Re-Thinking Smartness:
Designing More Ethical Connected Devices for the Home

William Seymour
Kellogg College
University of Oxford

Thesis submitted for the Degree of Doctor of Philosophy
Michaelmas 2020

Statement of Authorship & Acknowledgements

Previous forms of much of the work in this thesis have been published as co-authored papers. On each of these papers I was the principle author, and contributed the majority of the writing that appears in them. This also applies to the rest of the research process, with me being either wholly or primarily responsible for planning studies, collecting data, and performing analysis. Where content in this thesis has previously appeared in one of these papers, this is clearly marked at the beginning of the chapter.

But science is not done alone. I owe a huge debt of gratitude to my supervisor, Max Van Kleek, who has given me so much good advice, pushed me to do better when I grew complacent, and picked me up when I fell. He and my other co-authors have helped to guide and shape my work, teaching me how to plan and execute effective research projects. None of the work in this thesis would have been possible without them. I've also been fortunate enough to work with other amazing people in the human-centred computing group and further afield who's feedback and friendship have been a joy. Thank you to Martin Krämer, Reuben Binns, Petr Slovak, Dave Murray-Rust, Nigel Shadbolt, Ulrik Lyngs, Konrad Kollnig, Claudine Tinsman, George Challoub, Helena Webb, Ivan Flechais, Nitin Agrawal, Ge (Tiffany) Wang, Jun Zhao, and Marina Jirotka.

Thanks are also due to my wider Cybersecurity CDT family and EPSRC for keeping me in champagne and caviar for the past four years. Outside of the department I've received so much love, support, and encouragement from my family (hi Mum!) and friends. There isn't space for you all here, but thank you to Rhiannon Michelmore, Ben Falconer, Jamie Bayne, Esther Rzewski, Jamie Lee, Fijnanda Van Klingeren, Oli Sturrock, and Reino Niskannen.

Abstract

Modern smart devices are capable of incredible things: making life easier, more enjoyable, and more secure. But this 'smartness' often comes at the cost of devices harvesting data from the home, constraining how we use them, and changing the ways we relate to each other. This thesis explores what it means for a device to be smart, the ethical concerns that smartness causes, and ways that we might re-think smartness to better support people's needs and values. It does so using four lenses of *smartness*, *privacy*, *social actors*, and *respect*.

In order to better understand the problem, we begin by presenting the results of surveys and interviews investigating perceptions of what smartness is, and how it manifests across a variety of contexts. In doing so we identify concomitant ethical concerns, such as privacy and autonomy, and discuss the similarities and differences in how they operate across devices. We then follow up by addressing two of the identified problems in greater detail. We present the results of a six-week technology probe deployment designed to give people control over their connected devices by visualising and constraining their data flows. We show how participants preferences shifted over the course of the study and how, when given the right resources, people can learn and come together to solve privacy problems in the home. Secondly, we explore social concerns around voice interfaces, with a survey exploring correlations between trust, anthropomorphism, and relationship development with voice assistants. We show how people develop relationships with social devices in a similar manner to those between people, raising questions about the potential for social interaction modalities to be used to manipulate. The thesis then brings these lines of enquiry together by proposing the concept of respect as a lens for using standards of interpersonal interaction to evaluate interactions with smart devices. Practical and theoretical perspectives on respectful behaviour from a variety of disciplines are used to link the behaviours of intelligent systems to previous work on moral theory, agency, social hierarchies, and oppression.

Drawing on each of these four lenses, the thesis closes by discussing potential ways to re-think smartness, reaping its benefits whilst mitigating its problems. From rejecting smartness altogether, giving people greater control over their

devices, making devices more respectful, and using different conceptual models of devices, the thesis lays the foundations for more socially aware systems that use their smartness to support users in managing and enjoying life in the connected home.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

"We exploited Facebook to harvest millions of people's profiles. And built models to exploit what we knew about them and target their inner demons. That was the basis the entire company was built on."

— Christopher Wylie, *Cambridge Analytica* whistleblower

Contents

1.1 Motivation

After decades of existence in popular imagination[1], the notion of a 'smart home' as a domestic space enriched with connected digital devices is gradually becoming a reality. Contemporary devices are able to automate tasks around the home, allow for hands free access to information, and are accessible to all who share the home environment.

But this comes with a trade-off; smart and connected devices frequently intrude on the privacy of the home by surveilling people's daily lives. With clear evidence of the corporate appetite for big data, many commercially available apps and smart devices are designed to return information on users and their homes to manufactures and/or advertisers [40, 155].

However, the connected home of early research and science fiction was one that allowed a new way of living, with people able to own and utilise the data collected about themselves to pursue more efficient, safer, and fulfilled lives (e.g. [322]). Science fiction rarely considered how economic forces would drive these devices to exploit their position in the home to collect data for profit, typically identifying government surveillance as the major threat to privacy and freedom (e.g. Orwell's *1984*). Ubiquitous computing research largely operated under the assumption that inhabitants would have control over the data produced in their own homes, and that this new technology would respect existing social conventions around privacy [181]. Sadly, this is still far from being realised.

This harvesting of what Zuboff terms 'behavioural surplus' can be difficult to reason about—the small activities that individually make up people's lives have little meaning or value by themselves, making it difficult to see the long term societal disadvantages and privacy risks of a few corporations with vast behavioural data sets. This reflects a more general information asymmetry that often exists between users and companies, leading to an imbalance of power when a corporation has access to large amounts of data about an individual.

At best, these companies are blind to the scale and nature of their intrusion into people's lives. At worst, they are intentionally violating ethical norms in a quest for increasing profits and power. Both possibilities further entrench the paradigm of surveillance capitalism, which "unilaterally claims human experience as free raw material", automating human behaviour on a global scale in order to drive consumption [369]. For users, this makes the consequences of accepting devices—and their terms—into the home unclear. Add in the pressure of the 'take it or leave it' Hobson's choice offered by virtually

[1] For the curious, a 1989 episode of the BBC program *Tomorrow's World* contains surprisingly accurate predictions about the smart home of 2020: `https://twitter.com/bbcarchive/status/1212018067804311552`

all device privacy policies, and it is clear why many click accept with little thought.

What marks the contemporary smart home as one that is different from the early promises of research and science fiction then, is that it is *shared* with these devices and the companies that operate them rather than being owned and controlled solely by the people who live within it. The encroachment of surveillance capitalism into the home transforms it into a battleground where companies compete to extract value from its residents. This is a concern because the home is an essential private space, providing safety and isolation from the outside world [227]. Somewhere to relax and genuinely be one's self, the boundary of the home represents a major "'attack surface' of the digital on everyday life" [78] and it is one of the few remaining places where people can truly be alone.

The technology that enables this large scale data collection, as well as the utility and convenience offered by modern devices, is often encapsulated as 'smartness'. Referring to varying features such as sensors, wireless communication, and advanced machine learning, smartness is applied to a wide range of devices with vastly different capabilities: a light bulb is clearly 'smart' in a different way to a speaker or a phone. In this way, smartness represents an unknown and constantly shifting assemblage of technology in the home.

A notable characteristic of smartness in modern devices is the way that it smooths over all but the most basic aspects of state and configuration. Devices light up, make soft pings, and add or remove functionality through updates, all with little or no user input. A smart device might have microphones and cameras, for example, but it can be extremely difficult to determine when they are recording and whether recordings from the home are processed locally or streamed to the cloud. Adjusting the amount of data that devices share with companies and other third parties is similarly difficult or impossible. The continued shrinking of consumer technology combined with its increasing sophistication means that even the smallest of devices might contain invasive functionality.

Theses issues are unquestionably ones of privacy, but approaches that consider privacy solely as access control or restrictions on the collection and processing of information cannot explain the phenomenological unease that many people feel when using smart devices. As a response to this the literature offers understandings of privacy that incorporate other values such as autonomy and transparency [260]. There is also an increasingly social component to smartness; when asked about data collection, it is not uncommon for users to talk about the devices and companies that they share their homes with *as if they were people* [217, 303, 337]. The increasing popularity of devices with sophisticated speech recognition and synthesis further blurs the line between natural, virtual, and corporate persons. Early Human Computer Interaction (HCI)

research showed that computers are *social actors*: people readily apply social rules to human-computer interactions and naturally project characteristics like gender, complete with the associated biases [253, 255, 256].

Controversy over children's use of 'please' and 'thank you' with voice assistants serves as a modern example that we cannot help but treat computers as if they were human, even if we know this not to be the case. This further complicates our relationships with smart devices in a way that cannot be explained by only looking at functionality. Other questions such as how parents should use smart devices around their children [81, 330], or the extent to which devices should choose the content that we see [107] lurk in the background, and demand understandings of privacy beyond those of fact and law. These questions probe more deeply at the essence of what it is that smart technologies offer us, and touch on ethical concerns around autonomy, transparency, and the social order of the home.

It is therefore becoming apparent that smartness in its contemporary form is fundamentally misaligned with people's values, and addresses their needs in ways that are unacceptable. It impinges not only on people's privacy, but also their autonomy and social relationships.

Research and Regulatory Responses

Responses to these problems by the research community have targeted every aspect of how devices in the home are designed, used, and regulated. The nature of smart devices means that research in this area is truly interdisciplinary, with contributions from computer science featuring alongside those from sociology, law, communications theory, philosophy, economics, and psychology.

To inform regulatory efforts, a range of expository research has generated policy recommendations based on various facets of smart home ethics. These discussions about the application of concepts such as data protection [341] tend to consider ethical issues independently, but this can fall short when concepts overlap. Often being rooted in older wiretap legislation—intended to prevent specific harms for calls made between two phones—data protection frameworks often struggle to deal with the intimate, co-created data produced in shared spaces such as the home (e.g. the European General Data Protection Regulation only considers the rights of individuals [131]). Software developers have also come under scrutiny from fairness, accountability, and transparency scholars for being opaque in how they use data in the cloud-based systems that power many smart devices and services [39, 245, 339].

On the other hand, empirical research by the academic community investigating user concerns with these devices tends to work from the ground up. Explorations into the effects of smart devices on the home show the broad

impacts of these technologies on the routines and social dynamics that exist within the home environment as people endeavour to limit the impact of these devices on their privacy and interpersonal relationships [78, 144, 352, 365]. Socio-technical work on specific devices maps out more localised effects, such as the vastly different opinions of parents and teenagers over whether parents should be able to monitor teenagers' devices [81, 330]). Research in this area often focuses on questions of adoption (i.e. how smart devices are integrated into daily life) [166], and by extension, barriers to adoption [51].

The nature of connected devices as *products* has prompted the application of a rich collection of design practices to smart home research. Common approaches include speculative design [17, 99], functional prototypes [64, 237], design fictions [203, 294], and even fictional *research* [214]. These techniques—that present the imagined as real—allow for the mapping of alternate futures that can be used to steer the direction of current smart home technology away from the grasp of surveillance capitalism. From a more technical perspective, empirical research on how smart home devices operate has helped to build a common understanding of the design patterns and architectures of contemporary devices [3, 155, 267, 326]. Primarily useful in the context of privacy and security, these findings are also of great help when creating tools for the connected home.

1.2 Research Questions

Crucially, existing efforts fail to provide solutions for the long term design and development of ethical smart devices. While many of the *technological* predictions made by early research have come true, there is clearly still much work to be done to steer the future away from cyberpunk dystopias where technology, motivated by money and multinational capitalist corporations, is associated with destruction, mental regression, and dehumanisation [163]. Hopefully we can instead move towards the promises of earlier science fiction utopias, leveraging smart technologies to promote human flourishing.

Motivated by this challenge, this thesis seeks to answer fundamental questions about the nature of smartness and the ethical concerns it causes, as well as what can be done to 'rethink' what it means to be smart in a way that mitigates concerns both now and in the future. Echoing Taylor et al., we see smart devices "less as something to be designed as intelligent and more as a resource for intelligence" [322].

To this end, the thesis examines concerns around connected devices in the home using four lenses, narrowing the scope of its investigations to those of: *smartness*, *privacy*, *social actors*, and *respect*. The first of these, *smartness*, examines what it means for devices to be smart and how this differs between devices. In studying perceptions of smartness, it surfaces ethical concerns

about what smart devices are, and the commonalities between the devices and contexts that these concerns are centred on.

In doing so it becomes clear that concerns around smart devices have many dimensions that can make it difficult to reason about why a particular aspect of a device or its context of use give cause for concern. The second lens of *privacy* explores the foremost amongst these in more detail, asking how we can to help people form and act on privacy preferences regarding connected devices, and exploring options for making data flows of devices understandable and controllable.

The third lens considers the the role of devices as *social actors*, and how their social nature might distort the way that we interact with them. It highlights the relationship between people's propensity to anthropomorphise social devices with how much they trust them, and the extent to which behaviour towards devices might mimic that between people.

This leads us to seek a fourth perspective that can help answer the questions that arise about how devices *should* treat people, helping to highlight the similarities and differences between conceptualisations in terms of privacy, smartness, and social interaction. For this we turn to philosophy for clarity, using thinking on how humans should treat each other to systematically reason about how devices treat *us*.

Specifically, we answer the following research questions:

RQ1 *Smartness*—how do ethical concerns about smart devices relate to different types of 'smart' functionality?

- 1a What do people perceive 'smartness' as?
- 1b How do perceptions of 'smartness' relate to ethical concerns about smart devices?
- 1c How do these concerns operate across devices and contexts?

RQ2 *Privacy*—how can devices support users' privacy?

- 2a How can privacy-implicated activities of devices be made legible to users?
- 2b How can we help users interpret and form preferences about data disclosures by their devices?
- 2c How can we provide privacy controls that enable users to take action about disclosures by their devices?

RQ3 *Social actors*—how do voice assistants distort social interactions?

- 3a Do users of voice assistants exhibit relationship development with their devices similar to that measured between humans?

3b Are trust, relationship development, and anthropomorphism related for voice assistants?

3c What underlying assumptions exist in the design of contemporary voice assistants?

RQ4 *Respect*—how can the concept of respect help us anticipate and design for ethical concerns with connected devices?

4a What conceptualisations of respect exist outside of computer science that are potentially relevant to HCI?

4b How do these conceptualisations relate to existing design frameworks and principles?

4c What steps can designers take to create more respectful systems?

1.3 Contributions & Outcomes

Through answering the research questions laid out above, this thesis makes a number of direct contributions to the understanding of how smart devices are used and designed in HCI:

An Understanding of Smartness

Through an analysis of how non-experts consider smartness in the context of smart home devices, we unpack the key aspects of device functionality that shape people's perceptions of what constitutes a smart device. We show how these components of smartness operate across contexts, directly relating them to the ethical concerns that they cause or ameliorate.

A Template for Privacy-Empowering Technologies in the Home

We explore the ways that people can better understand and interact with their existing connected devices, developing a novel technology probe to demonstrate how the combination of data flow visualisations, a data protection curriculum, and high level firewall controls can help people to develop meaningful privacy preferences and realise them within the landscape of the smart home.

Identifying a New Link Between Interpersonal and Human-Computer Relationships

By adapting prior work on interpersonal relationships and bridging it with contemporary voice assistant research, we show that people's relationships with voice assistants develop in a similar way to relationships between people. We further identify statistically significant correlations between this

relationship development and both trusting behaviour and anthropomorphic tendencies in users.

Mapping the Landscape of Respect for Smart Home Designers

We bring together a wide range of conceptualisations of respect from work in philosophy and the human sciences, comparing them to established design frameworks and principles in HCI. Through the introduction of different configurations of respect, we show how designers of smart home systems can leverage this thinking from outside of HCI in order to build more respectful devices that promote human flourishing.

1.4 Thesis Outline

Chapter 2 sets out the prior work that this thesis builds on, beginning with efforts to understand and map out the evolving landscape of the smart home and its relationship to those who inhabit it. From situated explorations of individual devices to concerns over the ubiquity of surveillance capitalism, it discusses the high and low level approaches to smart home research, as well as the potential for intermediate approaches. This is followed by an overview of informational and digital privacy in the West, including a comparison and critique of major models of informational privacy. The chapter closes by focusing more broadly on computer ethics, exploring mechanisms of disrespect, marginalisation, and oppression that are perpetuated by digital technology, and how the HCI literature has previously considered the concept of respect. Through this comparison and analysis of existing work, the chapter identifies opportunities to advance our understanding of what smart devices are, how we might empower people to take control of their privacy at home, and explore connections between interpersonal interaction and interactions with voice assistants.

Chapter 3 goes on to summarise the major HCI theories and paradigms related to research presented in the thesis, discussing different approaches to research and design in HCI. As part of this, the chapter critically examines the benefits and drawbacks of these approaches and lays out how they compare to and could be developed by the work on the ethics of respect presented later in the thesis, paving the way for the theoretical work in Chapter 7. It then details the practical research methodologies employed throughout the thesis, including the use of surveys and interviews, with a reflection on the adaptation of protocols to individual experiments. Finally, there is a discussion on research ethics in the context of the studies undertaken, along with the steps taken to improve the replicability of the results obtained.

Chapter 4 asks what features of smart devices users consider as rendering them 'smart', and how these relate to the ethical concerns that they generate.

Beginning with a brief overview of the gap left by previous high and low level contributions, it presents the results of a multi-method investigation into what end users perceive smartness as consisting of. Using surveys and semi-structured interviews, eight types of smartness are identified and described. We then explore and discuss how these perceptions of smartness work to engender a variety of ethical concerns including privacy, autonomy, and disruption to the social order of the home.

Chapter 5 takes this more nuanced understanding of smartness and dives more deeply into concerns around privacy and transparency, exploring the potential for new tools that support and empower users. It details the development of Aretha, a privacy assistant technology probe that combines a network disaggregator, personal tutor, and firewall, intended to give people knowledge of data disclosures from their connected devices and the means to control them. The results of a six week home deployment show how the probe, with its novel affordances—and its limitations—prompted users to co-adapt, developing their own control mechanisms and suggesting new approaches to address the challenge of regaining privacy in the connected home.

Chapter 6 focuses on voice assistants as a class of smart device, exploring the ways that spoken conversational interfaces might influence how people relate to their smart devices. It does this by exploring how closely interactions with voice assistants might mirror the relationships we have with other people, and how this occurs alongside trust and anthropomorphism. The chapter reflects on the potential opportunities and dangers of using increasingly social devices to exploit the ways that humans are 'wired for speech'.

Chapter 7 brings together findings from the preceding chapters on the ways that people are treated by connected devices, seeking to unite them under a single notion. Starting from a grounding in the philosophical concept of respect, it draws from a diverse range of thinking from philosophy, sociology, psychology, and political science in order to explore what respect means and how it is relevant to the design of smart devices. This is then discussed in terms of its application to the design of smart and connected devices, with four key principles identified and expanded upon.

Chapter 8 discusses the findings of the thesis in a broader context, revisiting the research questions in light of the results obtained. It then reflects on the nature of smartness, suggesting ways in which we might re-think the concept in order to better align the future with the ideals of early smart home research and science fiction. In doing so it presents avenues for future work that move us towards a respectful and value sensitive future.

Finally, the thesis concludes in Chapter 9 with a summary of the overall research project, its goals, results, and future direction.

1.5 Terminology

Smart Device: refers to the category of internet connected devices sold for use solely in the home that are not marketed as general-purpose computers (e.g. excluding computers, laptops, and mobile phones). Chapter 4 contains a more in-depth exploration of what it means for a device to be smart.

Connected Device: refers to any internet connected device that might typically be found in the home. All of these devices are arguably smart, but it is often useful to be able to refer to them as distinct categories.

Smart Home and **Connected Home**: used interchangeably to refer to homes that contain one or more smart/connected devices.

VA: Voice assistants (often referred to by the distinct but overlapping categories of **Virtual Assistants**, **Conversational Assistants**, and **Smart Speakers**) are software agents that accept tasks and relay information through spoken conversation. Pioneered by Apple with the launch of Siri in 2011, voice assistants were later popularised by Amazon with the launch of Alexa in 2014. Major competitors followed suit with the subsequent launches of Cortana (2014), the Google Assistant (2016), and Bixby (2017). Today VAs are available on a variety of 'smart' devices, including speakers, phones, and TVs.

GDPR: the European General Data Protection Regulation / Regulation (EU) 2016/679 [131]. Implemented in 2018, the GDPR is a European Regulation governing the protection of personal data in EU member states. The GDPR sets out the grounds on which different types of information can be collected from individuals, how it must be stored, what processing may be done on that data, and the uses to which collected and processed data may be put. The GDPR repeals the existing Data Protection Directive (Directive 95/46/EC), and was implemented in the UK Data Protection Act 2018 to replace the 1998 act of the same name.

Data Controller: the company or individual that is legally responsible for how and why personal data is processed (defined in Article 4 of the GDPR [131]).

Data Flow: a connection between computers (in this thesis typically a smart device and a company server) through which data is transferred.

ICO: The Information Commissioner's Office is the UK national data protection authority, responsible for overseeing and enforcing data protection regulations. The ICO is an independent public body that reports to Parliament, and is sponsored by the Department of Digital, Culture, Media, and Sport.

1.6 Publications

The work presented in this thesis has previously been disseminated via the following publications:

Conference Papers

William Seymour, Martin J. Kraemer, Reuben Binns, and Max Van Kleek. 2020. Informing the Design of Privacy-Empowering Tools for the Connected Home. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (ACM).
https://doi.org/10.1145/3313831.3376264.

William Seymour, Reuben Binns, Max Van Kleek, Petr Slovak, and Nigel Shadbolt. 2020. Strangers in the Room: Unpacking Perceptions of 'Smartness' and Related Ethical Concerns in the Home. In proceedings of the 2020 ACM Conference on Designing Interactive Systems (ACM).
https://doi.org/10.1145/3357236.3395501

Max Van Kleek, William Seymour, Reuben Binns, Jun Zhao, Daniel Karandikar, and Nigel Shadbolt. 2019. IoT Refine: Making Smart Home Devices Accountable for Their Data Harvesting Practices. In Proceedings of Living in the IoT 2019 (IET).
https://doi.org/10.1049/cp.2019.0134

William Seymour, Max Van Kleek, Reuben Binns, and Nigel Shadbolt. 2019. Aretha: A Respectful Voice Assistant for the Smart Home. In Proceedings of Living in the IoT 2019 (IET).
https://doi.org/10.1049/cp.2019.0154

Max Van Kleek, William Seymour, Reuben Binns, and Nigel Shadbolt. 2018. Respectful Things: Adding Social Intelligence to 'Smart' Devices. In Proceedings of Living in the IoT 2018 (IET).
https://doi.org/10.1049/cp.2018.0006.

