for Defence Policymakers (2015): 27.

Williams, Bronwyn T. "The face in the mirror, the person on the page." Journal of Adolescent & Adult Literacy 47, no. 2 (2003): 178-182.

Woods, David D., Emily S. Patterson, and Emilie M. Roth. "Can we ever escape from data overload? A cognitive systems diagnosis." Cognition, Technology & Work 4, no. 1 (2002): 22-36. http://dx.doi.org/10.1007/s101110200002

Wooldridge, Michael, and Nicholas R. Jennings. "Intelligent agents: Theory and practice." The knowledge engineering review 10, no. 2 (1995): 115-152. http://dx.doi.org/10.1017/S0269888900008122

Wynn, D. C., C. M. Eckert, and P. J. Clarkson. "Modelling and simulating iterative development processes." In DS 35: Proceedings ICED 05, the 15th International Conference on Engineering Design, Melbourne, Australia, 15.-18.08. 2005. 2005.

Yang, Rayoung, Mark W. Newman, and Jodi Forlizzi. "Making sustainability sustainable: challenges in the design of eco-interaction technologies." In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 823-832. ACM, 2014. http://dx.doi.org/10.1145/2556288.2557380

Yang, Rayoung, and Mark W. Newman. "Living with an intelligent thermostat: advanced control for heating and cooling systems." In Proceedings of the 2012 ACM Conference on Ubiquitous Computing, pp. 1102-1107. ACM, 2012. https://doi.org/10.1145/2370216.2370449

You, Q., Jin, H., Wang, Z., Fang, C., Luo, J.: "Image captioning with semantic attention". In: Proc. of Computer Vision and Pattern Recognition. pp. 4651–4659 (2016) http://dx.doi.org/10.1109/CVPR.2016.503

Young, Emilie, and Russell Greenlee. "Participatory video prototyping." In Posters and short talks of the 1992 SIGCHI conference on human factors in computing systems, pp. 28-28. ACM, 1992. http://dx.doi.org/10.1145/1125021.1125047

Zanella, Andrea, Nicola Bui, Angelo Castellani, Lorenzo Vangelista, and Michele Zorzi. "Internet of things for smart cities." IEEE Internet of Things journal 1, no. 1 (2014): 22-32. http://dx.doi.org/10.1109/JIOT.2014.2306328

Zeng, Yi, Enmeng Lu, and Cunqing Huangfu. "Linking Artificial Intelligence Principles." arXiv preprint arXiv:1812.04814 (2018).

Zion Market Research, "Smart Home Market (Smart Kitchen, Security & Access Control, Lighting Control, Home Healthcare, HVAC Control and Others): Global Industry Perspective, Comprehensive Analysis and Forecast, 2016-2022," Zion Market Research (2017). https://globenewswire.com/newsrelease/2018/01/03/1281338/0/en/Global-Smart-HomeMarket-to-Exceed-53-45-Billion-by-2022-Zion-MarketResearch.html. Accessed : November 29, 2019

Zou, Cong, and Zhenzhong Chen. "Joint latent factors and attributes to discover interpretable preferences in recommendation." Information Sciences 505 (2019): 498-512. http://dx.doi.org/10.1016/j.ins.2019.07.064

END