Submitted Publications

William Seymour and Max Van Kleek. 2020. Exploring Interactions Between Trust, Anthropomorphism, and Relationship Development in Voice Assistants. Submitted to the Proceedings of the ACM on Human Computer Interaction (ACM).

William Seymour, Max Van Kleek, Reuben Binns, and Dave Murray-Rust. 2020. Conceptualising Respect: Supporting the Ethical & Inclusive Design of Interactive Systems. Submitted to AI & Society.

Extended Abstracts

William Seymour and Max Van Kleek. 2020. Does Siri Have a Soul? Exploring Voice Assistants Through Shinto Design Fictions. Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (alt.chi).
https://doi.org/10.1145/3334480.3381809

William Seymour. 2020. A Design Philosophy for Agents in the Smart Home. Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (CHI Doctoral Consortium).
https://doi.org/10.1145/3334480.3375032

William Seymour. 2019. Privacy Therapy with Aretha: What if your firewall could talk? Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems.
Winner, SIGCHI Student Research Competition (graduate).
https://doi.org/10.1145/3290607.3308449

William Seymour. 2018. How loyal is your Alexa? Imagining a Respectful Smart Assistant. Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems.
Runner up, SIGCHI Student Research Competition (graduate).
https://doi.org/10.1145/3170427.3180289

Workshops

Martin J. Kraemer, William Seymour, and Ivan Flechais. Responsibility and Privacy: Caring for a Dependent in a Digital Age. Privacy and Power: Acknowledging the Importance of Privacy Research and Design for Vulnerable Populations (CHI '20 Workshop).

Martin J. Kraemer, William Seymour, Reuben Binns, Max Van Kleek, and Ivan Flechais. 2019. Informing The Future of Data Protection in Smart Homes. New Directions for the IoT: Automate, Share, Build, and Care (CHI '19 Workshop).

William Seymour and Max Van Kleek. 2019. The Internet of Kant: Respect as a Lens for IoT Design. Standing on the Shoulders of Giants: Exploring the Intersection of Philosophy and HCI (CHI '19 Workshop).

Max Van Kleek, William Seymour, Michael Veale, Reuben Binns, and Nigel Shadbolt. 2018. The Need for Sensemaking in Networked Privacy and Algorithmic Responsibility. Sensemaking in a Senseless World (CHI '18 Workshop).

William Seymour. 2018. Social Acceptability and Respectful Smart Assistants. (Un)Acceptable!?! – Re-thinking the Social Acceptability of Emerging Technologies (CHI '18 Workshop).

William Seymour. 2018. Detecting Bias: Does an Algorithm Have to Be Transparent in Order to Be Fair? BIAS - Bias in Information, Algorithms, and Systems (iConference '18 Workshop)

Chapter 2

Background

"I have over 4,000 emails, pictures, addresses [...] People just submitted it. I don't know why. They trust me. Dumb fucks."

— Mark Zuckerberg, 2003[1]

Contents

This chapter situates the work presented in this thesis in relation to prior contributions. It begins with a discussion of smartness, from its roots as a historical, marketing, and ideological concept, to contemporary expository work on ethical principles, both as they operate across individual devices and

[1] `https://www.businessinsider.com/well-these-new-zuckerberg-ims-wont-help-facebooks-privacy-problems-2010-5?r=US&IR=T`

Parts of the content in this chapter have been adapted from previously published papers listed in Section 1.6.

contexts as well as the home in general. The discussion then moves towards the high-level phenomena related to our goal of mitigating ethical concerns in smart devices.

The second part of the chapter delves into one of the major concerns expressed about smart devices—privacy—in more detail, giving a brief history of informational privacy theories that attempt to categorise and reason about the nature of privacy as experienced online and through connected devices. It considers the efficacy of these theories in explaining peoples' interactions with connected devices in the home and the limits of privacy's ability to explain and solve ethical problems with smart devices.

Finally, we look towards the concept of respect as a potential answer to the problems highlighted around the design and use of smart devices for the home. Motivated by ongoing examples of disrespect that are perpetuated by digital technologies, we explore the different ways that minority and marginalised groups are excluded or disadvantaged by the structure and operation of these technologies, and how this manifests across communities. The chapter closes by reviewing previous contributions in HCI that have examined the concept of respect.

2.1 Smartness

As with any loosely defined technological concept (e.g. cloud computing or blockchain), narratives surrounding smartness and what it means for devices to be smart differ greatly based on context and culture. Smartphones, for example, have seen tropes of integration and dis-integration, pitching the ideal of the connected and productive phone user against fears of becoming addicted or out of touch [151]. The public/private sector balance that drives smart cities, on the other hand, often tends towards more hegemonic discourse, ignoring risks such as technological lock-in and the overreach of surveillance and profiling [186].

Similar to the smartphone, smart home devices are largely produced by the private sector, although perception differs in that smart home devices are mainly seen as novelties or conveniences, rather than enabling a new way of living to which the smart home is integral. As with smart cities, the neoliberal approach to smart device functionality continually situates devices as solutions, rather than as problems, a process which Kitchin describes as designed to "bring [detractors] into the fold while keeping [vendors'] central mission of capital accumulation and technocratic governance intact" [186]. But this approach of addressing symptoms rather than the cause of problems only suffices while the pace of development prevents evaluation of past attempts; it is unlikely that introducing a 'please and thank you' mode for Alexa, for example, addresses the fundamental concerns around raising children in an

environment with voice assistants, but it will likely be obsolete by the time this is fully explored.

Central to all of these narratives (smart phones, cities, and homes) is the echoing of the hopes and fears of contemporary society, often to the detriment of more nuanced accounts of how people use technology. The stories told by designers and marketers tend to produce categories that pitch practices and values against each other (e.g. Harmon's tropes of integration and disintegration), creating false dichotomies for users of those technologies [151], and designers themselves become caught up in the smartness dream, describing attributes that do not exist in contemporary devices (such as contextual awareness and wide-ranging interoperability) when asked what it means to be smart [183]. How these tropes are perceived matters beyond the realms of marketing and adoption; smart devices increasingly blur previously well-defined categories such as "human" and "machine", leaving them constantly in flux [204]. This drives users to 'purify' their definitions of these categories in order to keep them disjoint (see also [289]). In this way, we can think of smart homes as approaching "humanness and machine-ness as intra-actively produced"—Leahu notes that like users, researchers also exhibit this behaviour, naturally over-focussing on certain phenomena when looking for categories [204].

A significant portion of contemporary work on smart homes is (explicitly or implicitly) motivated by questions of adoption; i.e., identifying factors which might lead to people using or rejecting smart devices in their home. Within this framing, ethical concerns are often lumped in with a range of other factors seen as 'barriers to adoption' (a phrase often used throughout the literature, e.g. [51]). As Jensen et al argue, (drawing from [257]), *ethics* can be defined as that which relates to how an object *ought* to be designed according to 'ethical or moral codes', and is one of three broad kinds of desiderata for smart homes, alongside *reason* (relating to a smart home device's 'purpose'), and *aesthetics* (relating to how pleasurable it is to use) [172].

In-situ Investigations of Smart Devices and Associated Ethical Concerns

At a high level, investigations into smart home ethics can be divided into those with technical or social focuses, and the latter further into those that consider particular devices or contexts of use. All three have their places, with technical analyses outlining the mechanisms of data capture, processing, and retrieval that form the basis of concerns, and social investigations showing the impact of those mechanisms on people, their relationships, and the environments in which they live; none of the ethical issues explored in this thesis could be investigated with one of these approaches alone.

Studies focusing on the impact of individual devices often target privacy concerns presented by particularly relevant or high-impact devices such as connected toys [236], cameras [270, 330], smartphones [133, 304] TVs [224], and voice assistants [124, 218, 284, 325]). Broader studies that consider several smart devices that might be found in the home have studied data collection [144], household security behaviours [344], and privacy concerns [81, 132, 365] of smart device users. The findings of these studies often replicate those of more general work on privacy, such as the "nothing to hide" mentality and the tendency for users to perceive others as more vulnerable to surveilance than themselves. Many users are unaware or uncertain about what data their devices are collecting [224], and many adopt a position of dejected acceptance when they lack the tools and strategies required to mitigate the ethical concerns caused by their devices and apps [304]. But while devices with prominent cameras and always-on microphones are well represented in research outputs, there is a risk that problems with more mundane devices go un-investigated.

The other major perspective of social investigations into smart devices is the exploration of how specific interpersonal relationships are impacted by connected devices, which often highlight how technology can strain social arrangements and force people to adapt their routines. Power dynamics in the home are affected in different ways by technology, with several studies on specific relationships such as parents and children [81, 330], and the way that existing arrangements and etiquette are adapted in the face of technology [166, 283].

There has also been a sustained interest in understanding how smart devices are used and how they operate at a technical level. For end users there are two main projects oriented towards the privacy and security-conscious: Princeton IoT Inspector and Fing. The former is aimed at a general audience, prioritising ease-of-use and a simple feature set including visualisations of device activity, including a display of data destinations associated with advertising or tracking. The latter is primarily designed for securing and troubleshooting networks, providing tools for advanced users.

Research tools providing smart device analysis capability fall into three categories. Device fingerprinting in IoTSense [35] and IoTSentinel [239] was able to identify devices based on the network traffic they produced. Behaviour classifiers such as HomeSnitch [267], Peek-a-Boo [3], PingPong [326], and HoMonit [366] use machine learning to infer the reasons that devices are sending data to particular destinations at particular times.

A third kind of tool, including HomeNet Profiler [92] takes more holistic approach rather than device-specific analyses. Homenet Profiler was used to collect MAC addresses and statistics about all of the services advertised by devices across 2,400 French homes. Meanwhile, X-Ray Refine [337] was

	IoT Inspector[2]	HomeSnitch [267]	Fing[3]	HomeNet Profiler [92]	IoT Sentinel [239]	Peek-a-Boo [3]	PingPong [326]	IoTSense [35]	HoMonit [366]	X-Ray Refine [337]
For end-users	●	○	●	◑	○	○	○	○	○	◑
Actively Maintained	●	○	●	○	○	○	○	○	○	○
Open Source	●	○	○	○	●	○	○	○	○	●
Deep Inspection	○	○	○	○	○	○	○	●	○	○
Destination Labelling	●	○	○	○	○	○	○	○	○	●
Classifies Devices	●	●	●	●	●	●	●	●	●	○
Infers Behaviour	○	●	○	○	○	●	●	○	●	○
Live Visualisation	○	○	○	○	○	○	○	○	○	○
Active Controls	○	○	○	○	●	○	○	○	○	○
Anomaly Detection	○	●	●	○	○	○	○	○	●	○

Table 2.1: Comparison of popular home network analysers. Black circles indicate feature presence, with half circles denoting projects designed to be run by end users, but primarily used for research.

used to create data sharing models using static code analysis techniques to build high-level visualisations of smartphone users' entire exposure profiles resulting from their app use. A summary of all of these projects and their functionalities is given in Table 2.1.

Voice Assistants

Unlike how smartphones are seen to enable new ways of living [151], voice assistants have largely functioned as proxy input devices *for* smartphones (with the possible exception of playing music). Quantitative analysis of what people use voice assistants for suggests that the majority of invocations are comprised of listening to music, checking weather forecasts, searching for information, and controlling IoT devices (see Table 2.2). Requests for facts and weather forecasts can appear to extend existing smartphone usage, and control of IoT devices is almost exclusively supplemental to a smartphone app.

But there are other properties of voice assistants that complicate their role in the home. The always-listening nature of products such as the Echo and Google Home can lead to feelings of being observed [303], with demonstrated

Study	Device	Sample	Music	Search	IoT	Weather	Small talk	Reminders
[32]	Google	88	40%	17%	9%	6%	8%	6%
[13]	Google	88	26.1%	26%	-	-	-	-
[13]	Alexa	82	28.5%	19.4%	-	-	-	-
[301]	Alexa	75	25%	3.7%	14.7%	4.6%	2.6%	5.1%
[218]	Alexa	19	21.3%	7%	13.2%	28.7%	4.4%	8%

Table 2.2: Relative usage of Amazon Alexa and Google Assistant by category. Data collection was via device logs except for Lopatovska et al. [218] who used daily diaries instead of device logs. Ammari et al. [13] only provide statistics for music and search. Sciuto et al. [301] report a large amount of 'uncategorisable' commands (20.3%).

long term psychological effects [270]. Studies of voice assistant uptake in older adults show usage varying based on perceived need for social interaction and help with physical tasks [284, 325]. Indeed, there is something about the nature of computers that communicate with natural language, especially those that speak, that taps deeper into the human psyche. Voice control also raises novel challenges and solutions relating to equality of access. Voice assistants are often viewed as a key accessibility tool for visually impaired [285, 317], but also for those with cognitive or motor disabilities [2, 20]. Unusually, the 'power user' experience of many visually impaired voice assistant users also makes their design input valuable for *all* users [2]. The modality of voice also creates its own accessibility problems, such as making certain content less accessible (e.g. musical sub-genres which make use of non-standard words, spellings or characters [313]) and use by those with hearing diabilities.

Because users gender voice interfaces, applying stereotypes as they would with human interlocutors [255, 253], to give a device a voice is to give it a gender[4], and indeed many devices are explicitly gendered (see Figure 2.1). The industry norm of representing voice assistants as female has been criticised for reinforcing existing societal biases around the role of women in the workforce, portraying them as "obliging, docile and eager-to-please helpers" [347]. Market leaders have also come under fire for dealing poorly with sexually abusive comments that "intensify rape culture by presenting indirect ambiguity as a valid response to harassment" [114].

[4]Although there is at least one effort to create a genderless voice profile called Q—`http://www.genderlessvoice.com/`

- **Alexa** reports to be "female in character" [347].
- **Cortana** is a character in the Halo video games that projects itself as a 'sensuous unclothed woman' [347].
- The **Google Assistant** was described by an engineer as "a young woman from Colorado" [347].
- **Siri** is a Scandinavian female name [347].
- This phenomenon is not limited to the west—the market-leading Korean voice assistants all have female voices as the default and sometimes only options [162].

Figure 2.1: Gender identity and performance of major voice assistants.

Recent work has dug deeper into the social nature of people's responses to voice assistants, exploring personification [217, 288], effects of modality on information disclosure [63], effects on the social order of the home [283], and how older adults relate to voice assistants [284]. There has been some trouble, however, in determining the extent to which these results represent a shift towards deeper social relationships versus mindless behaviours triggered by the use of speech as a modality; observed interactions with voice assistants seem to sit between traditional human-computer and human-human interactions. Prior work suggests that voice assistants engender a sense of social presence in users [63], are seen as human in some sense [207, 358], and that interactions with voice assistants can often be positive even when failing to fulfil their functional objectives (i.e. the interactions themselves are satisfying) [218]. When smart devices and homes are personified, they may themselves become treated as social agents, with end-users ascribing them character traits such as 'my best friend', 'caring', or 'commanding' [237, 288]. However, other contributions consider the behaviours often associated with personification to be mindless re-enactments of over-learned social routines [116, 254].

Smart Meters and Grids

Research into smart energy meters and energy disaggregators offers an opportunity to engage with many of the same problems exhibited by smart home technology in a context where the actors and boundaries are much better understood. A rich body of qualitative research into smart energy meters (that make energy usage transparent) and smart energy agents (that provide mechanisms for varying energy prices based on time/demand) describes a set of concerns that closely mirrors those expressed around smart home devices. The long-established nature of domestic energy policy and pricing helps to

crystalise these problems in ways that the novel and fast-moving nature of smart home technologies make more challenging. Whereas smart home devices are designed to offer convenience and automation, these projects instead offer information and mechanisms to support decision-making in the home, perhaps considering the intelligent home "less as something to be designed as intelligent and more as a resource for intelligence" [322].

Autonomy is a primary issue that arises with the introduction of technology that makes demands on energy usage, either implicitly by signalling differences in pricing, or explicitly through mechanisms that require planning of energy usage, as adapting to the demands of energy systems disrupts normal household routines [53, 75, 143, 293]. While these systems are often perceived as helpful, there is a clear tension between them helping users to manage their energy usage, and managing users to promote the aims of energy companies.

Not only can these systems create additional work for users [75], but it can also threaten to shift the responsibility for high energy bills onto consumers (i.e. as additional unwanted tools that justify additional costs and responsibilities) [53], especially as energy companies are commonly distrusted and not perceived to be acting in the best interest of their customers [53, 293]. The data sharing relationship between the energy company and the user of the smart meter/agent is also made obvious, and the fact that the energy company has a clear business case for collecting the data appears to make consumers more wary, expecting something in return and potentially feeling spied upon [293].

Towards a Definition of Smartness

Given the popularity of smart devices with consumers and the rapidly growing body of research surrounding them, one might think there might be conventions around what is and is not a smart device—what set of features or use case turns something 'dumb' into something 'smart'. But throughout the literature, smartness is consistently used to either as an a priori (and unexplained) definition of a class of device, or as a prompt for participants who are then left to determine which devices are in scope[5]. Studies that focus on a more specific set of devices appear to offer more clarity, but again there is no discussion about what it is that makes a smart camera or speaker smart. Internet connected is occasionally offered as a starting point, but that would appear to exclude privacy preserving devices that offer computation and processing locally on the device itself (still arguably 'smart').

High-level Explorations: Policy, Psychology, and Capitalism

In addition to empirical research like the above which elicits concerns pertaining to particular devices in particular contexts (whether real or speculative), a

[5]Left to their own devices, if you will.

range of expository work addresses smart home ethics from the level of policy, principles, philosophy, cultural critique, and other perspectives. For instance, discussions about the application of legal regulations like data protection [341], or 'ethical design manifestos' (such as those analysed in [123]) tend to consider ethical issues independently from a 'bird's eye view', in contrast to the in-situ work described above that works empirically from the ground-up to investigate user concerns. Within such work, overarching principles and forms of harm are enumerated, such as non-discrimination, privacy and accountability (e.g. [34, 244]).

This work is useful in situating the ethics of smart homes in relation to more esoteric and theoretical discourse around long-standing moral principles and human rights, as well as drawing points of connection to potentially legally enforceable rights and duties. As legal obligations are increasingly mandated to be embedded in technologies from the outset, through requirements like privacy- and data protection-by-design [333], such work can potentially guide which principles might apply in a range of smart home contexts. However, they may not always reflect the concerns that are in fact elicited or expressed by end-users in reality, and they don't by themselves map how the design space of smart devices might relate to those ethical principles.

There are also several high level phenomena that directly relate to the understanding and rethinking of smartness. Patterns in the way that we perceive and interact with devices that use increasingly social modes of interaction, and the increasing technical sophistication of the sensors embedded within them, have generated fruitful new avenues of investigation for HCI researchers. We now discuss the resultant concerns about the extent to which contemporary devices are designed to interact socially with their users, and the trend towards business models built around tracking and data-harvesting.

Computers as Social Actors

Pioneering work by Nass et al. in 1994, on what would later become the Computers are Social Actors (CASA) paradigm, showed that people sometimes applied social rules (and took corresponding social actions) when interacting with computers even whilst being fully aware that such computer(s) were machines, not humans [255]. This has been described as a kind of cognitive dissonance (see [116]), in which individuals experienced difficulty rationalising their actions with computers during Nass' experiments. Further supporting this view included a series of experiments carried out by Nass, Steuer, and Tauber, which demonstrated that (1) people were prone to apply notions of 'self' and 'other' in interactions with computers; (2) draw upon and apply gender stereotypes in such interactions; and, finally, (3) exhibited automatic and unconscious social responses in particular settings [255] (see also [253, 289], later replicated for voice assistants in [217]).

This same methodology was used by Moon and Nass to show that social routines such as reciprocal information sharing [246] and mindless re-enactment of existing social scripts [254] also apply to human-computer interactions. This mindless interaction is reminiscent of overlearned social routines, where any task-components that seem applicable are blindly accepted as true, such as [116]. Early work on affective computing also recognised the potential for machines to manipulate and deceive using emotions, although it did not anticipate this affecting subsequent interpersonal interactions [277, 278].

Intertwined with our compulsion to treat computers as social actors is the phenomenon of anthropomorphism, the "universal human tendency to ascribe human physical and mental characteristics to nonhuman entities, objects and events" [363]. Anthropomorphism is frequently studied in the context computers and robots, where the language used to describe agents often forms a central part of the methodology [127, 217, 288]. It is commonly accepted that, as with CASA, anthropomorphism does not come from a deeply held belief that computers and robots actually possess human characteristics such as emotions or gender, but rather that people act as if this were the case. Given this, anthropomorphism is often cited as a mechanism for humans to interpret the actions of computers and robots [363], providing a scaffolding to allow for the development of mental models (similarly to how people apply folk psychology to animals [67]). Unsurprisingly, these projection effects have been found to correlate with a machine's perceived intelligence [28]. Work by Loughnan and Haslam further revealed that animals and robots are subject to distinct kinds of anthropomorphism (infra-humanisation and self-humanisation respectively) [219].

Together, CASA and anthropomorphic tendencies amplify existing mechanisms of attachment to inanimate objects, already a commonly exhibited human behaviour. Prior work in HCI has examined the use of smartphones as objects of attachment [93, 182], drawing a distinction between attachment to the *device* versus attachment to its *functionality* [125]. Building on this, the CASA paradigm and the ability of many modern smart devices to leverage speech as an interaction modality suggests that some of our relationships with them might actually be more social than functional [218], in contrast with smaller groups—such as the blind—that have more functional relationships with devices like voice assistants as assistive aids [2]. The one-sided nature of interactions with speech-powered devices might be more akin to the parasocial relationships described by Horton and Wohl, where "the interaction, characteristically, is one-sided, non dialectical, controlled by the performer, and not susceptible of mutual development" [154]. However, while parasocial relationships capture the non-developmental aspect of interactions with contemporary voice interfaces, the term implies a level of perceived friendship that is often missing from the interactions reported by existing

research, which is better characterised as a fluid position between friend and machine that changes over time based on interaction outcomes [284].

Surveillance Capitalism

As discussed above, the most prevalent ethical concerns expressed about smart home devices relate to the capitalist business models that underpin the modern smart home, relying on the identification, extraction, and monetisation of so-called 'behavioural surplus' (i.e. traces about behaviours that would otherwise go uncaptured or unused). Described by Zuboff as "unilaterally [claiming] human experience as free raw material", surveillance capitalism describes the trend towards big data business models built upon the extraction and exploitation of this unmined information about people's behaviours, and the ubiquity of connected platforms that are able to centralise control over it [369].

This had led to the creation of large-scale architectures for extraction, including a retooling of existing portions of internet architecture to facilitate omnipresent surveillance. This trend has spilled over from the digital world to the physical one (such as smartphones with GPS, check-ins, and activity tracking), the elements of which Crabtree, Tolmie, and Knight describe as the "attack surface of the digital on everyday life"[78]. While the public purpose of these technologies is delivering better products, they often serve the explicit goal of collecting more behavioural surplus in order to drive behavioural models and advertising revenues [369].

The typical route to market that Zuboff describes represents a more aggressive version of Pierce's *foot in the door* devices and *hole and corner* applications, where products are launched seeking forgiveness rather than permission, and legal or regulatory challenges are delayed until products become ubiquitous and indispensable [279]. At the same time, apps and devices make use of psychology and 'dark patterns'[6] to make usage habitual [87, 103] (as, ironically, do the apps that have been created to help people manage their usage [221]). The situation is further complicated by difficulties in creating legal frameworks that might constrain unacceptable practices and the self-regulatory approach taken by the West.

The size and influence of data collection industries can make them difficult to effectively regulate. In July 2019, the UK Information Commissioner's Office (ICO) released a report on advertising technology (adtech), describing how many of the real time bidding processes used by the industry violated conditions in the GDPR around consent and special category data [329]. The report gave the industry a six month period to change its behaviour, but seven

[6] `https://www.darkpatterns.org`

months on only some companies had responded with proposed changes[7]. While many companies continue to act in breach of data protection law, the ICO has been slow to take regulatory action.

Projects such as Databox [80] and Solid[8] represent long-term and direct responses to the mechanisms that enable surveillance capitalism, enabling more respectful models of data storage and attempting to strike a balance between edge-based data storage and the wider digital economy [79]. By allowing users to store (and thus control) their own data, the project aims to reverse the "crisis in trust" that has arisen as a result of centralised data silos, as well as the trend where users do not own their data. Crabtree et al. use the Databox project to examine the benefits of decentralised data storage and ownership in the context of the GDPR [80]. With a focus on adding accountability to IoT ecosystems, the ability of databox to provide mechanisms for ongoing consent, transparent data processing, control over data flows, data portability, and local processing is shown to satisfy the external accountability requirement of the GDPR.

What *is* Smartness?

Smartness has been used in many ways across research, marketing, and everyday discourse to mean many things. Research into smart homes has similarly taken many different forms, from individual devices such as voice assistants *in situ* to broad-reaching discussions over surveillance capitalism. Smart devices and smart homes interact with many broad paradigms of existing HCI research, most notably CASA, allowing studies to draw on a rich array of existing contributions.

However, prior work has been content to skirt the topic of smartness itself, either by limiting investigations to individual kinds of device or considering smartness as a self-evident descriptor. This limits research efforts, creating confusion when smartness is used to mean different things across studies. If findings about smartness as "internet connected" [365] in one context are compared to findings about smartness as "a resource for intelligence" [322], results and themes are likely to be divergent to the extent that they may seem unrepeatable (and thus potentially less sound). It is only through gaining an understanding of how people consider and reason about smartness that we can truly begin to understand its affects across the broad range of configurations of families and living spaces.

To the extent that smartness has likely become somewhat socially constructed, with it's use in policy and marketing across cities, homes, and individual

[7] https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2020/01/blog-adtech-the-reform-of-real-time-bidding-has-started/

[8] https://solid.inrupt.com

devices, user perspectives are needed when disentangling its many facets. These perspectives influence how people perceive and interact with smart technology, and can be as helpful or harmful in that respect as concrete functionality. But previous contributions on smart home ethics discussed above tend to either take a high-level approach in exploring how established ethical principles relate to the concept of the smart home, or focus on a small number of devices in narrow contexts. The results gained from these studies are therefore similarly segmented into those that are narrow (e.g. smart meter agents raise concerns about autonomy [75]), or broad (e.g. smart homes in general threaten principles of accountability [34] or discrimination [341]) in scope. Both approaches have merit, providing concrete data related to specific problems and informing policy and regulatory approaches respectively. But both lack the power to generalise the ethical concerns that might arise from different kinds of devices that are rooted in their smart functionalities. This reveals an exciting opportunity to develop an understanding that sits at a level between the two, exploring what smartness *is* and the human values that might be supported and hindered by different conceptualisations of smartness. This will be explored further in Chapter 4.

2.2 Privacy

The most obvious and readily offered defence against widespread data collection and surveillance capitalism is privacy. But far from being a simple catch-all solution to invasive devices and platforms, privacy is a wickedly complicated concept in disarray [308]. Virtually non-existent in the lives of ancient and prehistoric societies, where communal living around a fire was necessary to stay warm, privacy has since changed greatly in the West; over the next 7000 years, the introduction of separate houses, rooms, and beds allowed more of people's lives to remain unobserved, and greater consideration was given to the rights that people might enjoy with respect to privacy [307].

Very broadly conceptualised as the ability to withhold knowledge about or access to something, a myriad of philosophical and legal works have argued for the protection of things considered private matters from the state and society. In his 1859 text *On Liberty*, Mill warns about the dangers of "gross usurpations upon the liberty of private life" [240]. At a time of transition towards greater democracy away from all-powerful monarchs, Mill saw the potential for the "moral coercion of public opinion" to impose on the individual just as much as a monarch could [222]. His arguments for privacy are partly to defend the ability of individuals to craft their own identities (a now commonly recognised conception of privacy as the protection of identity or personhood [308]) and preserve a diversity of thought and opinion that was ultimately beneficial for society. *On Liberty* also discusses the circumstances under which individual freedoms can legitimately be interfered with, which

Mill claims is only when an individual causes harm to others (including by omission, such as failing to take care of dependants) [222, 240].

A few decades later in the US, developments in photography and its use by journalists led Warren and Brandeis to advocate for legal protection of the "right to be let alone" [342]. The article is widely considered to be the first discussion of privacy tort in the United States, shifting interpretation of existing non-disclosure law away from duties conferred by relationships– which journalists did not have with the people they photographed—towards the impact of such photos on their subjects [292].

The article was followed in 1960 by Prosser's *Privacy* in the California Law Review [287]. In it, he divides privacy into four torts:

1. Intrusion upon the plaintiff's seclusion or solitude, or into his private affairs.
2. Public disclosure of embarrassing facts about the plaintiff.
3. Publicity which places the plaintiff in a false light in the public eye.
4. Appropriation, for the defendant's advantage, of the plaintiff's name or likeness.

Prosser's position as a prominent legal scholar ensured that privacy was treated much more seriously in a legal context than it had been following Warren and Brandeis. But his concerns that privacy would come to envelop other, already well established sections of the law (such as defamation), and risk to defendants where damages were intangible led his work to take a more conservative approach [292]. Prosser discounted criticism that the right to privacy might be broader than was represented in his four cases, and Richards and Solove argue that in doing so he dramatically hampered the development of privacy law in the United States, leaving current laws struggling to deal with more nuanced contemporary understandings of privacy [292]. Indeed, the article concludes:

> This is not to say that the developments in the law of privacy are wrong. Undoubtedly they have been supported by genuine public demand and lively public feeling, and made necessary by real abuses on the part of defendants who have brought it all upon themselves. It is to say rather that it is high time that we realize what we are doing, and give some consideration to the question of where, if anywhere, we are to call a halt [287].

Thankfully, Prosser did not have the final say, and international law has since recognised (but often not defined) a universal right to privacy:

- "No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks."
  - *Art. 12, 1948 UN Universal Declaration of Human Rights*
  - *Art. 17, 1966 UN International Covenant on Civil and Political Rights*
  - *Art. 16, 1989 UN Convention on the Rights of the Child*
- Everyone has the right to respect for his private and family life, his home and his correspondence. There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others.
  - *Art. 8, 1950 European Convention on Human Rights.*

And more specifically with regards to informational and online privacy:

- "Whereas data-processing systems are designed to serve man; whereas they must, whatever the nationality or residence of natural persons, respect their fundamental rights and freedoms, notably the right to privacy, and contribute to economic and social progress, trade expansion and the well-being of individuals [...]"
  - *1995 EU Data Protection Directive 95/46/EC*
- "The protection of natural persons in relation to the processing of personal data is a fundamental right [...] The processing of personal data should be designed to serve mankind. The right to the protection of personal data is not an absolute right; it must be considered in relation to its function in society and be balanced against other fundamental rights, in accordance with the principle of proportionality.
  - *2016 General Data Protection Regulation (EU) 2016/679*

But perhaps having once been a simpler concept, privacy in the modern world is a perfect example of a portmanteau word that can be unpacked into a variety of sub-concepts and meanings. Described by Solove in 2008 as "a concept in disarray' [308], contemporary conceptions of privacy range from being left alone, through to a way to protect and craft one's personality, or even as a form of intimacy [308]. In contrast to these definitions, this work is concerned

	Low M.	Medium M.	High M.
Low K.	Marginally Concerned	-	-
Medium K.	-	Struggling Amateur	Technician
High K.	Lazy Expert	-	Fundamentalist

Table 2.3: Dupree et al.'s clustering of participants into privacy types based on knowledge (K) and motivation (M) [101].

with *informational privacy* (i.e. concerning information about oneself) and this section discusses the major theories and understandings of privacy that have influenced work on modern connected homes.

Theories of Privacy

Westin's Privacy Categories

Early research suggested that people may be divisible into a small set of distinct privacy 'types' that described attitudes and behaviour towards informational privacy. Through a number of surveys, Westin categorised participants as privacy 'fundamentalists', 'pragmatists', and the 'unconcerned' [194, 348]. Privacy fundamentalists (25–34% of those surveyed [194]) described those who felt very strongly about their privacy, believed that it had been eroded, and were more likely to advocate for stronger consumer privacy protections. Privacy pragmatists (55–63%) also cared about privacy, making decisions based on the characteristics of information sharing practices, only supporting tougher legal standards when they felt that voluntary measures were ineffective. Westin's final group, the privacy unconcerned (8–20%), showed little concern over potential abuse of shared information, and were largely happy to make disclosures in return for any type of benefit.

As might be expected, the simplicity of Westin's categories has garnered substantial criticism. Confirmatory studies have found that the definitions of fundamentalists, pragmatists and the unconcerned often fail to align with privacy and security behaviours [356] or views on specific related topics [184]. In the words of King, "how helpful is it to categorize someone as a Privacy Pragmatist if that categorization can't be used to predict where that person's opinion falls on the issues of the day?" [184]. See also discussion on the 'privacy paradox' below.

Dupree et al. later explored different categorisations based on Westin's original types, using a matrix of knowledge and motivation (Table 2.3) [101]. *Fundamentalists* are cast in the same light as by Westin, deeply sceptical of external security guarantees and their views on privacy extend beyond the computer. *Lazy Experts*, share the knowledge of fundamentalists but lack the motivation to act, rationalising their inaction with the belief that they are not targets. Lazy

Experts employ good privacy and security practices, but will use them to reduce the amount they need to interact with security, opting for sociability over protection if made to choose. They are also more likely to teach others than fundamentalists. *Technicians* take active steps to keep informed about privacy and security, but will trust gut instinct until presented with enough evidence. Amateurs occupy the middle ground, taking steps to protect themselves but lacking the level of knowledge and motivation to distinguish good advice from misinformation. Finally, the *Marginally Concerned* have mostly second hand knowledge and only act when prompted by an external event.

Boundary Regulation Theory

Contrary to Westin and other's theories, a key observation by Irwin Altman in 1975 was that informational privacy was not static or governed by sets of rules. Instead, Altman conceptualised privacy as a more fluid, dialectic process through which one exerts "selective control of access to the self or to one's group" [11]. In his Boundary Regulation Theory, Altman captured what he saw as:

- The dialectic quality of privacy—a synthesis of moving in and out of contact with others rather than simply as an exclusionary process of keeping others out
- The optimisation nature of privacy—whereby the amount of contact one desires with others fluctuates over time, making it possible for the same event to feel crowding, comfortable, or isolating depending on circumstance
- A process of boundary regulation—where boundaries between individuals and groups are constantly being negotiated and redrawn to match changing levels of desired social contact

The notion of privacy as a dialectic process is similar to Goffman's framework of self presentation, where one's identity is defined by what we share with others (and, therefore, what we keep private). But whereas Altman saw privacy with others as something that one moves in and out of to reconcile it against shifting preferences, Goffman suggests that the process of crafting one's identity is co-constructed according to the cultural and social settings in which interactions take place (i.e. that in exposing a presented character to a group, further construction and development is not something one remains solely in control of):

> In analyzing the self then we are drawn from its possessor, from the person who will profit or lose most by it, for he and his body

> merely provide the peg on which something of collaborative manufacture will be hung for a time. And the means for producing and maintaining selves do not reside inside the peg; in fact these means are often bolted down in social establishments. There will be a back region with its tools for shaping the body, and a front region with its fixed props. There will be a team of persons whose activity on stage in conjunction with available props will constitute the scene from which the performed character's self will emerge, and another team, the audience, whose interpretive activity will be necessary for this emergence. The self is a product of all of these arrangements, and in all of its parts bears the marks of this genesis [141].

But while Altman's approach expertly expresses important aspects of interpersonal privacy, it struggles when applied to the models of online data collection described in previous sections of this chapter (understandably so given that it preceded the invention of the World Wide Web by 14 years). A prime example of this is the way that third parties influence decisions around data processing—in order to use a new device or platform, one must first agree to the terms of service. There is typically no option to negotiate how data is shared, and social pressures often exist around the use or non-use of certain platforms (e.g. being able to switch to a more privacy-preserving messaging app is of little utility if other social contacts are not using it). Schwartz argues that self determination is heavily influenced by this model of personal data processing: "the external conditions of data use begin by affecting what it means to agree to information processing, and end by helping to form the conditions of social and individual life" [300] (similar to what Brunton and Nissenbaum describe as the "fantasy of opting out" [50]). In this way, the focus on individual choice in boundary management can fail to acknowledge the potential asymmetries that exist around the processing of personal data.

Privacy for the Networked World

In response to this, Palen and Dourish build on Altman's theory of privacy as a dynamic and dialectic process, adapting it for a networked world. Recognising the ways that information technology is able to disrupt and destabilise the regulation of boundaries, they proposed three boundaries central to the management of privacy in a networked world: disclosure, identity, and temporality [271].

The first, the *disclosure* boundary, describes how we choose to withhold or disclose information about ourselves when creating a public 'face'. Violations of the disclosure boundary occur when information about us is shared without our knowledge, such as photos taken by others or a web search revealing

teenage forum posts that were believed to be lost. The second boundary is the *identity* boundary. When technology mediates our communication with each other, it often disregards the different selves that we present to various social groupings, blurring the line between publicity and privacy. Examples include the resharing of information that was 'already public' in a different context, or digital trails left by cookies and other tracking mechanisms that interfere with the way we regulate our presented selves ('the freedom to make and remake oneself' [10]). Finally, the *temporality* boundary describes the ability for information technologies to create tensions between the past, present, and future ways we interpret and act upon information disclosures.

The disclosure and identity boundaries then, capture the complex nature of personal information as both a set of facts and an expression of the self in a way that Westin and Altman did not [141, 308]. Their use of "genres of disclosure", evocative of work on genres of organisational communication by Yates and Orlikowski [361], also begins to incorporate notions of context into privacy decision making; there are clearly norms and expectations around different types of information disclosure that inform how they are interpreted, and when one might feel that the 'promise' of a genre was broken [271]. The genre of a disclosure—such as a cookie or personal blog post—establishes a relationship between "forms of disclosure and expectations of use" [271], but still fails to adequately capture the full context of these events (such as the difference between an email to a doctor and an email to a tax accountant).

Privacy as Contextual Integrity

More recently, Nissenbaum further develops this understanding of context as a core component of informational privacy, conceptualising it as 'contextual integrity' based on norms surrounding the appropriateness of information in context and flows of information between different parties [262]. Contextual integrity expresses this through the following vocabulary [262]:

- *Contexts*—structured social settings characterised by canonical activities, roles, relationships, power structures, norms (or rules), and internal values (goals, ends, purposes).
- *Actors*—the individuals, groups, or organisations who are sending/receiving/subjects of data flows, and the capacity in which they are acting (e.g. as a parent or as a lawyer)
- *Attributes*—The type or nature of the information being disclosed (e.g. health records or school transcripts)
- *Transmission Principles*—constraints on the flow of information between parties (e.g. confidentiality or reciprocity)

Achieving privacy according to contextual integrity means successfully regulating the transmission of information to various actors so as to maintain the appropriate sharing norms of the contexts in which they operate [14, 229]. This offers a more sophisticated means of breaking down the promises of different genres in a way that can be analysed and compared.

The existence of context-specific norms for information disclosure has been borne out by empirical studies showing that end-users' views on disclosive actions (such as data flows to companies) are influenced by both social and physical context(s) from which such information are gleaned, as well as other factors including the kinds of information being disclosed, perceived norms around disclosures, individuals' relationships with those whom their data are shared, and the purposes of disclosure [30, 36, 104, 187, 206, 209, 211].

When violations of such norms are revealed, the consequences are feelings of betrayal and being "creeped-out", whilst when such norms are respected, people feel "in control" or benefited [304, 331]. Repeated violations of expectations, and the inability to tell whether disclosures are occurring, have been associated with feelings of being perpetually listened to or watched, and even feelings of helplessness and resignation, highlighting potential harmful phenomenological effects associated with long-term violations of privacy [202].

Privacy as Accountability

A less common, but nonetheless interesting conceptualisation of privacy is as (freedom from) *accountability*. Social accountability happens frequently during normal interactions: we give reasons for being late, and label our intentions when there is ambiguity ("it was meant as a joke"). By doing so, we attempt to recast other's interpretations of our actions by giving them additional information or context [57].

While this may seem far removed from the other conceptualisations of privacy discussed here, it captures the external pressures that act on people to share information [97] (a major shortcoming of Altman's theory and those derived from it around online data collection). In the context of daily life, accountability is also used as a tool to "manage the attack surface of the digital on everyday life" [78] by holding others in the household accountable for their use of technology when it disrupts interpersonal social activities. Good examples of this privacy at the level of the household are provided by Porcheron et al. when studying how use of voice assistants is embedded in family life [283].

But while accountability offers a powerful mechanism for protecting the social order from digital disruption, *too much* accountability can be problematic when we consider the role of privacy in individual's identity management and construction. In a culture of sharing, one can feel constantly observed and pressured to share information, and the scope for 'guilty pleasures' or

personal thoughts all but disappears. This nothing to hide, nothing to fear mentality ignores the fact that there are many human activities for which privacy is a prerequisite (e.g. writing a diary). It is therefore important that the actions that we are accountable for are managed and kept private from others.

In the context of device mediated interactions with companies, the calculus is somewhat different. We might initially struggle to hold companies to account because the actions that they take through devices in the home are opaque and unknown. This highlights a potential shortcoming in the application of accountability to actions not undertaken by natural persons; I hold you to account because you could have acted differently. But with UI dark patterns, Skinner boxes, and platforms designed to build habitual use, the ability of people to detect manipulative or otherwise unwanted behaviour by technology is compromised.

But when this behaviour is detected, accountability acts for the same reasons as in interpersonal relationships. Feinberg identifies three reasons for which an entity's failure may be accounted for: (1) insufficient skill or ability; (2) insufficient care or attention; and (3) improper intentions (e.g. cheating on an exam) [112]. The latter is the most common complaint against technology in the connected home, but also a cause for dispute. If a company acts legally in order to maximise profit, to what extent can it be said to be acting with improper intentions?

The corrective measures available against companies also differ from those present in interpersonal relationships. The distant and non-human nature of corporate entities means that solutions can often be separated into those signalling disapproval (e.g. a boycott or call for regulation) and those that interfere with the actions taken locally by the company (e.g. installing an add blocker or obfuscation behaviours [50]). These obfuscations hold companies to account by sabotaging their accounting for our own behaviour—Goffman writes about how a 'good' account (i.e. one that succeeds in recasting perceptions of one's actions) is not always a true one [142]. Obfuscation takes this proposition and applies it much more generally.

The Seven Veils of Privacy

O'Hara encapsulates many of the dimensions of privacy described above in the *seven veils* framework [268]. Here, privacy concerns can operate at the level of (1) conceptualisation, accommodating the many diverse theories of what privacy 'is'; (2) empirical facts about whether particular conceptualisations of privacy are met in a given situation; (3) phenomenology, describing the way that similar actions can *feel* very different depending on context; (4) preferences, similar to Altman's privacy as an optimisation process; (5) norms, the level at which Nissembaum's theory of contextual integrity sits; (6) laws/rules

as dictated by states and organisations, usually to aid in the protection of level 4 concerns; and (7) rights, higher level guiding principles that steer the development of laws and norms, closely linked to conceptualisations at level one.

As can be seen above, the seven veils encapsulate many crucial aspects of the other approaches discussed in this section. While they do not take an ethical position or prescribe certain behaviours, they are nonetheless useful in unpacking the different types of concerns that people have around privacy. They also highlight why privacy research can be incredibly challenging; the study described in Chapter 5, for instance, explores privacy perceptions (3) and generation of preferences (4) through the use of ground truth (2) as well as knowledge of laws (6) and rights (7). The interconnectedness of the veils of privacy not only requires different skill sets (obtaining facts might require technical expertise, but social science methodologies are required when investigating perceptions), but also acknowledgement of the influence that these layers have on each other.

The Privacy Paradox

But the relationship between privacy preferences, intentions, and behaviour is complicated. For decades research has highlighted frequent discrepancies between people's stated preferences towards online privacy and their behaviour [321]. This phenomenon, where people who claim to value privacy take actions to the contrary, is commonly known as the *privacy paradox*, and has been a major source of contention in the privacy literature.

There have been various attempts to explain (and explain away) the apparent discontinuities of the privacy paradox (see [191]). Some studies conclude that the paradox arises as a result of information asymmetries—that people are prevented from making rational choices about privacy due to incomplete information (e.g. Norberg et al. consider users to be broadly unaware of how their data are used [263]). Acquisti and Grossklags take this further, asserting that even with complete information about this, people would unlikely be able to process all of it and make rational decisions in the face of cognitive biases [5].

These cognitive biases are often cited as the cause for study participants repeatedly valuing immediate and tangible rewards over long term abstracts risks [4, 191]. Borrowed from economics, hyperbolic discounting describes how people inconsistently discount future events more heavily when they would occur sooner; many people would happily forgo a future good to preserve their privacy but would be unwilling to make the same sacrifice for something provided immediately. This occurs alongside other biases including systematic underestimation of personal risk compared to others

(optimism bias) [19] and overconfidence in personal knowledge about privacy risks and mitigations [170].

But this *homo economicus* model of privacy often fails to consider the nuance and context involved in making privacy judgements. A common criticism of these approaches is that they take measurements of abstract privacy perceptions and contrast this with behaviour under concrete conditions. In this vein, Barkhuus argues against the way that privacy is often generalised in contemporary research on the basis that notions of privacy are so contextually sensitive that it is impossible to derive *general* answers to questions of privacy. She shows how asking participants questions about topics which are broader than they seem can lead to inaccurate generalisations and incorrect conclusions [24]. Others have concluded that researching privacy in experimental isolation is unhelpful [96], and that privacy attitudes are relatively unpredictable across different scenarios [356].

Supporting Boundary Management

As discussed above, treatment of privacy in the HCI literature is sprawling, complicated, and often contradictory. Barkhuus's criticism of studies sorting people into broad privacy categories (based on the conceptualisation of privacy as hiding information from others) helps to explain why Westin's privacy types fail to convincingly describe privacy behaviours. She goes on to praise the theories of Altman and Palen & Dourish due to the way that they fit technology into already existing cultural practices, but ultimately advocates for Nissenbaum's theory of Contextual Integrity due to the way that it accounts for external norms and values. Altman's conceptualisation of privacy as the process of boundary regulation was undoubtedly a significant step forward; elements of his theory persist in the models generally accepted today. But both Altman's original theory, and Palen and Dourish's extension of it, fail to fully consider the role of others in shaping our perceptions and behaviours. This lack of complete autonomy over the self—creatively described by Goffman as "the peg on which something of collaborative manufacture will be hung" [141]—is key to understanding (and explaining) how people act towards matters of privacy as part of families, societies, and communities.

In response to this, Allen describes how the external factors present when making privacy decisions mean certain types of (usually government) intervention are *required* in order to preserve individuals' ability to make meaningful privacy decisions [10], thereby supporting people in managing privacy boundaries. But even this approach fails to capture the phenomenological aspects of informational privacy (the third veil); knowing that one is being tracked around the web *feels* very different to browsing the internet tracked but unaware. If privacy tools are to truly help people in managing the boundaries observed and crossed by their devices, then these aspects need to be

taken into account. How people feel about privacy matters beyond well-being: secure practices that are perceived to be insecure are likely to be disregarded, and insecure practices that are perceived as secure become habitual. These perceptions propagate through the media as well as informal support networks (e.g. those with lower security motivation are more likely to receive information second hand [86]).

What makes devices in the home environment particularly interesting in the context of boundary regulation is that they are able to observe users across many different contexts (e.g. as working, relaxing, as friends, parents or children). When devices become responsible for filtering and propagating information across boundaries in this way they become responsible for managing complex disclosures, often unwittingly infringing on people's privacy in the process. Whereas humans have the required social and cultural understanding to manage information across boundaries, contemporary devices lack the contextual understanding to do this effectively.

Relating Privacy Preferences, Intentions, and Norms

Research taking a more socio-technical view would therefore suggest that privacy preferences, intentions, and behaviours need to be interpreted alongside the competing pressures and incentives that users face in real-world contexts. As a result, research into online privacy should not divorce the elicitation of preferences and provision of controls from the ebbs and flows of daily life. In a study on user perceptions of smart energy systems, for example, Goulden notes a distinction between perceptions and reasoning of users as *consumers*, for whom energy is "a good to be expended in pursuit of personal goals", and as *citizens*, who contextualise their usage and practices in relation to wider norms and societal values [143]. These modes of thinking map directly to Wilson's functional and instrumental views of smart home technology, and we will return to them later when discussing methodologies that allow us to transform abstract problems that concern citizens into concrete ones that concern consumers in Chapter 3.

A major theory as to why end-users' privacy preferences are seldom predictive of subsequent action is that end-users' preferences are not sufficiently supported by an understanding of privacy risks or awareness to justify their views. Studies have demonstrated that end-users (and even many experts) lack awareness of the ways in which their data are transmitted and used, whether via the Web, through smartphone platforms and apps, or connected devices in the home [5, 367], and that people cite trust in manufacturers and regulatory action to protect them [320]. Such findings are unsurprising given that user tracking and information disclosures are rarely disclosed in meaningful ways to end-users of connected systems [320].

Indeed, when information disclosures are made apparent in a form that users can understand, individuals have been shown to articulate more specific and actionable privacy preferences [133], as well as views on ethical, economic (business models), and political dimensions of the data economy [336, 337]. Relatedly, users often cite the infeasibility or unavailability of alternatives as reasons for not taking further action despite awareness. In fact, greater awareness of the reality of the nature of data collection, in the absence of available mitigations, may also create a sense of helplessness and resignation [337]. Such findings suggest that people need to be aware of available controls and have strong beliefs about their efficacy in order to be motivated to act, as demonstrated in the context of security behaviours [205, 320, 343]. It is notable looking back at the technical projects above that aim to inform privacy concerns around smart devices (see Table 2.1) that they only address the transparency of data flows, leaving out the contextualisation and provision of controls. This would suggest that there is an opportunity to improve significantly on prior work by combining existing approaches with the other techniques identified in this section.

The Limits of Privacy

When we examine the 'human factors' of digital privacy, particularly in the smart home, there is a tendency to characterise them as a series of seemingly unsolvable problems. The privacy paradox suggests that (broadly speaking) people do not care about privacy, expending little effort to protect their devices and data and trading even this off when offered promised or real functionality in return [321]. One of the reasons given for this is that people lack awareness of behaviours that are concerning, as well as the knowledge to identify such behaviours as concerning (i.e. they would not know to be concerned even if they could see them) [5]. At the same time, people often feel like they have no way of changing their circumstances ('the fantasy of opting out') [50, 304]. Given the relative lack of success of initiatives addressing one of these issues, it would appear that the way forward lies in a combination of the three: increased transparency of device's privacy-relevant behaviour, education interventions around what these behaviours mean and why they are concerning, and a means of enforcing one's will onto those device behaviours.

These efforts appear to flow against the design of many contemporary smart devices, where "smooth swipes, soft pings, and gentle buzzes [...] shield us from computational and infrastructural complexity" [280], and where compliance with privacy regulations has, in cases, been partial. This suggests that initial efforts in this area will likely come from third parties, providing information about devices from observations and integrations while also providing the context required to interpret device behaviours. Being external to device manufacturers also affords a level of neutrality and trust that is

lacking in the major players in this space. A review of existing projects with similar goals highlights that none combine the three approaches identified above to address issues of privacy (transparency of data flows and behaviour, knowledge of norms & laws, and effective controls with which to enforce preferences). This opportunity to create devices and services layered on top of existing smart devices will be explored further in Chapter 5.

But there is a limit to how much privacy alone can address the problems present in modern smart devices. Section 2.1 began to uncover the different types of ethical concerns present in smart home devices, and while privacy was the most prevalent amongst these, other issues like accountability and distortion of the social order of the home also greatly impact people's daily lives. Scholarly work on digital technology often attempts to adapt the concept of privacy to take on and subsume these additional concerns, as when discussing privacy as accountability or privacy as identity management, but in doing so it risks losing the nuance of what makes the original behaviours so problematic. Viewing obfuscation behaviours as a means of disrupting a company's account of private digital activity does not so much consider new issues previously outside the realm of privacy but rather consider conventional privacy problems through a new lens.

Perhaps unsurprisingly then, privacy has become the sole tool used by regulators to address these problems. The general approach to the regulation of smart devices in the West is very much centred around privacy, partly due to the adaptation of existing (non-digital) privacy laws, where lawmakers have traditionally been reluctant to impose restrictions on what can happen within the boundaries of the home, and partly due to the ability of privacy to (partially) express some of the other concerns described above. As a result, the GDPR is concerned primarily with data protection and mandates 'privacy by design'; it contains no mention of accountability by design, or how digital technology should preserve the existing roles and relationships within the home. Instead, companies are made accountable to consumers through the publication of privacy policies and the state through reports to privacy bodies (such as the ICO), and the effects of the GDPR stop at the boundary of the home without considering the effects of data collection and control on interpersonal relationships or the management of co-created data. Other approaches are therefore required in order to fully engage with the other issues presented by smart devices, moving away from privacy as a blunt one-size-fits-all approach to regulation and towards a model that more holistically encompasses the platforms behind devices and their influences on life in the home.

2.3 Ethics & Respect

The problem spaces highlighted thus far, and the associated investigations conducted in later chapters of this thesis, are concerned with asymmetries between the people using (or affected by) connected devices and those who design, manufacture, and truly control them. But computer ethics extends far beyond privacy, autonomy, and the social order of the home. There are vast inequalities that exist in virtually all of today's societies. Too often, these are perpetuated and systematised by AI systems, many of which power smart and connected devices (e.g. gender performances by voice assistants that reinforce existing gender stereotypes). Motivated by all of these issues, the goal of this thesis is the rethinking of 'smartness' in a way that truly works to the benefit of the people using and affected by smart technology. There are many moral theories and concepts that could be used to categorise and justify different types of design decisions, but as the thesis progresses the results of its research point to the concept of respect as being the most effective concept that can be used to anticipate and design for ethical concerns with connected devices.

This section begins by focusing on these societal asymmetries, mapping out some of the most common ways that people are disrespected, marginalised, and excluded by modern societies and motivating the subsequent adoption of respect as a lens through which to analyse and address the ethical concerns uncovered by prior work. We then examine previous treatment of respect in the HCI literature.

Exploring Disrespect and Social Exclusion/Marginalisation

A crucial first step in the process of understanding respect and its importance is to explore its inverse—disrespect. Disrespect has a deep connection to, and defining role within, processes of marginalisation and social exclusion. These terms refer to the ways in which individuals are denied full and equal access to rights, opportunities, and resources normally available to others on the basis of a person's race, gender, sexual orientation, religion, disability, cultural background, or other attributes or affinities [6]. Where people are not seen as deserving of fundamental human rights because of their group membership, such discrimination is dehumanising; "if personhood is seen as being not simply human but human in a way that is valued and meaningful, then individuals can be persons to a greater or lesser extent" [164]. Work examining the dehumanisation of marginalised groups has shown it to be present in stereotypes around race and ethnicity [138], gender [66], age [46], and disability [164] (amongst others).

Marginalisation can occur at any point that a person "deviates in any way from the perceived norms of a population" [349], which means that people

may experience being marginalised in different (and multiple) ways at various stages of their lives based on factors over which they largely have no control. Despite such connections, there has been little work looking at respect as it relates broadly to marginalisation abstractly, as opposed to that of specific groups. Thus we instead examine the roles respect plays in specific groups – in particular, the ways that disrespect is seen to articulate, communicate, reinforce, enact, and justify violence based on race, gender, age, and disability. The literature on disrespect is vast and diverse, and as such this section is intended only to outline the major issues in the space in order to motivate the subsequent use of respect as a lens with which to create devices that truly benefit their users.

Race

Disrespect acts out, communicates, and reinforces racism in numerous ways, oppressing, marginalising and dehumanising racial and ethnic groups. In Glasgow's philosophical essay "Racism as Disrespect", he describes how racism "subtle or overt, intentional or unintentional, conscious or unconscious" [136] can be expressed through belief, attitudes, or behaviours, intersecting with the varied conceptualisations of respect outlined above. From internal to external, this can be: 1) a conscious, cognitive view of nonwhites as inferior, non-equal or "other" creating disrespect through beliefs and representations; 2) attitude-based racism that involves expressing or holding views of dislike, disdain, or antipathy towards nonwhite people; or 3) racist actions and behaviours that harm or disadvantage people directly or indirectly. Racism can be perpetuated despite good intentions, for example a well-meaning individual with racist beliefs or attitudes may take action with the intention of expressing care that, through its paternalistic nature and lack of ascribed agency, perpetuates the racist attitudes or beliefs that motivated it.

Van Dijk and others contend that the ways in which racism is most often expressed have become less overt and more subtle/indirect over time [334]. Essed used the term 'everyday racism' to capture a variety of recurring slights, abuses, and put-downs which people of colour regularly experience [108]. These microaggressions are difficult to deal with due to the fact that they can be readily denied as being racist or justified for other reasons (e.g. by depicting the recipient as criminal) [334]. So in interracial contexts, disrespect can be include perpetuating or failing to acknowledge 'older' forms of racism, ignoring the fact that Whites are often privileged in that they can take respect for granted throughout the larger society whilst people of colour have to earn or prove they are worthy of receiving respect from Whites [269]. Disrespect also surfaces in glossing the exhausting and constant experience of being the recipient of everyday racism, failing to consider the lived experience of many people.

Gender

As with racism, discriminatory practices and behaviours towards those identifying as women and minority genders pervade society. Disrespect is shown in the way that women are generally paid less than their male counterparts for the same jobs [241], with a deficit of evaluative or recognition respect meaning they are often passed over for promotion, particularly at high levels (the 'glass-ceiling') [76]. Interpersonally, workplace harassment was and is often inadequately dealt with by employers, and occurs against a backdrop of majority-male perpetrated physical and sexual violence in the home and the community, with victim blaming being common [192].

Where Goffman's deference looks at respect as including being addressed in one's preferred manner, the influence of gender roles is deeply entrenched, embedded in language in the form of gendered nouns, outcry over gender neutral titles and pronouns, and use of the generic masculine (e.g. using mankind to refer to the human race) [275]. This 'male as default' approach excludes others from consideration in the design of many everyday things, including in safety-critical environments such as automotive safety [212] and healthcare [230].

Alongside discrimination against women, transgender and non-binary people are often excluded entirely by governments, organisations, healthcare systems, and society [216, 252]. Many of the microagressions suffered by transgender and non-binary people relate to a lack of respect, such as violations of directive respect and conventions around demeanour-based signalling when others use incorrect or offensive forms of address, ignoring avoidance rituals by asking personal questions or exoticising them, and lack of recognition respect when others deny their experiences [251].

Disability

The representation and oppression surrounding disability often fundamentally differs from that of other forms of discrimination. In many contemporary societies the state not only legislates for obligations on employers and others to ensure the equal status of people with disabilities, but also defines and operationalises what it *means* to be disabled (e.g. through the disbursement of social security) [164]. This arises in part from the medicalisation of disability during the late 19th Century, which framed disability as the 'cause' of people's functional impairments, rather than societal factors [25]. Paradoxically, equal treatment under the law can often lead to disabled people and the disrespect they face being ignored, as "they are designated so as to disappear, they are named so as to go unmentioned" [316]. In this way the treatment of the disabled 'as if they were normal' erases any difficulties they do have, thus reinforcing inequality [164].

In a personal and social context, being different because of a disability often means being *less*, and being ascribed less agency, leading to treatment that in extreme cases invalidates and infantalises those with impairments [368]. By being perceived as 'sick, suffering, diseased, and in pain', those with impairments represent some of the darkest fears of the able bodied [157]. These perceptions can lead to disrespect—failures of evaluative respect—about contributing to society as well as when benefiting from it. Barnes writes about how modern associations of economic contribution with personal worth can give rise to perceptions of disabled people as 'useless', and how when they appear happy despite being unable to fully enjoy the social and material benefits of everyday life they are applauded for their 'exceptional courage' [25].

Age

People are often exposed to prejudice around age and ageing from an early age, with fairy tale and media portrayals of older adults emphasising stereotypes of decline and incompetence [364], and the idea of 'reading ages' and school grades/years serve to associate age with development and (in)capability [258]. Fear of ageing contributes to explicit and internalised ageism in a process beginning during childhood where the young slowly internalise ageist conceptions as they come to associate older adults with the disabilities that often arise with age [258]. Consequently, by the time people reach old age, they are likely to have spent decades internalising the ageism implicit in society, leading to greater acceptance of the stereotype of older adults as more passive and dependant [56]. Examples of the types of poor treatment shown to older adults include not being taken seriously by healthcare professionals [165], abuse in the context of care [197], and being treated as less able to hear and understand during conversation [134].

Colonialism, Culture, and the 'Other'

Disrespect as a form of social exclusion and marginalisation also manifests through cultural imperialism and colonial ideology or discourse. Postcolonial theory highlights various ways in which dominant cultures, and globalised imperial hegemony, erase different cultural meanings and ways of knowing [296, 149]. Here, 'The West' can be understood as a historical rather than geographic concept, as a discourse through which power operates [149]. Through a process Said called 'orientalism' [296], an implicit binary and reductionist depiction is created of non-Western culture as Eastern 'Other', or 'subaltern' [312]. Through cultural imperialism, members of subordinated and colonised cultures are made 'to experience how the dominant meanings of a society render the particular perspective of one's own group invisible ... and mark it as the Other' [362]. Even attempts to be respectful to colonised groups or

minority populations can be misguided, where they are informed by colonial ideology or serve to exert soft power over religious minorities through 'paternalist multicultural respect' [195]. As Ahmed has argued, in failing to 'ascribe personhood to any non-White Other, even their 'love' for the Other is demeaning, and their 'respect' is intrinsically disrespectful' [7].

Intersectionality: Overlapping Marginalised Identities

Originating in work by Black, Latina, and Native women since the 18th century [71], intersectionality was coined in 1989 by Crenshaw, who demonstrated the problematic ways in which anti-discrimination law treats race and gender as discrete categories [83]. Intersectionality considers the diverse collection of overlapping identities that situate the relationships between individuals, technology, and social systems, examining how the interactions *between* identities contribute to the ways that people are treated (rather than simply considering these identities in isolation). These bottom-up constructions of identity "[problematize] the notion of static categories and the systems of power that get to create them" [354]. Sharing roots in feminist theory, there are clear parallels between intersectionality and respect as care. Both take a more personal approach to considering people and their experiences, in particular around combinations of circumstances. Previous work on integrating feminism into HCI [23] has inspired similar efforts with intersectionality, although uptake has been slow: a literature review of almost 14,000 CHI papers by Schlesinger et al. found only 24 contributions that focused on more than one category of identity (such as race, gender, or class) [299, 354].

Respect and HCI

In the HCI literature there has been very little discussion of respect in prior works. Lea Kissner frames respect as prioritising "the privacy, security, and anti-abuse" of users of technologies [185], recognising that as platforms scale to millions or even billions of people, so to do human needs and feelings. The diversity of human vulnerability needs to be considered when building such systems, and this can only truly be accomplished through deeper understanding of people and societies.

Patrizia Marti discusses respectful design as that which "redresses the balance of how technology takes into consideration the unique skills of the targeted users including their values, motives and beliefs." [228]. The paper does not elaborate on what is meant by respect, but stresses the need for cultural and emotional sensitivity when confronting social issues.

While these contributions are somewhat vague in their use of respect, using it implicitly as a wrapper for a selection of values rather than a concept in and of itself. They do, however, both speak to a need for greater consideration of

the impact of smart devices on peoples' relationships and emotions during the design process, highlighting the social side of respect that is often missing from the application of privacy to problems of smart home ethics.

Whereas people have social responses to and understandings of technology (such as conceptualising companies as people [337]), privacy-based approaches mainly focus on disclosure of information and the subsequent consequences. In this way respect is more closely aligned with the design approaches in HCI that specifically take users and values into account—we will explore the relationship of respect with these approaches in the next chapter.

Chapter 3

Methodology

"If you have something that you don't want anyone to know, maybe you shouldn't be doing it in the first place"

— Eric Schmidt, 2019 *CNBC Inside the Mind of Google* Interview

Contents

This chapter gives an overview of the theory and methodology that informs the thesis, much of it specific to HCI. It begins by outlining the methological approaches that drive and inform the research presented in later chapters, relating these to the work on respect presented in Chapter 7. It then details the practical approaches used to gather data, in a way intended to be complementary to the more specific methodology sections accompanying individual experiments. The chapter concludes by describing the efforts taken to ensure that the research presented in this thesis is transparent, ethical, and reproducible.

3.1 HCI Frameworks and Principles

Since the inception of computers and subsequently the field of human-computer interaction, both have evolved significantly. Initially inheriting may of its techniques from industrial engineering and ergonomics (the 'first wave' of HCI), the field has become increasingly occupied by the human side of the equation. With the second 'wave', focus shifted towards the ways that people interpret and process information using concepts such as sensemaking [281], finally leading to the third 'wave' that digs deeper into the lives, identities, values, and cultures of those who use digital technology, as well as the ways in which that technology affects their lives [98]. This process of change and development has resulted in the emergence of a number of different approaches towards the design process that determine what is prioritised and how HCI work is undertaken.

In this section we briefly outline the major frameworks and principles in contemporary HCI that relate to the goals of the thesis, comparing and constrasting them to each other and the lens of respect developed in Chapter 7. This provides the basis for the methodological stance taken by the research presented in the thesis and situates the use of respect in the design of smart devices relative to existing approaches, justifying its use as a means to shape the evolving landscape of smartness.

User-Centred Design

Popularised by Norman and Draper in 1986, user-centred design (UCD) focuses the design process on those who will be using a system [265]. By asking what their needs, goals, and abilities are, UCD aims to align the process of using a system with concepts and actions that the user is already familiar with.

Encompassed within UCD are a number of different design practices, such as participatory design and co-design, that see designers sharing the process of design with those who will be using their products. These techniques are frequently used within HCI to help designers understand the intricacies of wicked problems in various contexts (e.g. [359, 360]).

Another widespread technique that falls under UCD is the use of personas, as championed by Alan Cooper [73]. These and similar techniques aim to prompt designers to integrate their solutions with existing aspects of a user's life and workflow, rather than seeing them as someone to be designed *for*.

While this is an important step towards respectful design, in which representation and identity are important, it arguably does not go far enough towards understanding the wider context in which the use of a technology is situated. In the 80s and 90s, software might have had a relatively narrow range of uses

and users (e.g. on office suite used by employees for business on desktop computers), but the advancement of home, mobile, and ubiquitous computing forces the consideration of a much wider range of people and the circumstances they might have. Connected homes are a prime example of this; a smart speaker might be used via voice or a smartphone to assist with calendar scheduling, children doing homework, compiling a shopping list over the week, or remembering to turn off the bath taps.

In this way user-centred design can be homogenising—designers adapt ideas to fit infrastructure and routines that are already in place (e.g. the question is how to represent a tree model of file and folders [73] rather than how to create alternatives that fit a new use case), and in co-design scenarios users similarly re-purpose solutions they are already familiar with (e.g. private browsing applied to connected cars [360]). There is an inherent risk of paternalism and stereotyping when designers attempt to solve problems for people who's lives they have knowledge of but have not experienced. This would suggest that in order to effectively and respectfully serve the people who use their products, designers need a much richer understanding of users and their lives.

The central tenets of UCD have clear resonance with some of the basic framings of respect. For instance, in centring users and prioritising their needs above other ends, UCD could be seen as embodying a version of respect for persons articulated by the categorical imperative; in designing systems that address people's needs, we treat them as an end in themselves, rather than as a means of achieving someone else's objectives. In practice, this might mean prioritising the fulfilment of those needs (e.g. maintaining social connections) above any other interests (e.g. driving user engagement for profit, or enabling manipulation by advertisers).

As well as the generic orientation of UCD reflecting certain kinds of respect, we can also consider how specific practices within UCD may or may not be respectful. Particularly relevant here is the use of personas to direct the design process towards the unique needs of a system's users. While undoubtedly useful in eliciting and understanding requirements, using personas necessarily abstracts away from individuals by representing a large number of people as a small number of personas with definable characteristics. Designers are not unaware of this; *About Face* warns about the dangers of using stereotypes rather than archetypes, and suggests using personas with gender, ethnic, age, and geographic diversity where this is "unimportant to the design or its acceptance" [74]. However, as explored in the background material on disrespect, showing respect in terms of representation and design might be perceived differently depending who is respecting who—attempting to do so without knowledge of current and historic injustices risks furthering the idea of a colourless, genderless design subject. The focus on concrete personas over the abstract user moves towards the care conceptualisation of respect as

one that is concerned with individuals (Cooper even uses the word 'empathy' to describe the benefits of using personas [74]). But this still fails to design for actual individual users, with reducing the focus on designing for edge cases often described as a *benefit* of following a UCD approach.

Finally, the varied and rich conceptualisations of respect that have emerged since Kant's work might also yield as yet unconsidered dimensions of 'user needs'. For instance, the kind of needs emphasised in *care* respect, such as to be looked after and feel cared for, or the need to feel like one belongs to a group [90], may be examples of latent needs that are typically not articulated within UCD practice or addressed directly by technical systems. Furthermore, devices can meet user needs in various ways, and doing so respectfully might entail more than just meeting them functionally. Rather, as articlated by perspectives such as dramaturgy, respect can require not only doing the right thing for someone but conveying certain appropriate sentiments to them (e.g. deference) in the process; user needs relating to respect might therefore have to be met at an experiential and emotional, as well as functional level.

Human-Centred and Value-Sensitive Design

If our criticism of user-centred design is that it can be homogenising—designers adapt ideas to fit infrastructure and routines that are already in place, and users often re-purpose existing solutions during participatory design techniques (e.g. private browsing applied to connected cars in [362])—an alternative approach would be to have designers gain a much richer understanding of their users and their lives.

Going beyond the steps taken in user-centred design, human-centred design (HCD) calls for a much more in-depth understanding of the users of a system. By immersing themselves in the life, work, and culture of the people who use their creations, designers aim to more fully comprehend the needs of complex individuals navigating complex circumstances. In this way, it is possible for designers to understand more about the problems faced by their users than the users themselves may realise, and subsequently to adapt technology to best serve their needs and desires.

But HCD still suffers from the general inability to envisage *new* ways on interacting with platforms and devices. Perhaps then, rather than designing for the varied situations that users find themselves in, we can design according to the *values* that they possess? This value-sensitive design (VSD) approach embodies the philosophy of third wave HCI and provides a clearer framing for the many competing dimensions that factor in to the success of design solutions.

Of course, the question when designing for values is *whose* values are to be designed for? While every design artefact reflects human values to some

extent, VSD emphasises the well-being of those who use technology: "what is important to people in their lives, with a focus on ethics and morality" [120]. This brings external considerations, such as usability, into scope as factors that require consideration alongside those purely ethics oriented. In this way, value sensitive design prompts us to understand the culture, beliefs, and values of a given set of people when designing for them, and in doing so sidesteps the need to make normative declarations about particular theories and solutions (e.g. designing for people's particular conceptions of autonomy rather than one chosen beforehand). Doing this avoids the problem encountered with UCD which is more concerned with solving perceived problems—working at the level of values gives the latitude to address underlying problems whilst being cognisant of *why* issues arose in the first place. Of course, what values are deemed to matter most to users will inevitably be filtered through designers own perceptions, particularly where these are balanced against safety, well-being, or economic factors. There is a clear risk of paternalism here too, and Friedman suggests that the balance should often be struck in favour of ethics, particularly given the tendency for profit-seeking to ride roughshod over moral concerns [119].

VSD is much closer to many conceptualisations of respect in the way that it considers the wider ways in which technology can help or hinder those affected by it, as well as its ability to enrich people's lives and promote their flourishing. For this reason VSD represents the starting reference point for the research presented in this thesis, which is iteratively developed to arrive at the conceptualisations of respect for smart device design presented in Chapter 7.

One question to be considered in this thesis is whether respect could itself be a value that could be designed for in the context of VSD. One thing that the varied and sometimes conflicting accounts of respect described in Chapter 7 make clear is that it does not make sense to design for respect as a singular concept—one can refer to 'respect' in conversation and have context dictate the type of respect one means, but this is not specific enough for design. Instead, designers should consider the individual conceptualisations of respect in the context of the product being created, such as designing around respect for or from a particular group in a particular way in line with work by Mills on self-respect and respect for others post-Kant. Proposals have also been made to integrate the ethics of care into VSD in much the same way as with respect [338].

As mentioned above, a problem that arises when practising VSD is how one selects the values that are most relevant to those who will use a system. There are countless cultural, social, and personal values that *could* be designed for, but most are irrelevant in the context of an individual product (e.g. local etiquette might dictate that a shopkeeper initially refuse payment for an item, but this is unlikely to be appropriate when shopping online). Respect as a

lens offers a powerful tool in helping to narrow down this potentially infinite selection. There are a number of values that underpin the different kinds of respect described in Chapter 7, such as equality, autonomy, identity, and care; exploring which kinds of respect are most relevant to the situation at hand therefore leads to a much smaller and well defined set of candidate values. In addition to this, we believe that the ways in which VSD encourages designers to better understand people's culture, beliefs, and values prioritises a type of approach which is closer to respect as care than other design methods. In this sense, respect is not (just) another collection of values to be considered within VSD, but a way to structure and relate to other values. Finally, the understanding of care respect developed in Chapter 7 is helpful in balancing a duty of care for others with the risks of paternalism.

Fairness

While issues of societal bias, discrimination, and fairness in sociotechnical systems have long been studied in a range of disciplines [121, 128, 160], the last decade has seen a significant growth of interest in these topics amongst computer scientists, often borrowing from or in collaboration with other disciplines such as law, philosophy and social sciences [274, 102, 26, 37]. Alongside transparency and accountability, fairness has emerged as a key design goal for machine learning systems [65]. Like respect, fairness is a concept which admits multiple interpretations across a variety of disciplines. Some of these have been implicitly or explicitly appealed to as the philosophical grounding for various statistical fairness metrics proposed within recent fair machine learning literature.

Some understandings of fairness more closely align with respect than others. 'Distributive' or 'allocative' fairness, understood in terms of the distribution of benefits and burdens across people in society [102], can be contrasted with *representational* fairness, which addresses how systems stereotype, denigrate, under-represent, and erase certain groups [82]. An example of representational harm would be language models which replicate gender biases [45, 58]. This distinction reflects the debate in political philosophy about equality of distribution vs equality of recognition [117]; an ML model might be unfair in the sense that it doesn't allocate benefits and burdens fairly, and also, separately, in the sense that it represents social groups, cultures, and identities in harmful ways [37].

These differing notions of fairness bear a more or less close relation to respect. Distributive fairness, which has arguably been the dominant influence on fair machine learning research to date, is arguably only tenuously connected to respect; it is possible to imagine a system which ensures benefits and harms are applied according to some egalitarian principle, but which nonetheless disrespects certain groups. For example, predictive text systems modified

to ensure equal accuracy between majority and minority dialects might be distributively fair, but still perpetuate representational harms if their predictions for a minority population are perceived as cultural appropriation, or reproduce psychosocial injustices [220]. Meanwhile, fairness understood in terms of representational harms seems more closely related to respect; echoing Darwall, Crawford describes the problem of recognition as one of respect for humanity and personhood, and the overlaps with the concept of recognition respect as described above are clear [82].

Another difference between fairness and respect in these contexts is that the former generally concerns individuals in so far as they are members of groups, whose outcomes are compared. Even concepts like 'individual fairness' [102] which aim to ensure that similar individuals receive similar outcomes still essentially equate an individual with their position in the feature space of a machine learning model, which is based on generalisations from other people in the training data [38]. In this sense, individuals are not treated as individuals, but on the basis of generalisation; and arguably, as means to ends (in this case, the end of achieving a 'fair' model). As such, they may violate, or at least fail to satisfy, the Kantian notion of respect as the genuine treatment of people as ends in themselves rather than solely as means.

Accessibility

The HCI community has a long history of undertaking work on accessible, inclusive, and universal design, moving away from the notion of accessible systems as a 'special need' towards a more inclusive approach centred around the lives and goals of those at the margins [199]. Core to inclusive and universal approaches is the expansion of the target group of users being designed for. Instead of just considering who can use a product, these 'counter-exclusionary' methods identify who *cannot* use a product, and therefore what needs to be improved to widen access [199].

The idea of universal access (or 'direct accessibility') as the belief that everyone should be able to use a product or service without assistance, regardless of their particular needs or disabilities, is a concept deeply concordant with the deontological respect for persons that forms the basis for human rights: as a member of the human race, each person should expect to be able to use a product, and designers have a moral duty to accommodate this in their work. An example of this might be Amazon's Alexa software, which can be interacted with through a wide variety of languages, input modalities (e.g. speech, touch, keyboard), and form factors, allowing users to choose how they wish to use the software. Indirect accessibility, where products facilitate access through compatibility with assistive technologies, might be seen as closer to other types of respect where people recognise a characteristic of another that requires a certain response (recognition respect), with designers

recognising the specific needs of the people using their systems and adapting them accordingly (e.g. screen reader compatibility for the blind and partially sighted).

Beyond this, designing for accessibility and inclusivity are closely related to feminist perspectives on respect as care, and accounts of respect as ascribed agency (i.e. taking actions that affirm the agency of others, such as by allowing them to choose whether they wish to be assisted or not). This directly relates to the ways that design decisions intended to increase accessibility can actually constrain the agency of those using a system by taking the choice to *ask* for help away from people. Similarly, care respect asks that we understand a person and their needs as an individual but connected being, and in doing to precludes the prescription of functionalities 'for' particular disabilities (e.g. a 'vision impaired mode' that increases the font size). Instead, it might suggest approaches based on *directive* respect, allowing people to customise the behaviour of the systems they use regardless of the extent to which they identify with a particular label, if at all.

Here we can see how respect can help elucidate the values underlying accessibility as a design goal, as well as how it can help draw out the tensions between two differently motivated approaches; some grounded in the belief that products should be designed to be usable by everyone because they are human beings, and others that recognise a range of disabilities and allow people to adapt their own individual approaches to use their devices comfortably.

3.2 Practical Methods

This section details the methods used to gather research data throughout the thesis. Following Carroll's paradigms of HCI [60], the research presented here is distributed between the paradigms of *description*, observing and understanding how people respond to smart devices, and *invention*, creating real and fictional design artefacts that drive the research process.

A popular and more obvious choice is to survey or interview people living in homes where smart devices are already embedded to gather perceptions and self reported behaviours [355], potentially as part of longitudinal or mixed methods studies. This can take the form of direct measurement of participants [270], or the subsequent uploading of device logs [301]. In order to study the adoption of new technologies, households can also be provided with new devices, with researchers observing how they become integrated into daily routines and practices [325]. Finally, there is a rich collection of design-led work on smart homes, often incorporating methodology from critical design.

Surveys

The surveys presented in this thesis were conducted using JISC Online Surveys[1], a survey platform previously known as Bristol Online Surveys. JISC is a UK non-profit organisation founded to support post-16 education and research. It is funded by the UK Government. As Online Surveys provides only basic analysis tools, results were exported for offline analysis in R. Recruitment was handled by Prolific Academic[2], a UK company providing participant pools to researchers. Prolific protects participant's personal information (providing demographics and a randomised identifier), and ensures that participants are fairly compensated for their time. Both JISC Online Surveys and Prolific Academic are fully GDPR compliant.

The welfare of participants is obviously a core concern when conducting surveys. The surveys undertaken as part of this thesis follow the British Psychological Society Ethics Guidelines for Internet-Mediated Research[3], built around four main principles:

1. Respect for the Autonomy, Privacy and Dignity of Individuals and Communities
2. Scientific integrity
3. Social responsibility
4. Maximising benefits and minimising harm

To this end, the ethical review process undertaken for surveys included consideration of participants' comfort, well being, and autonomy. This included making questions optional when asking about potentially sensitive or uncomfortable topics, as well as restricting the focus of questions to what was strictly necessary (e.g. asking how participants use technology to support their relationships, rather than asking about the relationships themselves).

For a seemingly simple task, constructing surveys requires great care. Questions must be precisely constructed in order to convey meaning to non-experts in a few sentences. Particularly for subjective measures, pilot surveys and within-subject configurations can help to reduce the variance associated with question interpretation. The lack of supervision can mean that survey participants may lose engagement with tasks. Attention checks can help identify participants who give low quality answers, but treat the symptom rather than

[1] `onlinesurveys.ac.uk`

[2] `prolific.co`

[3] https://www.bps.org.uk/news-and-policy/ethics-guidelines-internet-mediated-research-2017

the cause—it is better to distil surveys down to key questions, or to break longer surveys into segments than can be completed over several sessions.

Subjective or experiential questions are often measured using Likert-type scales (e.g. *approve, somewhat approve, neither approve or disapprove, somewhat disapprove, disapprove*) [210]. While traditionally five item scales, there is evidence to suggest that extending these to seven point scales leads to more accurate results as participants need to interpolate scale items less frequently [115].

Interviews

Participant interviews conducted for studies in this thesis were semi-structured and conducted either at the participant's home or at the Department of Computer Science in Oxford. Interviews were conducted by one or two researchers, audio recorded, and transcribed by those same researchers prior to analysis. References to named individuals and locations were redacted, and audio files were destroyed following transcription.

Analysis was conducted using NVivo 12, which contains several affordances for the efficient labelling of transcripts and structuring of codes. NVivo projects can also be easily shared or printed for comparison with the work of other researchers. Thematic analysis was carried out on interview data following Braun and Clarke [47], who define the following major stages of the analysis process:

1. *Familiarisation*—where the researcher listens to, reads, and rereads interview content in an attempt to immerse themselves in the data collected.

2. *Coding*—researchers generate an initial code for each unit of research data, describing its features in relation to the research questions being answered

3. *Searching for themes*—Using the familiarity gained from the previous steps, researchers construct themes that encapsulate many codes into an overarching narrative.

4. *Reviewing themes*—Newly created themes are reviewed for consistency and integrity (i.e. whether they accurately reflect the collected data). Themes may split, merged, or removed as part of this process, and it may be necessary to return to the previous stage in order to search for new themes.

5. *Defining and naming themes*—In order to fully define each theme, researchers construct a concise overview (often illustrated with quotes, naming each theme in a way that clearly conveys its narrative.

6. *Writing up*—Identified themes are used to weave a story that will be used in research outputs, comparing patterns and data points both within and across themes.

The light and flexible nature of thematic analysis as a research methodology makes it well suited to initial research explorations, such as the research questions given in Chapter 1. Other common qualitative social science research methods (e.g. grounded theory [135]) often require more research data in order to produce comprehensive models of behaviour. These approaches are therefore more appropriate for follow up investigations that seek to explain the results of earlier exploratory ones, rather than as an initial tool.

Another important consideration when interviewing participants is developing rapport. Conducting interviews in locations familiar to participants can help to put them at ease, and staring with broader 'warm up' questions helps to break the ice, especially helpful given that interviews are typically more one-sided than normal conversations. The results of doing this well can be surprising—often participants will volunteer additional anecdotes and information at the end of the interview, and these 'just one more thing' exchanges are often amongst the most valuable (e.g. the car mechanic example in Chapter 4).

The use of semi-structured protocols allows for investigation of interesting and relevant topics that might be hinted at during the course of an interview. Especially when the study is seeking to explore perceptions, these hooks into people's mental models can serve as a useful aid when interviewing subsequent participants.

Adapting Interview Protocols: A Retrospective

When asked directly, people often find it difficult to give a detailed response to questions about privacy and other ethical concerns in the home. As a researcher, one might spend a lot of time pondering privacy or autonomy concerns around certain products and the effect that it has on daily life. However, for the most part, people do not spend large amounts of time pondering questions of data disclosure, or why a particular device makes them feel uneasy. This can make it difficult to organise one's thoughts and give responses under pressure, often leading to surface level answers, recital of opinions from elsewhere that they believe the interviewer wants to hear, or non-answers. Some examples of this arose in the pilot surveys for the study in Chapter 4:

"Bit of a silly question if I'm honest."
"It doesn't have a soul!"
"It's a robot."

To combat this, the methodology used in the study adapted, with participants provided 'functionality cards' that served as mental (and physical) scaffolding they could use to articulate their thoughts. In this way we were able to preserve interactions that might otherwise have been lost, such as participants reaching out to touch a card before pausing and changing their minds, whilst still being able to capture spontaneous responses to the interview questions.

The methodology used in Chapter 5, however, was designed to encourage participants to think through the concepts discussed in the interviews more thoroughly. The six week deployment of the technology probe allowed for new concepts and thoughts to be internalised and used to reinterpret previous events. The repeated introduction of privacy and data protection concepts over the course of the study led participants to apply these to the world around them:

> It's funny. I was actually thinking about that as I was driving back from [activity] this morning, I was thinking that you were going to ask me a question about that. [P2]

This does not diminish the usefulness of more spontaneous responses; allowing participants time to develop their answers was appropriate in the context of the study because the research questions asked how well participants would internalise the content made available to them. In other cases, as with Chapter 4, existing subconscious perceptions are likely to be lost through overthinking.

Vignettes

Vignettes are sketches of fictional scenarios, often used as a research tool in surveys and interviews to collect situated data on values, beliefs, and norms. Vignettes are common in sociology and have more recently been applied to HCI [77, 169]. During data collection, the respondent is invited to imagine the sketched scenario, drawing on their own experience to give a response. In utilising fictional details, this method is similar to speculative design approaches, which have been used to study views on the ethical implications of technology use [203, 213, 335]. While this approach necessarily moves away from the lived experiences usually sought from qualitative research methods, responses to vignettes have been shown to often be predictive of behaviour [169]). Responses to imagined scenarios can also be used to understand more general underlying attitudes and perceptions.

Research through Design

Critical Design and Design Fiction

Design work plays a central role in HCI research. The forward-looking nature of HCI is readily apparent when research strives to design the products of the future. Beyond this, HCI scholars often practise research *through* design, creating design artefacts in order to observe how people use and respond to them. Freed from the constraints of profitability, this type of research allows for exploration and critique of existing and proposed products.

These *critical design* artefacts often reside in a grey area between the believable and parody, divorcing familiar tropes and displaying them in new light. Critical design asks provocative questions about the values underpinning our lives in ways that can only be achieved through discomfort; there is clearly no market potential for a robotic armpit [146], but the grotesque combination of silicon and pheromones highlights the absurdities and dangers presented by anthropomorphism. *Speculative design* and *design fiction* are subsets of critical design focusing on the creation of physical artefacts and fictional products/worlds respectively.

Sometimes the intention of a critical design artefact is to present a counterfactual, asking what kind of society would create such a product. In this way it can be provocative, presenting the viewer with a dilemma as to the seriousness of an object which satirises the contemporary world; by creating these objects, designers look to create images of the future which can act as catalysts for debate about what people really want (rather than suggesting that particular futures themselves are desirable) [100].

Unlike with speculative design artefacts, which can exploit their nature as products to engage us in societal questions where the answer has a direct effect on our lives, in design fictions the viewer is invited to take the inspiration given to them by the designer and extrapolate from that what they believe a scenario would be like if it were to exist. More closely related to conventional media, design fiction "works in the space between the arrogance of science fact, and the seriously playful imaginary of science fiction" [42]

Within HCI speculative design and design fictions have been used to explore futures with personified technology. Smart homes that socialise with inhabitants, designed to be "extroverted and cheerfull", push at boundaries already present, but invisible, in current home devices [237]. Inspired by prior work by the Alt.chi community reflecting on the philosophical and cultural assumptions embedded in popular understandings of artificial intelligence [49] and treatment of techno-spiritual research in HCI [54], this section uses design fiction to explore alternative interpretations of how we might live with autonomous assistants.

Contemporary work, such as Dunne and Raby's *Technological Dream Series: No. 1, Robots* explore this in a modern setting [99]. Robot 4 is extremely smart, but is trapped in an underdeveloped and helpless body—a critical reflection on how neediness is used to engender a sense of control over smart and sophisticated technology. But if Asimov's robotic slaves were designed to take something fundamentally human—subjugation—and turn it into something robotic, modern smart products (including virtual assistants) appear to be performing the same process in reverse. Assistants such as Alexa take advancements in computation and turn them into something distinctly human-like.

To explore this in the context of respect, the thesis presents a work of design fiction in Chapter 8, speculating on the darker side of respectful devices. This allows for engagement with devices that are disrespectful in a way that would be problematic in empirical research; it is obviously acceptable to trial respectful devices with human participants (e.g. as in Chapter 5), but experiments on the effects of harmful technologies are fraught with ethical issues (e.g. long term surveillance in [270]). It also allows for the presentation of the 'platonic ideal' of a device that literally has a mind of its own in a way that even a Wizard of Oz experimental setup might struggle to replicate.

Design & Technology Probes

Utilised extensively in HCI research since the turn of the century, the design probe (and its close relative the technology probe [161]) is an exploratory method used to gain insight into contexts that might otherwise be difficult to observe. Design probes can broadly be described as a collection of objects and tasks which participants must manipulate in some way. Their main focus is user participation—by taking an active part in the design process users are prompted to reflect on their habits and behaviours. This also allows for probes to address issues that are more personal or intimate in nature, with emphasis on user participation allowing probes to better capture the individual contexts of users, including cultural and social aspects [232].

The increase in popularity of probes in HCI has been marked, and has led to many eclectic variations on the original theme. Many different practices surrounding the use of design probes in HCI have emerged since their introduction in 1999 [130]. Projects using probes often have wildly divergent ideas about what aspects of the original methodology are essential, and what can be altered [44]. In addition to this, the place of design probes within the overall argument of a work can vary from the cornerstone to an incidental piece of the puzzle; include the fact that some contributions consider any study or survey that contains open ended questions to constitute a probe and the lines become increasingly blurred [44].

Chapter 5 used a technology probe to test the efficacy of a range of privacy interventions. The probe formed the centrepiece of the experiment, and was

installed in participants' homes to be used on a daily basis. The focus, as with the original probe studies, was for participants to engage with the probe in a way that allowed for them to deepen their understandings of the activities that take place in their daily lives. In this way it echoed the three goals of the original technology probes [161]: the social science goal of capturing how people develop and enact privacy preferences in the rich context of the home; the design goal of exploring innovative ways of empowering people to take control of their privacy at home; and the engineering goal of field-testing prototype hardware and software. This approach was chosen over alternative study and probe configurations due to the degree of interactivity it allowed us to have with participants as they learned and discovered more about their home networks. Delivering insights and teaching through the probe allowed for a much more integrated experience with respect to the three design goals, as well as for repeated contact with the visualisations and curriculum material that would not have been possible with a more light-touch or researcher-led approach.

Understanding Human-Computer Trust

When exploring ethical issues arising around smart devices, it is frequently desirable to understand the extent to which people trust in the systems they use. Trust informs perceptions and influences behaviour in various ways, but has underlying components that can make it difficult to meaningfully measure. Trust in computer systems, or human-computer trust (HCT) is an important and widely applicable factor in most types of interactive systems. Having originally been developed from models of interpersonal trust from psychology and supervisory systems [249] (indeed, many models such as [1, 297] still include an individual's innate propensity to trust [109] as an important factor), prior work in HCI and other fields has shown the importance of trust in many situations involving systems exposing users to some level of vulnerability (e.g. online shopping and e-government) [31].

In systems that provide analysis or decision support, HCT is even more vital. In this context, trust can be thought of as a combination of *confidence* in a system combined with the *willingness* to act on its provided recommendations [223]. Abbas et al. further break this down into system reliability, intelligibility, and level of automation amongst others [1]. A recurring theme in the literature is the lack of consensus around the nuances of HCT mechanisms [31, 223].

In HRI, trust is often combined with other factors, such as animacy and anthropomorphism, in order to evaluate user perceptions of prototype robots (as in [29]). A meta-study of trust in HRI highlights robot-related factors as being significantly more impactful on trust than human or environmental factors [150], but Salem et al. further break down the effect of reliability

on robot trust, showing that while frequent errors impact users' *perceptions* of robots, these altered perceptions may not substantially affect behaviour towards them [297].

3.3 Research Ethics

Ensuring that research involving human participants is carried out in an ethical manner is of the utmost importance. Studies on privacy and the treatment of sensitive information can easily involve topics that participants are uncomfortable recalling or disclosing to researchers. Asking about previous bad experiences with privacy, for example, might bring up traumatic memories of doxxing[4], harassment, or abuse.

There are several ways in which these risks can be reduced. Building rapport with participants helps build trust, making it easier for them to decline to talk about uncomfortable topics. Similarly, taking opportunities to remind participants, both in person and in writing, that they are free to choose what they share, or if they share at all, whilst contributing to research studies helps to foster a non-judgemental atmosphere. Finally, decoupling research incentives from data sharing helps to remove any financial pressures that may push participants to share more than they are comfortable with. Where personal stories are shared, care needs to be taken to anonymise them during transcription, removing references to identifying people and places. Special care should be taken if data sets are to be made available alongside publications, as supposedly anonymised data can still identify individuals when paired with other information (e.g. reporting data as collected from "a local hospital" when there is only one in the vicinity of the author's institution).

An element of this process also exists to protect researchers themselves. Naïve use of modern recruitment platforms can lead to researchers effectively meeting strangers online and visiting them at home, a risky approach with the potential for serious consequences. There are a number of strategies that can help to mitigate this, such as arranging meetings in public places, or by arranging regular check-ins with colleagues, with clear protocols for missed communications—likely a call to the police. In the same way that participants should not feel pressured to disclose personal stories, researchers should not feel pressured to undertake fieldwork that they consider overly risky.

A curious case arises when researching privacy *empowerment*. If one is attempting to gather data on the severity and scale of online surveillance undertaken by devices and companies that are not transparent, how else to achieve this but through additional surveillance undertaken by the researcher? While stringent research ethics policies address concerns of *quis custodiet ipsos custodes*,

[4]Having private and/or identifying information about one's self published online.

the level of intrusion can still be difficult to justify. Further still if the resultant discoveries might cause distress to participants; there may be a widespread understanding that Facebook is tracking people, but fewer expect to be told that they can be precisely located due to oversights in the Tindr API. In these situations, what is appropriate will differ according to circumstances. For the study in Chapter 5 we collected detailed information on internet traffic in the home, but balanced this by (1) providing out-of-band methods for participants to access the internet without being surveilled; (2) prompting participants to use the provided redaction tools to remove connection records that they did not want to share, both before and during home visits; and (3) making it clear verbally and in writing exactly what data was going to be collected, how it was going to be used, and what opportunities participants would have to tailor the data analysed as part of the study.

In the interests of transparency, all of the work presented in this thesis involving human participants has been approved by either the Social Sciences and Humanities Inter-divisional Research Ethics Committee or the Computer Science Departmental Research Ethics Committee under the following references:

Study	Reference
Chapter 4	CS_C1A_18_027
Chapter 5	CS_C1A_19_028
Chapter 6	CS_C1A_18_018

3.4 Promoting Replication

Unlike in psychology, many published HCI studies are not accompanied by data sets that can be used to verify claims and scrutinise methodology. Indeed, many studies also lack the materials required in order to replicate experiments, such as survey questions, interview protocols, and source code for software. This lack of transparency effectively precludes replication studies, as well as the comparison of theories and results across systems, cultures, and other demographics.

Data preservation, at least, is now mandated by a large number of funding bodies (e.g. 10 years for EPSRC funded projects). But when data is archived there needs to be guarantees on *how* it will be preserved; if data is "owned" by a researcher, what happens to that data if they leave the institution? Does the organisation storing the data have the funding and capacity to guarantee that it will be available 10 or 100 years in the future? How do these practices scale when the data is measured in terabytes or petabytes?

In line with the philosophy of transparent and reproducible research, as much of the data collected for the research presented in this thesis as possible has been made available for scrutiny and reuse by others. This is archived by the Center for Open Science on the Open Science Framework (OSF), which has a data management plan to ensure the availability of data stored on the platform for at least 50 years. Growth of the Center's preservation fund is also a major long term priority. Data from this thesis is preserved in the following OSF repositories:

Data	Repository
Chapter 4	`https://osf.io/c3aes`
Chapter 5	`https://osf.io/6j8hc`
Chapter 6	`https://osf.io/53q6j`

In order to carry out the study described in Chapter 5, bespoke research software was required in order to combine data capture, an educational curriculum, and firewall capabilities. This software was developed by myself, and to allow for reuse, extension, and replication the source code is available under the GNU General Public License[5]. This is hosted on Github at `github.com/OxfordHCC/Aretha`, allowing other researchers to download, copy, and contribute back to the code base.

[5] `https://www.gnu.org/licenses/gpl-3.0.en.html`

Chapter 4

The Relationship Between Smartness and Ethical Concerns in the Home

""With your permission you give us more information about you, about your friends, and we can improve the quality of our searches. We don't need you to type at all. We know where you are. We know where you've been. We can more or less now what you're thinking about."

— Eric Schmidt, 2010 [1]

Contents

[1] https://www.theatlantic.com/technology/archive/2010/10/googles-ceo-the-laws-are-written-by-lobbyists/63908

4.1 Motivation

The notion of a 'smart home' augmented by connected digital devices has existed for decades in the popular imagination and is gradually becoming a reality. Devices such as thermostats that learn your preferred temperature, and speakers you can interact with through conversation, are designed to wrap up 'smartness' in a familiar dumb packaging. Given the growing body of research about smart devices and the fact that many people are at least generally aware of one or more smart home devices, one might think it would be obvious what qualifies them to bear the smart label.

But, as shown in Chapter 2 this is not the case. While early research into smart devices tended to focus on the broad potential for smart devices to "imbue homes with intelligence" [322], many more recent contributions partition the space into overlapping but distinct definitions of *autonomous systems* [259], *home assistants* [124], or more narrow definitions describing single types of device; definitions of smart home devices that go beyond 'internet connected' [365] or providing 'smart home service functions' [183] are few and far between.

At the same time, there are very real ethical concerns about the privacy, autonomy, transparency, and social effects of these devices once they are integrated into everyday life. The home as an environment contains finely balanced social equilibria that technology is apt to disrupt. Power dynamics within familial units and between cohabitants are altered by technology that acts as a gatekeeper of information, and for many, the home is the most private space in their lives, a place where they do not expect to be surveilled.

These concerns are often cited at a low level in relation to specific products or deployment settings, or at a high level, considering the ethical implications of smart homes in general. Neither approach adequately considers the broader features shared by multiple devices from a variety of categories and contexts, and while useful for uncovering the implications of smart device usage *in situ*, or ethical concerns writ large, may be of less use in understanding how

A version of this chapter has previously been published as: "William Seymour, Reuben Binns, Petr Slovak, Max Van Kleek, and Nigel Shadbolt. 2020. *Strangers in the Room: Unpacking Perceptions of 'Smartness' and Related Ethical Concerns in the Home.* In Proceedings of the 2020 ACM Designing Interactive Systems Conference (DIS '20). Association for Computing Machinery, New York, NY, USA, 841–854. DOI: `https://doi.org/10.1145/3357236.3395501`"

different constituent elements of smartness, such as automation, or remote control, operate across different contexts to generate ethical concerns.

Instead, this chapter aims to explore an intermediate level of analysis, answering the following research questions:

RQ1a What do people perceive 'smartness' as?

RQ1b How do perceptions of 'smartness' relate to ethical concerns about smart devices?

RQ1c How do these concerns operate across devices and contexts?

And in doing so it makes the following contributions:

1. Presenting user formulations of smartness: how non-experts consider smartness in the context of smart home devices, and how these concepts of smartness interact across categories of devices
2. Analysing how these formulations of smartness relate to ethical concerns expressed in the academic literature, as well as how ethical concerns can be caused or ameliorated by features of smart devices and formulations of smartness
3. Demonstrating how these unpacked formulations of smartness can both reveal and bridge ethical concerns between specific contexts, devices, and locations, in a way not possible through highly specific or very broad approaches

This is important in the wider context of the thesis, as better understanding the links between functionality, context, and concerns is vital for building tools that address the shortcomings in contemporary smart devices; through understanding what current devices do wrong we can begin to make them better. In the longer term, the formulations of smartness presented in this chapter help establish a design space of capabilities and user expectations, usable by creators of future smart home systems to guide the design of new 'smart' things.

Between Situated Ethical Concerns and Abstract Ethical Principles

If we consider the works on smart home ethics presented in Chapter 2, they broadly fall into two categories: empirical studies of ethical concerns arising in response to particular smart devices in particular situations, or expository research exploring how long-standing ethical principles might relate to smart homes in general. Any insights gained therefore tend to be either highly

specific (e.g. smart meter agents raise concerns about autonomy [75]), or else highly generic (e.g. smart homes in general threaten principles of accountability [34] or discrimination [341]). Both specificity and generality have their place; specificity has high ecological validity within the context in question, whereas broad ethical principles are useful for informing policy or design initiatives which have to be adaptable to multiple contexts that cannot be specified in advance. However, they are less useful if we want to understand how ethical concerns might be generated from different kinds of smart devices in different contexts.

Taking an intermediate approach allows for an exploration of the human values supported and hindered by different conceptualisations of smartness (see [122]). Normative ethical principles vary by society and culture, and the extent to which they are embedded in devices is determined by designers (modulated by law and regulation). Previous studies have examined user perceptions of specific devices that might be considered smart (e.g. biosensors [238], conversational agents [207], and home automation [259]). Meanwhile, design tools such as KnowCards[2] and Envisioning Cards[3] help designers to make abstract concerns more concrete. To this end, this chapter examines how end-users conceptualise 'smartness' beyond specific devices and contexts, and how particular formulations of smartness relate to the particular ethical concerns that arise through device usage.

4.2 Methodology

In order to investigate the different ways in which users perceived devices to be 'smart' we designed a set of surveys targeted at device owners, followed by interviews with a mixture of smart device users to explore the links between perceptions of smartness and ethical concerns they had about smart home devices. To inform our survey and interview design we examined related papers, purposefully casting a broad net with respect to characterising ethical concerns according to the definition used by Jensen et al. (how an object ought to be designed according to 'ethical or moral codes' [172]).

All parts of the study were approved by the institution's ethics review board, and all participants were compensated at or exceeding the UK Living Wage. Materials from the study, including interview transcripts, are available at `https://osf.io/c3aes`.

Survey: Unpacking Smartness

Even whilst talking about natural intelligence, the idea of "smart" can have a variety of meanings and connotations, making it a prime example of a

[2] `github.com/betteriot/betteriot-knowcards`

[3] `www.envisioningcards.com`

portmanteau word. However, in the context of connected and algorithmically augmented devices the adjective has been applied to refer to a much larger set of meanings. In order to start unpacking the different perceptions of smartness in the context of smart home devices, we conducted an online survey to capture user understandings about what makes 'smart' devices different to non-smart devices. Based on prior literature we chose six different types of smart home devices covering a wide range of functionalities, locations, and usage contexts in the home: light bulbs, speakers (with voice assistants), security cameras, TVs, kitchen appliances, and thermostats.

Prior to the main survey we conducted two pilot surveys that directly focused on smartness in smart home devices. Understandably, when presented with more abstract questions about whether they trusted their smart devices or considered them intelligent, participants experienced difficulty disentangling perceptions of smartness (e.g. smartness as intelligence vs ideological or historical constructions). These responses reflect the fact that cultural representations of smart technology are influenced by a mixture of positivist marketing, science fiction, and lived experiences, rather than discrete narratives. To this end, the final survey questions approached smartness from a number of different angles. For example:

- Q7 and Q8 explored smartness more as a marketing concept
- Q23 considered smartness more as an inherent property of the device
- Q13 and Q21 attempted to probe how conceptions of smart devices withstood participants' practical experiences with them

We also included the 5 perceived intelligence questions from the 'Godspeed' scale [27], designed to measure perceived intelligence in a human-robot interaction context, in an attempt to draw out other ways in which users might see devices as being smart. A summary of the survey questions is given in Table 4.1.

We ran online surveys on the Prolific Academic platform with 20 participants for each device type. Participants owned the device in question, were 18 or over, and were UK residents. Three researchers independently used thematic analysis to generate initial codes for the 120 responses, which were then discussed and combined to produce eight distinct types of smart functionality. These were used to generate the interview protocol.

Interviews: Exploring Ethical Concerns Associated with Dimensions of Smartness

The goal of the interviews was to explore how the different dimensions of smartness uncovered in the survey might relate to users' ethical concerns

Table 4.1: Sample of questions used to generate the six surveys

Q7	Why did you choose your smart [device] over other options? Was it being 'smart' a factor?
Q8	If you were describing your smart [device] to a friend, how would you complete the following sentence: "My smart [device] is like a regular [device], except..."
Q13	If someone you knew was interested in purchasing a [device], would you try and convince them to get a smart [device] over a regular equivalent? What arguments would you use for or against?
Q21	Recall a time when your smart [device] didn't do what you expected. Did this change your view of your smart [device]?
Q23	How do you think your smart [device] could be smarter in 5 years? 15 years?

about smart devices. A simplistic approach to examining how perceptions of smartness relate to ethical concerns might be to present these types of smartness individually, and to ask people to respond with ethical concerns for each. However, since our aim was to examine how types of functionality operated to create ethical concerns across contexts, we wanted participants to be able to use the functionalities as a form of mental scaffolding to reason about devices. Inspired by the use of design cards in HCI to stimulate and structure thinking during the design process, we used the types of smartness from the survey analysis to create *functionality cards* that named each type of smartness and gave a brief description (see Figure 4.1). An additional set of cards depicted the smart home devices used in the survey[4]. Rather than beginning with abstract concerns and asking how they might apply to products, as in the design process, we did the reverse—beginning with concrete devices and functionalities in order to draw out associated ethical concerns.

University mailing lists, the Call for Participants platform, and a pool of existing participants were used to recruit participants who had previous experience with smart devices (either their own or via friends and family). As with the surveys, participants were UK residents aged 18 or over. Interviews took place at the University of Oxford and lasted 30 to 45 minutes.

At the beginning of the interview, the six device cards were arranged on the table in a random order, and the eight functionality cards were placed face down to the side. In order to establish a baseline for each participant's experience with and understanding of smart devices, we began each interview

[4]For the kitchen appliance category we initially chose a smart oven, but switched to a smart fridge after two interviews ([P01] and [P05]) due to the oven producing some confusion amongst participants.

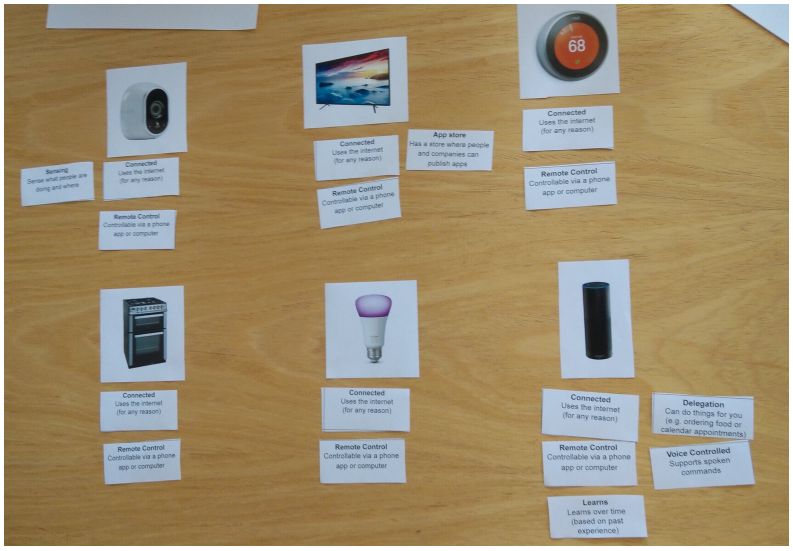

Figure 4.1: Device and functionality cards after being arranged by [P05].

by asking which smart devices they had used, either currently or at some point in the past. We also asked them to identify the devices and describe what they felt made them smart. Participants were introduced to the functionality cards with a brief explanation and asked to arrange cards next to devices they thought had that capability (see Figure 4.1). Participants were then asked if they would have any concerns about using the device as it was, and whether their concerns would be alleviated by putting it in different locations in the house, or by adding or removing functionality cards.

In order to go beyond participants' immediate personal reactions to these devices, we then reset the functionality cards to a standard configuration and asked them to respond to a set of vignettes depicting smart home contexts. Vignettes are sketches of fictional scenarios, used as a research tool to collect situated data on values, beliefs, and norms, common in sociology and more recently applied to HCI [77, 169]. In an in-depth interview, the respondent is invited to imagine the scenario, drawing on their own experience. In utilising fictional details, this method is similar to speculative design approaches, which have been used to study views on the ethical implications of technology use [203, 213, 335]. Relatedly, Nilsson et al. used breaching experiments with scenario-based design and contra-vision to draw out problems with home automation by creating "provocative views on the home of the future" [259].

Our vignettes were drawn from papers on ethical concerns in smart homes. For each of our vignettes, we asked participants how they thought smart devices might be used by the individual characters in the vignette, which device(s) they imagined being most concerned about and why, and whether

adding or removing functionality from those devices would alleviate those concerns (represented by placing or removing functionality cards on the table). The four vignettes were as follows:

V1 *Parents and teenagers:* A household in which the parents used the smart devices to monitor what their teenage children were doing, adapted from [64, 81, 330].

V2 *Revenue maximisation:* A home where the manufacturers of each device attempted to generate as many revenue streams as possible, adapted from [53, 273, 293].

V3 *Vulnerable people:* The installation of the devices into an environment where vulnerable people were living, such as a care home for older people, or an assisted living facility, adapted from [91, 235, 285, 286, 313, 352].

V4 *Rental accommodation:* A rental accommodation where each of the devices had been installed by either another tenant or the landlord, adapted from [233, 270, 365].

Interviews were recorded and transcribed[5], before being coded according to references to devices and types of smartness from the device and functionality cards. Photos were taken of each participant's arrangement of cards and included for reference in the analysis. An iterative thematic coding process was used to identify common themes relating to dimensions of ethical concerns. Three researchers independently coded disjoint subsets of the transcriptions, before convening to consolidate themes and derive a set of joint codes addressing the varieties of ethical concerns. These were re-applied to the data and are presented in the next section organised by types of smartness. The categories of ethical concerns themselves are elaborated on in the discussion.

4.3 Results

Survey: What Does Smartness Consist of?

The 120 survey respondents had an average age of 34 ($\sigma = 11$); 81 identified as female, and 31 as male. Six declined to provide their age and/or gender identity.

As expected, different conceptions of smartness were often entangled. For some, merely being marketed as smart meant that their device must be better

[5]The car mechanic example attributed to P11 in the discussion was given after the interviewer believed the participant had finished talking, and is thus recreated from memory (with consent).

Table 4.2: Types of smart functionality identified from the survey results

Functionality	Description
Apps	Can install apps from an app store
Voice Control	Responds to spoken commands
Connected	Connects to the internet
Remote Control	Via an app or computer
Sensing	Knows what's going on
Responsive	Responds to sensing or other stimuli
Learns	Behaviour is based on past interactions
Delegation	Executes tasks on the user's behalf

in some sense, whereas for other there was more concrete functionality (described below). Smartness was also presented more neutrally as a trade-off ("made me realise how much more complicated it can be compared to just a normal tv" [tv16]) or even negatively ("a lot of people have this belief that smart must mean better" [appliance03]).

The survey results made it clear that for certain devices, such as the TV and the speaker, there was an obvious capability or feature that users associated with that device being "smart". Looking across the results from different devices also led us to other ways that respondents consistently described smart functionality (such as being able to control the device via a smartphone). Analysis of the survey results yielded eight distinct features, ranging from more concrete to more abstract, that were considered to contribute to device smartness[6] outlined in Table 4.2.

Apps were the most common smart functionality given for the TV, and the second most commonly mentioned for the smart speaker (after voice control). **Voice control** was the most commonly cited smart element of respondents' speakers, often mentioned in terms of what users could do with voice control, rather than as a feature in and of itself.

Internet connectedness was rarely mentioned explicitly during the descriptive parts of the survey, but in response to Q14, 83% of respondents (99) said they believed that their device was connected to the internet, with a further 6% (7) stating that they were unsure. **Remote control** was a common theme, particularly amongst the more functional devices such as the thermostat, light bulbs, and kitchen appliances.

Sensing was commonly mentioned when describing the functionality of the thermostat and camera in Q8: *"It can sense when you are returning and turn itself*

[6]It is interesting to note that some of these categories partially correspond to attributes used in European Parliament discussions about smart robots [332], and practitioner definitions of smart home devices [183].

Table 4.3: Demographics of interview participants

P#	Gender Id.	Age	P#	Gender Id.	Age
P01	F	18-24	P09	M	25-34
P02	M	35-44	P10	M	25-34
P03	M	25-44	P11	M	45-54
P04	M	18-24	P12	F	35-44
P05	F	35-44	P13	M	25-34
P06	F	35-44	P14	F	25-34
P07	M	55-64	P15	M	18-24
P08	M	25-34			

on'' [thermostat10]. **Responding** to stimuli came across from responses about cameras, but was also mentioned frequently in the context of smart speakers (*"It responds well and is able to complete many different tasks"* [speaker12]).

Delegation was a common theme with smart speakers and thermostats, with users mentioning being able to create reminders and purchase products through the device, as well as thermostats being better at controlling the heating than they were. **Learning** was a popular response to Q23, with suggestions about learning preferred TV programs, learning about people in the house, and even becoming more human-like.

Interviews: Connections Between Types of Smartness and Ethical Concerns

A total of 15 participants were interviewed, ranging from those who had only used others' smart home devices (P15) to those who had all 6 example devices and more (P09). The gender identities and ages of the participants are shown in Table 4.3.

Delegation

Among the kinds of tasks and activities for which delegation was seen as most welcome—that is, for which participants expressed the fewest reservations–were those that could be described as "monitoring activities". Delegation in this case was seen as an opportunity to 'keep track' of things so that people wouldn't have to. For kitchen appliances, such activities were often safety-related, such as smart ovens detecting or preventing fires. These activities were also discussed for other devices specifically designed for safety or security, such as cameras, including those which facilitated the remote monitoring of elderly relatives. In this way, delegation was seen to enable people to continue living independently who might otherwise need more full-blown paternalistic help:

> I've got a friend whose mum has got early onset dementia and she's got [...] sensors in the door. So one day, her mum went to bed and her phone pinged at 11 o'clock saying did you know your mum's left her door open, which is brilliant because she can get her next door neighbour to go and shut it. [P02]
>
> The more automated things are then the more it could be useful for people who don't have a capacity to understand time. [P12]

Delegating monitoring activities was seen as a means of enabling people to live more efficiently—by simply making it possible for those who didn't have the time or effort to pay attention to their energy usage:

> [A smart thermostat] could be very useful because I know some people they just don't care—they just set 20° every day, they waste so much money and energy without even realising. [P05]

For other contexts, however, participants expressed reservations toward delegation. For some this was down to a puritanical notion that delegation could encourage laziness: *"if everything gets done for you [...] you don't really make any effort. I think I like some effort"* [P05]. More commonly, participants expressed hesitation regarding delegated tasks that had potential value-laden consequences in the real world. In such cases, there was a concern that devices might make bad choices resulting in loss: *"it could be a surprise if you find loads of rotting food sitting at your doorstep"* [P07]. Communicating or messaging others was also seen to be complex and fraught with risks: *"it's a bit weird that something has so much power as calling someone or sending a message to someone–"* [P05]. However, when the parameters for these these actions were well-specifiable and the mechanisms were perceived as reliable, people were less concerned:

> I would find it very useful if tasks that are pretty simple to understand can be [delegated], because then I know these things are on autopilot, almost like a direct debit, so I know it's being dealt with [...] in an OK way. [P11]

Learning

On the one hand, participants were happy, even expectant, that devices learn in certain contexts. These usually related to: *direct learning*—comprising learning by example (such as how to heat a house based on daily adjustments); via explicit instruction (such as mood lighting); or in cases of reminders and recommendation (e.g. *"you'd want your TV to learn what you like to watch"* [P02]).

But participants expressed being less comfortable when the learning related to information that they had not intentionally given the device, or was about things that were distinctly personal, such as a person's thoughts, activities, or preferences: *"I have reservations about technology learning how we think and how we act"*. [P02]

Some felt that learning preferences was akin to passing judgement (*"I wouldn't want my fridge to judge what I was eating"* [P13]), while some participants made a distinction between learning from the individual's own data in isolation as opposed to with data from others:

> The huge power of smart devices is that they don't just learn from me, but they learn from millions of people—that's a bit freaky. [P07]

Participants expressed further reservations for devices learning things they initially did not expect the device was capable of, and discovering things learnt caused their perceptions about it to change from positive to creepy or invasive. In many cases, this was because participants saw this kind of extended learning going too far: *"it's like it's got a mind of its own, that weirds me out"* [P01].

As with delegation, the ability for devices to learn was often contrasted against ideals of control, but also of transparency, with participant's lack of understanding about *how* and *what* devices were learning leading them to cease feeling in control of them: *"You might not want it to be learning. You might want to be in more control of it"* [P04].

In the context of delegated responsibilities, learning was also seen as problematic because it could impinge on a person's autonomy–the ability to change and control one's shopping for instance. It was also seen that getting the system to anticipate such changes would in some sense cross the line:

> Just because you bought it last weekend doesn't mean you want to buy it this week, you might change [what you want to eat]. Either it's going to order things that are not right, or it's going to know far more about me than you would like it to. [P07]

Connectedness

Devices being connected and remote controllable was described as being core to their smartness (in the words of P13, *"that's part of what they are"*), and often one of the key features that drove people to purchase and use them. In addition to being necessary for the delivery of new features, participants also saw it as required for updating devices and any apps that ran on them.

Participants clearly understood that being able to connect out would also mean that others might be able to connect in, making connectedness an obvious locus for concerns about privacy, with all participants describing fears about data being stolen or hackers compromising and controlling devices such as cameras. Often, when asked which functionality cards they would remove to alleviate these concerns, participants would start by touching the connected card before pausing as they realised that doing so would remove a key part of what made the device desirable:

> So if we remove the internet connection it's fine. But then, that's the smartness right, that's the smartness gone. [P11]

In this way, connectedness was seen as an all or nothing concept—a question of *how much* data to send to remote parties, rather than *how many* remote parties to send data to—where removing it would cause devices to cease being smart. Thus the more connected a device was seen to be (i.e. the more data participants believed it was sending over the internet), the greater their concerns about it were. But in other cases, participants suggested that devices might be able to communicate only *within* the home, forgoing the ability to operate them remotely in order to alleviate concerns stemming from it being connected to the internet:

> The camera, I think that if it wasn't connected to the internet that it would be good [...] because traditionally people had closed circuit televisions—that was the point of it, that it's closed circuit and no one else can get in. [P02]

Voice Control

Participants' concerns over voice control manifested primarily as issues around privacy, in particular perceptions of such devices as always listening and attending to what was spoken. This was perceived as unsettling, not only because eavesdropping was human-like behaviour they were not accustomed to devices having, but simply by creating a constant feeling of *presence*:

> Because Alexa is something that you talk to, it feels like it occupies a room [...] it's not the fact of having the microphone, it's the fact of having the entity that speaks back to you. [P13]

Voice control was also seen as problematic due to a lack of awareness about what exactly *what* such devices were doing or capable of doing. Such 'fear of the unknown' resulted in some assuming that the extent of the data collection

via voice control outstripped their understanding: *"You don't know what else it's recording ... they get some useful marketing information, much more than you're even aware of really"* [P07].

Voice control was discussed most positively in the *vulnerable users* scenario, in terms of the value it brought to individuals with special needs. For instance P01's autistic cousin who had too many questions for his mother to cope with, whereas *"if he's asking a machine, it will just keep answering his questions."* [P01]. Similarly, participants reflected that their own concerns about voice assistants might be offset by their utility if they became less able: *"Forty years from now, I might have difficulty doing things and appreciate something voice controlled [...] but right now I don't"* [P06].

Sensing & Responsiveness

Participants often discussed *sensing* and *responsiveness* in conjunction; for the purposes of the results, we therefore report both here. Sensing capabilities were seen as a risk if they might inadvertently compromise the privacy of others in their environments, including their own family members—*"I wouldn't really want to know what my wife was up to"* [P02]—and that doing so could undermine relationships, demonstrating *"a lack of trust between the parents and the children [...] you've gotta have that kind of trust, otherwise they're not going to respect you."* [P02].

Privacy risks associated with sensing seemed to be strongly related to their degree of fidelity, or at least the extent to which information about people could be unambiguously determined. On one hand, sensors with inherently limited capabilities such as motion detectors and occupancy sensors, often associated with smart lighting or thermostats, gave rise to the fewest concerns. At the other extreme were sensors with sophisticated inference capabilities, such as smart security cameras with the potential of facial recognition capacities; P07 argued that while it would be acceptable for the smart camera to sense the presence of a person, *"if it can sense* who *it is, that would be a bit freaky"* [P07].

Remote Control

The convenience of being able to control smart devices remotely was seen as a key benefit of their smartness, especially for home thermostats, lighting and security cameras. But it was also seen as an opportunity for mischief:

> Sometimes on holidays, my boyfriend stays home and I connect to the camera. I say 'Hello!' and I scare him. [P05]

Others described using remote control to subvert traditional means of monitoring by family members (particularly in the context of V1), for instance P01

imagined remotely switching off her home lights to deceive her parents into thinking she had already arrived home whilst still out at night.

But this was often accompanied by the realisation that if *they* could control the device, then perhaps others could too: *"with remote control it's not always a given that you're the person who's in control"* [P02]. In addition to harmless mischief, remote control might tempt landlords or overbearing family to interfere in users' lives against their wishes (V4).

Apps

App stores were often seen as enabling third party developers to unlock the promised potential behind smart devices, since *"they have many many more ideas about what can be done with a device than the company that makes the device"* [P11].

While this was seen as essential—*"ecosystems like that must exist"* [P13]—the introduction of third parties via app stores was identified as a potential source of concern, by opening up the surveillance capabilities of devices to actors beyond the manufacturer. Some participants worried that this additional functionality might itself enable other types of privacy violations, for instance if a *"rogue app that you could install [on Amazon Echo] to listen and take the microphone and store what's being said."* [P11]. For these reasons, some suggested removing the app store to ameliorate privacy concerns: *"the app store, if you get rid of that, then big companies won't be able to track all of your habits"* [P04].

Moreover, in response to the *revenue maximisation* scenario, participants described a conflict of interest on the part of manufacturers, who would be incentivised to compromise the user experience in an attempt to make money, *"encouraging people to buy more things on the apps store"* [P12].

4.4 Discussion

The previous section identified ethical concerns associated with key forms of 'smartness' across different devices through vignettes that grounded their use in particular contexts (summarised in Table 4.4). It showed how the ethical concerns of participants around smartness were often centred around the control of knowledge, either for themselves or their devices, and how all behaviours became increasingly uncomfortable as they drifted towards those characterised as human. Another key dimension to smartness was that being 'smart' represented devices (and by extension their users) being a part of something bigger, for better or worse. Here we delve into discussion of the ethical concerns themselves, examining the extent to which they persist, vary, and operate among the constituent forms of 'smartness' and contexts, and how they relate to previous conceptualisations from the literature.

Table 4.4: Connection between types of smartness and ethical concerns

Smartness	Ethical Concern
Delegation	Autonomy, Transparency, Uncanny Behaviour, Conflicts of Interest
Learning	Privacy, Transparency, Uncanny Behaviour
Connectedness	Privacy
Voice Control	Privacy, Autonomy, Uncanny Behaviour, Conflicts of Interest
Remote Control	Privacy, Autonomy, Social Order
Sensing & Responding	Privacy, Autonomy, Transparency, Social Order
Apps	Privacy

Privacy

Privacy was the most prevalent ethical concern, appearing across all types of smartness except delegation. This is rather unsurprising; previous research into smart devices in a variety of contexts has almost invariably brought up privacy (of the 35 papers used to design the vignettes, all mentioned privacy in some form). However, the manner in which privacy arose, the kind(s) of privacy referred to, and the meaning(s) attached to them subtly differed between the different types of smartness.

While *connectedness* had clear privacy implications associated with the possibility of unwanted data exposure, *voice control* raised privacy issues at a more phenomenological level; it was not necessarily just the fact that Alexa could facilitate privacy-violating data flows, but that having a device listening out for spoken commands induced the feeling of being watched by an "invisible person" in the room (P14). Through distinctions like this, we can see how different dimensions of smartness may impinge upon privacy at different levels, from *facts* about whether or not certain data is flowing, to *phenomenology* (how it feels), to *preferences* or *norms* [268].

In articulating how *app stores* and *remote control* might threaten privacy, references were made to the ways in which social contexts may dictate the appropriateness of certain information flows and incursions into private spaces [261]. P05 felt using the remote control on her home webcam to surprise her boyfriend while he was home alone was acceptable within the contextual norms of their relationship, but agreed it would be a violation for a landlord to do the same to a tenant. In a different way, as P04 cautioned, app stores open up the possibility of data sharing relationships with unknown third party actors beyond the manufacturer; like with smartphone app ecosystems, these third parties may come to be seen as social actors potentially violating

contextual integrity through opaque leakage of personal information [337, 336].

Autonomy & Control

As outlined in the results, delegation, voice control, and sensing/responding gave rise to discussions pertaining to *autonomy*, understood as the ability to make one's own choices without external influence, or 'self-governance' [70]. For instance, P07's feeling that delegating tasks to Alexa might restrict his freedom to choose what to eat for dinner or watch on TV, or P05's puritanical concern that too much delegation would eliminate meaningful effort. However, the effects of devices on autonomy were not always understood to be undermining. In some circumstances, as P12 remarked, the ability to delegate to a device or rely on its capacity to sense and respond, could *support* autonomy by enabling people to continue living independently.

This apparent contradiction—that features like delegation can both undermine and support autonomy—might be explained by reference to subtly different notions of autonomy they appeal to. P07's objection to having his meal choices decided for him reflects autonomy as 'negative liberty', or freedom from interference, while P12's example that delegation could support cognitively impaired people to live independently might reflect 'positive liberty', understood as the ability to take control of one's own life [33].

Participants also suffered from a lack of options when exercising control over *how* they used smart home devices. While the main ways that participants talked about resolving tensions around device functionality appeared to mirror exaggerated smartphone tropes of (dis)integration [151], for the devices considered in the study these categories really were total, reflecting the all or nothing aspect of device functionality; participants (often implicitly) saw their options as being to 'give in', integrate the device, and accept all of the device's features (desirable *and* undesirable), or to dis-integrate, discard the device, and benefit from none of them.

Transparency and Accountability

Similarly to [259], the issues around transparency and accountability that participants described in the interviews ranged from simple problems of visibility (e.g. what a device was sensing or recording), to more complex questions about why a device might take a particular action (e.g. as a result of learning, or suspected conflicts of interest between the user and the manufacturer).

Voice control and connectivity offered clear examples of the former, with participants frequently expressing confusion over when speakers were recording, or what information they stored about interactions. This echoed prior findings around devices poorly conveying the limits of data collection [224,

236], and the problems of poor mental models about smart devices [365]. A more detailed analysis showed that users often had good mental models about what devices were *capable* of capturing but little knowledge of what they were *actually* capturing (e.g. The Echo *could* record everything, but is it?), and it was from this that transparency concerns arose, rather than a lack of understanding about capability.

Of the more complex questions, while participants did not expect to understand the learning process, the repeated contrasting of learning with control did suggest that the lack of affordances about learning prevented them from building good mental models, giving rise to the feeling of no longer being in control. For P02, who questioned the appearance of adverts after a conversation in the vicinity of a voice assistant, a simple concern about transparency was elevated to a more complex concern about the lengths manufacturers were willing to go to in order to turn a profit. In some cases, it appeared that users had attempted to deal with these transparency problems by being much more sceptical about devices and manufacturers. Examples included P07's suggestion that companies are collecting 'far more than you know' and P04's belief that companies would automatically be tracking users, a type of dejected acceptance previously seen in this context [304].

Conflicts of Interest

Some types of smartness were connected to potential commercial conflicts of interest that devalued devices' perceived usefulness. Some of these featured quite familiar and prosaic forms of influence, such as adverts on their fridge LCD display, or a standby screen on the TV: *"It would be [...] akin to putting up a corporate poster in your kitchen."* [P02]. More complex conflicts of interest arose in relation to voice assistants. The fact that they made it easier for users to purchase from one vendor than others was seen by some as going directly against their will, and reduced the usefulness of the device. This is a typical example of what economists term the agency problem [147], a common theme in the literature around smart meters and smart energy grids [75, 143, 293]. Along these lines, P11 made an analogy to how customers of his car maintenance business would normally go to their preferred mechanic rather than other options that were cheaper or otherwise more convenient, and other participants also saw the potential for becoming trapped in an ecosystem.

There was also an alternative interpretation, which framed this concern in the context of *competition*. Participants suggested that offering reduced options when delegating was acceptable, so long as the provider of a device did not abuse their dominant position in the market: *"A lot depends on if it's the same company that sells you the device, is also the same company that's providing services behind the device"* [P07]. This was mirrored in a more general concern that once users had purchased one device, manufacturers would use their foothold in

the home to sell users other devices. Examples included voice assistants for other rooms, or additional thermostats that would only work with the system the user had originally purchased. In these cases, participants worried about their ability to purchase whichever device they liked in the future would be curtailed by devices being inoperable with each other.

Social Order

As in many previous studies of multi-user device interaction in smart homes (e.g. [237, 283, 365]), a prominent set of concerns pertained to devices disrupting or otherwise failing to respect the natural social order(s) of the home. Much as in Zeng et al.'s study [365], one major class of these concerns dealt with power and control—both within/among inhabitants, as well as relating to actors external to such spaces. In particular, there was a discussion of the potential hazards of the privileges that *owners as administrators* of smart devices (including landlords) might have over inhabitant-users. This was particularly the case where devices conferred asymmetric power or information access and remote control to owners, including sensing/listening capabilities.

From the perspective of the owner with such privileges, respecting the 'politics of control' meant devising ways to negotiate or self-regulate the use of devices to prevent accidental breaches of the trust, as in [283]. In doing so, participants discussed having to take deliberate action to *avoid* spying on, or otherwise violating the expectations that family members or those subject to surveillance had about them. In some cases, such efforts were made difficult or impossible by devices not adequately supporting appropriate access restrictions, such as smart TVs failing to differentiate among multiple owners, or smart speakers like Alexa keeping unified audit logs of all users' interactions.

A special class were caring relationships, discussed through our vignettes focused on parents and their children (V1), as well as with elderly relatives (V3). In such relationships, concerns centred around achieving a balance between devices enhancing the safety of the cared for, and risks of undermining their autonomy and privacy in the process. In the parents and teenagers scenario, it was widely agreed among participants that potential harms arising from the use of such devices included them being used in ways that were overly controlling of their children, as P02 described.

This was in contrast to responses to the elderly relatives vignette, in which participants were most willing to use the various kinds of smartness to support relatives in re-gaining self-sufficiency. Here, potential ethical concerns regarding relatives' autonomy were seen to be outweighed by the advantages these devices afforded. These two vignettes also highlighted the differences in perspective between our participants, and how trying to adopt a one-size-fits-all approach to smartness-related ethical concerns might lead to the marginalisation of certain voices. Whereas the other vignettes saw more focused themes

emerge, with participants drawing on shared cultural backgrounds and public discourse, social order concerns were far more likely to be related to highly personal experiences, creating a diverse patchwork of justifications informing what was deemed acceptable.

The Uncanny Valley of Social Smart Things

On the one hand, participants referred to some devices as if they were social entities [256], as observed in several prior studies [237, 288]. Examples of this include P04 worrying about being bossy to Alexa, or P14 describing it as an invisible person in the room. But on the other hand, tasks such as learning and delegation were described as unsettling because they were *too human*, making a device appear to have a *mind of its own*. In this way, devices sometimes fell into an uncanny valley, where they were perceived as social enough to trigger social behaviours in users, but still machine-like enough to create dissonance [248]. These uncanny valley responses also relate to the purification behaviours described in [204], where the blurring of human and machine causes users to alter their internal categories to eliminate overlaps or grey areas.

Because these sentiments were linked to normative values about the roles that machines should or should not have in people's lives, it is possible that these reactions will diminish over time, as P02 chillingly remarked with advertising on the door of the smart fridge: *"maybe it's just something that we need to get used to. Something that society will just grow to accept"*. The way that voice control was described as lazy mirrors the way in which many other technologies that are now integrated with everyday life began as curiosities; social use of the telephone was originally characterised as 'frivolous and unnecessary' [272].

Limitations

While it did not prevent major issues and themes arising in the interviews, the exploratory nature of the study meant our participants constituted a relatively small set of independent adults. This narrower set of views risks erasing accounts of smartness from marginalised groups, as evidenced by the willingness of participants to justify the use of smart devices for older adults with little consideration for their concerns. Additionally, while we attempted to mitigate this with vignettes that were applicable to a wide span of social situations and lifestyles, there are countless devices and use cases beyond those of which our participants had experienced.

4.5 Conclusion

The rise of 'smartness' as a buzzword used to describe devices for the home comes from a rich tapestry of influences. From the promises of science fiction

stretching back decades, to crafted marketing messages that pitch devices as the perfect solution to problems (both real and manufactured), as well as serving different ideological visions of the home and technology's place within it. Investigations of particular devices in specific contexts of use can provide rich insights into particular user's idiosyncratic concerns *in situ*, and abstract discourse may help ground those concerns in light of longstanding principles. But both approaches are either too specific or too generic; neither enables reflection on how different dimensions encapsulated by the 'smart' moniker relate to ethical concerns across devices and contexts.

The analysis presented in this chapter, exploring the design space of 'smartness' with a mid-level approach, represents the building blocks on which subsequent chapters of this thesis will build. The formulations of smartness given by the study participants help cut through the nebulous use of 'smart' both in the marketing of these devices and in academic research and contextualise it with respect to their relationships as partners, parents, and carers, as well as in relation to their own particular lived experiences. While understandings of smartness are unlikely to have neat boundaries (especially as the concept shifts across cultures and time), by outlining some of these rough edges and concerns we provide a starting point for designers to better anticipate how other devices and contexts might influence ethical concerns about their own (e.g. how general concerns over automation might interact with specific concerns about the deice itself). In this way, unpacking smartness allows for its subsequent *repacking* in order to help "manage the potential 'attack surface' of the digital on everyday life" [78], supporting designers in designing for the values of their users. The next chapter builds on this beginning, focusing on how we might empower people to 're-pack' the smartness in their own devices to better serve their own preferences and needs.

Also of interest were the ways that participants traded off the benefits and drawbacks of using devices that mediated or otherwise supported their relationships with others. There was much more hesitance towards using sensing devices around children than older adults, and mixed opinions on the use of such technology in relationships between adults in the same household. That the design of smart devices enables this behaviour is interesting given the previously referenced observation by Taylor et al. We will build on the understanding of allocation and balancing of care and tenderness through technology developed here in Chapter 7, with the goal of re-examining the results presented in this chapter in the context of philosophical thinking on respect and the design of smart devices.

Chapter 5

Connected Home Privacy and the Development of Privacy Preferences

"And I think the feeling of computer literacy among the populous is the thing that, for me at least, gives me the most comfort—that that centralised intelligence [won't] have the least effect on our lives without us knowing it."

— Steve Jobs, 1981 *ABC News* interview

Contents

5.1 Motivation

A major concern that surfaced in the previous chapter concerned the transparency of connected devices in the home. Connectivity was a unifying aspect of smartness as identified by the participants, but also a unifying concern across devices. Being connected meant that devices could be sending any and all information given to or collected by them over the internet, but were they? These concerns were combined with scepticism over the interests that devices were acting in, where a device's manufacturer might collect data related to their users not to improve the device's functionality, but rather to increase their own profits.

In some cases, as with P14, participants described the uncanny and unsettling feelings created by devices that were always listening. Although the ultimate consequences of intrusion by connected devices into the private spaces of the home have yet to be fully understood, researchers and journalists have documented the ways that devices are seen as creepy [124], or make users feel constantly monitored or spied upon [84, 315]. Recent scandals involving products revealed to have hidden sensors not disclosed to end-users, including microphones in smart mattresses [309] and home hubs [126], have further heightened anxieties around digital devices in the home. Such incidents have caused end-users, consumer rights advocates, and privacy and security researchers to press for restrictions on data collection, and the establishment of mandatory software quality, security standards, and duties of care [168].

Chapter 2 highlighted three key reasons why addressing privacy in the home is challenging. End-users generally lack an awareness of how their devices collect data about them, including the kinds of data collected, and the ways such data are disclosed and used by various first- and third-party entities [133, 336]. Second, even when such data uses are made apparent, individuals lack both breadth and depth of knowledge necessary to form informed preferences about such disclosures. Such knowledge includes both a broad, contextual understanding of data sharing norms, uses, and data protection requirements, on one hand, and specific, detailed knowledge on the other, such as of disclosure risks, or the reputations, business models, and security practices of the companies handling their data [104, 367]. Finally, even when such preferences

A version of this chapter has previously been published as: "William Seymour, Martin J. Kraemer, Reuben Binns, and Max Van Kleek. 2020. *Informing the Design of Privacy-Empowering Tools for the Connected Home*. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20). Association for Computing Machinery, New York, NY, USA. DOI: `https://doi.org/10.1145/3313831.3376264`"

have been formed, users rarely have the ability to take meaningful action to improve their privacy, due to a lack of options or effective means of exerting control [320]. When combined, these factors suggest that residents of today's connected homes may be stuck in a negative cycle of dis-empowerment when faced with the challenge of managing their privacy.

Motivated by these interrelated problems, this chapter presents the results of a 6-week deployment of a *technology probe*, a functional prototype sent "to find out about the unknown [...] to hopefully return with interesting data" [161]. The unknown, in this case, was the design space of *privacy-empowering technologies* for the connected home—future privacy tools that might go beyond static privacy options or labels to support dynamic situational awareness of their informational exposure, and to help end-users build up a rich, conceptual understanding of connected privacy and risks over time. We wanted to explore how such tools might provide targeted, contextualised information relating to their particular situations to maximise relevance, and also to provide the means of taking immediate action, thereby helping users cross both their privacy *gulfs of evaluation* (i.e., understanding their current situation), and *execution* (i.e., taking meaningful action) [264]. Technology probes have been seen as ideal for such initial investigations as they constitute "simple, flexible technologies with three goals: the social science goal of collecting in-context information about the use and the users, the engineering goal of testing the technology, and the design goal of inspiring users and researchers to envision future technologies." [266].

Through our technology probe, *Aretha*, we sought to jointly explore social science questions relating to privacy risks and preference formation, engineering questions relating to the feasibility of deriving understandable models of data disclosure from network traffic flows, and, finally, design questions pertaining to the capability and interaction design spaces of privacy-empowering technologies. In doing so we answer the following research questions:

RQ2a How can privacy-implicated activities of devices be made legible to users?

RQ2b How can we help users interpret and form preferences about data disclosures by their devices?

RQ2c How can we provide privacy controls that enable users to take action about disclosures by their devices?

5.2 Designing the Technology Probe

The background literature discussed in Chapter 2 provides a challenging set of social science, engineering, and design considerations relating to privacy-

empowering technologies in the connected home. These drove us to formulate the following key design goals for such technologies, and in turn, our technology probe:

DG1 *(Legibility)* To make the privacy-implicated activities of devices legible to their users, in the forms of both real-time and historical records of all information flows and their destinations.

DG2 *(Interpretability)* - To help users interpret and form preferences about data disclosure activities by providing them with a rich conceptual understanding in the technology and business models operating behind their connected devices, including disclosure norms, risks, and purposes.

DG3 *(Actionable Choices)* - To provide *in-situ* privacy controls that enable users to directly take meaningful action on information flows to reduce exposure.

We began the process of designing our probe by referring to the existing analysis tools examined in the background material (see Table 2.1). Three broad categories emerged: fingerprinting devices based on the network traffic they produce [35, 239]; behaviour classification using machine learning [3, 267, 326, 366]; and those taking a more holistic approaches rather than a device-specific analysis [92, 337].

While the latter two projects come the closest to the subject of this work, they only address DG1 (*Legibility*), stopping short of providing educational material to increase users' understanding, or providing actionable controls *in-situ* for reducing exposure. This indicates an unexplored design space for privacy-empowering tools that provide real-time legibility, support users to interpret their privacy exposure through better understanding of the technology and business models, and provide active controls. This design space was explored more thoroughly through the design of the Aretha technology probe. Source code for the probe software is freely available under an open source license[1].

DG1: Legibility

Given the lack of transparency around how connected devices in the home collect and disseminate data, we felt that a crucial first step was to provide them with a view of the ground truth about how their devices disclosed data. Aretha was connected to the home router and acted as a WiFi hotspot, capturing packet headers of traffic passing through it (this reduced the amount of collected data and avoided the challenges of inspecting encrypted payloads [290]). A third party service was used to obtain information about the

[1] `https://github.com/OxfordHCC/Aretha`

Figure 5.1: Time series visualisations used in the probe interface.

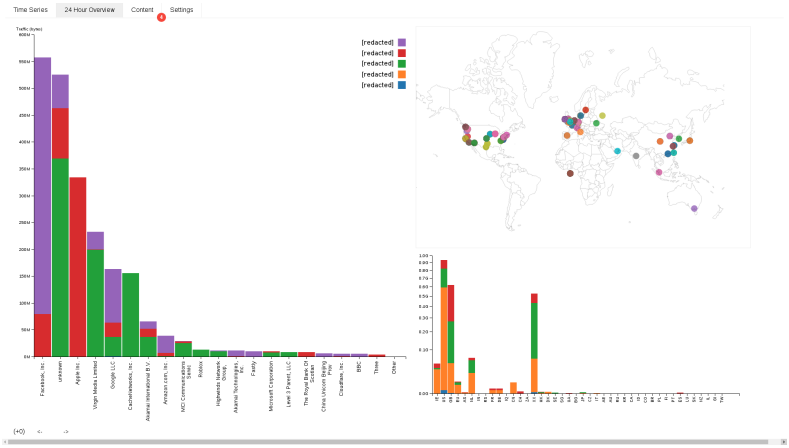

Figure 5.2: Aggregate visualisations used in the probe interface.

owner, location, and threat status of external destinations. Unlike IoT Inspector [155], Aretha included data from other connected devices in the home (e.g. smartphones and tablets) alongside IoT devices.

The interface was based on that of X-ray Refine [337], which had been usability-tested in previous lab studies. We adapted this to present information captured from the network both on a timeline, showing the ebb and flow of data throughout the day, and as an aggregated overview showing the total information exposure of the household (see Figure 1).

DG2: Interpretability

Previous studies have established that end-users often lacked the depth of understanding of networked privacy necessary to take views on risks and preventative actions [177, 357]. Thus, we felt that in order for end-users to be able to interpret and effectively evaluate device disclosure behaviours, Aretha should provide supporting educational and informational material of at least two kinds: essential foundational knowledge about networked privacy concepts and risks, and background information about the specific companies handling their data. To address the first, a short educational curriculum inspired by FoxIT [133] was assembled from content from the BBC, the UK Information Commissioner, and various other sources, with the aim of providing a broad overview of networked privacy, including Internet basics, how and why devices sent data, and information about data breaches, data protection, and kinds of privacy risks. To provide grounding for essential abstract concepts, concrete examples were constructed, when possible, from users' own home disclosure profiles. Such examples included illustrations of which devices within their homes sent encrypted versus unencrypted data, and for what purposes these data were being disclosed. Three iterations of the curriculum were piloted via online surveys (total 14 UK residents aged 18 or over), which let us evaluate the inclusion and presentation of different types of information.

DG3: Actionable Choices

The third goal was to enable users to exercise control by being able to take privacy remediation actions directly within the Aretha display. Towards this end, we faced the challenge of choosing or designing an appropriate method of control; we sought a simple, yet effective method that could serve as a basis for more sophisticated approaches. Since the concept of a firewall was likely to already be familiar and was easily explainable, this seemed a best initial fit. Unlike firewalls for use by experts (which often work at the level of IP or MAC addresses) we designed Aretha's firewall to be user friendly by using device and organisation names in directives (e.g. 'block all traffic between ¡Sam's iPhone¿ and ¡Facebook, Inc¿'). To make such directives easy to specify, a simple drop-down-based graphical interface was created that directly embedded the ability to create directives into Aretha's visualisations.

Pilot Field Test

To identify usability problems and improve the design of our probe prior to our study, we deployed our initial version within a smart home exhibit at a major building research centre in the UK[2]. Seventeen members of the public

[2]The prototype did not contain the firewall tool.

P#	Age	Gender Id.	Education
P1	25-34	M	Postgraduate Degree
P2	35-44	M	A-Level
P3	45-54	F	Bachelor's Degree

P#	Self-Reported Connected Devices in the Home
P1	**Laptop**, **Phone**, **Tablet**, Watch
P2	**Laptop**, **Phone**, *Alexa*, TV, Camera, IR, Light, Socket
P3	**Laptop**, **Phone**, **Alexa**, TV

Table 5.1: Demographic information for the study participants (top) and the types of connected devices they reported owning (bottom). Italics show devices users reported intending to connect to the probe and bold those actually connected according to network traces.

were invited to spend time with the prototype system in the fully-equipped smart home test bed, connecting their mobile phones and viewing the data destination visualisations on-screen. Feedback from the pilot was essential in optimising screen layouts to prioritise the visualisations participants found most useful and improving reliability by identifying software and hardware bugs.

5.3 Methodology

We designed the main study to consist of a six-week technology probe deployment with three families. As the goals of this study were exploratory, we aimed for a longer-term deployment with a smaller, carefully selected set of households. We felt a six week duration would allow sufficient time for participants to acclimatise to having the probe in the home, and for use of the probe to integrate with household rhythms and routines. It was also seen as sufficient time for participants to be exposed to and internalise the educational curriculum which was delivered during the second phase of the study, as well as for members of the household (other than the primary participant) to interact with the probe. All materials from the deployment study can be found at `https://osf.io/6j8hc`. The home deployment was approved by our University's IRB, and participants received £200 in shopping vouchers for taking part in the study.

Participant Recruitment and Selection

Recruitment was done through a two-step process. First, we recruited an initial pool of interested participants using the Call for Participants platform,

reaching out to several smart home hobbyist communities online, and through a mailing list of participants from previous smart home research studies. Interested participants were invited to complete an initial survey with the connected devices they owned, who they shared their living space with, and the geographic area where they lived. Second, we drew on this pool of interested candidates to select three households that represented a variety of technical expertise, family structures, and location of physical residence. Written consent was collected from the main participant in each household, who was required to explain the study to cohabitants and obtain oral consent (we provided written material to help with this). Demographic information on participants is given in Table 5.1.

Study Phases and Structure

At the start of the study, the same researcher visited each household for installation of the probe. The hardware of the probe consisted of an Intel NUC running the Aretha software, an HD display, and a standard keyboard and mouse. The NUC was connected to the participant's broadband router via Ethernet, and was configured to act as a WiFi hotspot to which participants were asked to connect their home devices in lieu of their regular WiFi access point. The interface was always on and continuously visible on the provided display, with users able to toggle between the two screens shown in Figures 5.1 and 5.2.

Before installation of the equipment, participants were interviewed at home for approximately 30 minutes. This semi-structured entry interview was designed to capture participants' relationships with their connected devices, including questions on their conception of what privacy meant to them, what data they thought their devices shared with whom, and the extent to which they felt in control of their data online. While participants were encouraged to connect as many of their devices as possible to the probe, the initial briefing made it clear that this was not obligatory and that they could redact information collected by Aretha at any time. Additionally, some smart devices required an Ethernet connection and thus could not be connected to the probe.

The deployment was split into three stages lasting two weeks each, with participants instructed to interact with the software approximately every other day (see Figure 5.3). In the first stage Aretha functioned only as a passive display, encouraging participants to experiment with it to learn more about their devices and to familiarise them with the research software. The second stage added the delivery of the educational curriculum. The final stage of the deployment gave users the ability to use Aretha to block their devices from communicating with companies of their choosing. Participants were contacted at least once during the study to capture feedback and reactions that might have otherwise been forgotten by the end of the six weeks.

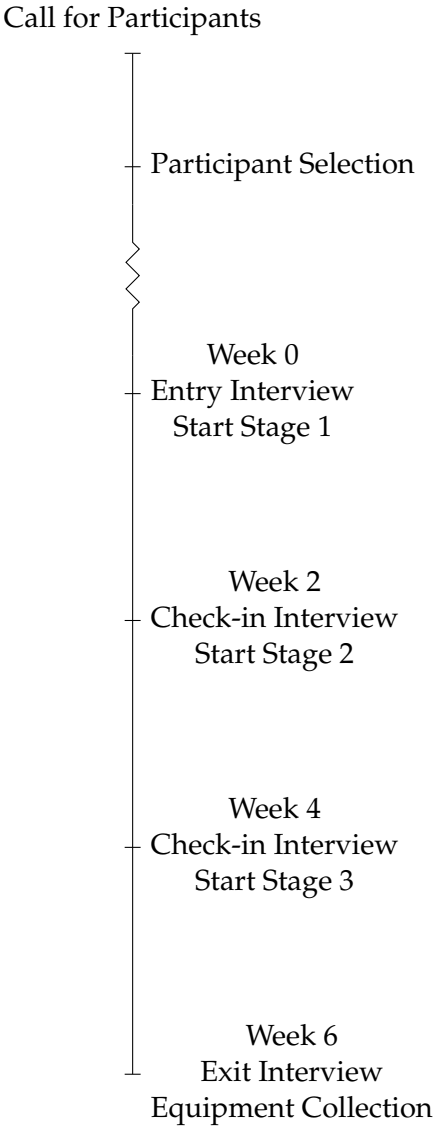

Figure 5.3: Structure of the home deployment

Exit interviews were conducted at the conclusion of the study and lasted approximately 40 minutes. Questions included what participants and their cohabitants reactions to the probe were, whether the participants had consciously changed their usage of any devices as a result of the probe, and what they had found most surprising about the data shown by the visualisations. Questions on perception of privacy and control from the entry interview were repeated, with participants given copies of their earlier responses to reflect on and revise if they wished. Participants were given the opportunity to review and redact information recorded by the software, and device MAC addresses were deleted at the guidance of our IRB.

Analysis

Interviews and check-in sessions were transcribed and anonymised by the researcher who conducted the session. Two researchers independently familiarised themselves with and coded the transcripts following [47], meeting to review and merge themes to generate the final analysis. A preliminary round of coding was also conducted before the exit interviews to allow emergent themes from the entry and check-in interviews to be further interrogated. Network traces collected by Aretha were referenced during the coding process. While these alone proved to be of limited utility in understanding how the probe was used, they were effective in allowing us to triangulate critical events reported by participants.

5.4 Results

All three probes operated continuously during the 6 week period except that of P1, which experienced a network-card failure that made the system unavailable for approximately four days during the first phase of the study before being rectified by the participant. All probes actively recorded network traffic (packet headers) during the study, recording 112 million packets from 19 devices to 428 unique destinations across 41 jurisdictions worldwide. On average, the top three destinations by volume per household represented 74% of the total traffic observed by Aretha.

Thematic Analysis of the interviews yielded five major themes: *changing conceptions of privacy*—how participants' perceptions of what privacy meant to them changed as the study progressed; *personalisation of privacy concerns*—how participants' understandings of what privacy meant to them became more concretely related to their individual routines and practices; *household members' interest in privacy*—how our participant's cohabitants expressed interest in the probe itself as well as the privacy practices of the household; and *everyday practicalities impacting adoption of the probe*—how participant's usage of the probe (particularly the firewall) was affected by routine happenings in

the household. In the following subsections, we briefly outline the journeys taken by each household and describe these themes in the context of each.

Household 1

The first participant lived with his partner in their studio flat, and was technically savvy. P1 was responsible for administering the technology in the household, but online accounts were kept separate between them. The privacy assistant was installed in the main area of their studio apartment, where it was visible throughout the study and used by both the participant and his partner.

Believing that privacy was important to him, P1 had difficulty elaborating further, mentioning tangential practices when asked about the steps he took towards it (such as unsubscribing from email lists and using antivirus software). Going into the study P1 repeatedly voiced concerns that his long term use of the internet meant that he had passed the point of no return with regards to controlling the dissemination of his data online:

"I've been on the internet for so long that it's just not going to happen. Obviously now I can be more defensive and be cautious, but I just think it's out of control now"

As new companies appeared on the visualisations, P1 developed his own information gathering process built around the UK national company registrar. Largely unconcerned about the actual data being collected, it was the *relationships* between the companies shown by Aretha that were of importance:

"So the systematic things I like to do when I research about a company: how many people are involved, what are their sponsors, what's the share price now, how has it changed in the past, what's their next project, what's their previous project, did it change anything, were they successful?"

During the study P1 changed his broadband supplier, and mentioned being curious about how his choice of ISP might impact his data exposure (network traces showed a 275% increase in avg. daily traffic to the ISP). Having the probe display in the main area of the apartment allowed it to trigger conversations between P1 and his partner about their data flows. Having not previously been a part of the administration process, his partner was able to educate and inform herself about what their devices were doing:

"She used to open web pages and then instantly go and check the time series, to just see that Yahoo was open, was there a change?"

In what became a theme across the participants in the study, P1 found the firewall to be an ineffective way of controlling his data flows. This was mainly due to a fear of it interfering with his digital activities, which were "everything", instead leading P1 to seek his own behavioural control mechanisms.

The curriculum content about the data economy led him to reflect on the corporations that were indirectly benefiting from his online activities, prompting him to use the ethical track records, details of ownership, and shared interests uncovered during his research to help align his purchasing choices with his responsibilities around sustainability and ethical consumerism:

"So [the privacy assistant] is doing that by default, and all I'm needing to do from the device is actually to spend my money where everything's fair, everything's fine [...] so I don't want [...] one day delivery from Amazon because they're not paying their people fair. [...] and if I know that when I go on Amazon it's also sending data to X Y and Z, I'm also not going to buy from X Y and Z because that's my grudge against the company now"

In addition to considering the management of his data that was already online as a problem that was too big to handle, P1 expressed scepticism of GDPR-based data rights as an effective method of controlling his exposure—if behavioural models of him still existed just without his name, then what was the difference? P1's Aretha observed 218 distinct companies operating in 37 countries, highlighting the sheer amount of effort required to meaningfully invoke his data rights. He proposed outsourcing the task to an independent organisation that would be able to handle revocation requests en masse:

"*If someone [...] comes and tells me that we're happy to remove everything from the internet that belongs to you and give you a new birth on the internet [...] then I might be interested to start from scratch*"

Household 2

The second participant lived with his partner and five children in a fully equipped smart home. Many aspects of the house were automated, and P2 was entirely responsible for keeping everything working. The privacy assistant was installed in the living room of the house, allowing for passive observation whilst watching TV or carrying out other leisure activities.

The setup of P2's home was by far the most advanced of the three households, with smart devices touching most aspects of family life. This way of living, to which smart devices were integral, not only foregrounded the reliance of the household on connected technology, but also highlighted how much friction was exerted by adding privacy protecting measures to an already complicated setup. P2 partly envisaged Aretha as a way to validate his own previous privacy and security measures on their home network.

P2 began the study with a more developed conception of privacy, describing it as a matter of 'give and take', with him combining the manufacturer's description of what devices were doing with "a little bit of faith" that they would operate as specified. While P2 did keep track of the data destinations

that appeared over the course of the study, his son was also monitoring the display and holding devices (and P2) to account:

"*My son is quite technical and is looking at the main bar graph a lot trying to see which bits of traffic are interesting to him, like just now he asked why we're sending so much data to the weather service, which is a good question!*"

Most surprising was both the amount of data sent to large technology firms, outside of the EU (54% by volume), and the frequency with which data was sent during periods of no or low active device usage. As with P1, P2 found the delayed impact of blocking decisions interfered with the routines and challenges presented by family life:

"*Because you then have to go away and see what changes happen a little bit later on [...] we've got the five kids running around and sometimes I was reticent to click something in case it affected something and I wouldn't get a chance to come back to it*"

This fear came true when blocking access to a specific company removed comments on YouTube videos. He also noted how the sheer number of destinations observed by Aretha made the firewall difficult to use: *"I found that I got a bit lost on occasion"*. But unlike P1's behavioural control mechanism, P2 did not go on to develop their own control mechanisms in place of the firewall. Instead, P2 mentioned having previously investigated a variety of different options and expressed a lack of control over the family's data flows that did not change over the course of the study.

However, by the end of the study, P2 was able to extend his original definition of what privacy meant to him. This centred around the *purpose* of products and behaviours, particularly how companies balanced meeting user needs against fulfilling other business purposes such as generating advertising revenue, a balance he was adamant needed to remain in the favour of the user. Vocabulary centred around proportionality was used to compare the data flows shown by Aretha with his previous expectations, reflecting that to some extent this was a symptom of the highly interconnected nature of the modern world wide web.

In contrast to P1, P2 drew a clear distinction between information sent to the authorities versus information collected by private companies, and highlighted the importance of keeping insights from devices anonymous and aggregated rather than hyper-personalised:

"*It's not so much that they know that three thousand people have watched this particular YouTube video, it's do they know that I have watched this YouTube video?*"

Household 3

The third participant lived with her partner and daughter. While the online accounts associated with devices in the house were registered to P3, their partner was responsible for administering devices. The privacy assistant was installed in the hallway between the kitchen and living room of their house, making it easy to glance at throughout the day.

Similarly to P1, P3 had difficulty describing what privacy meant to her beyond it being important, focusing on a previous bad experience where an organisation had published her personal details without consent. Overall, P3 reported being less engaged with the privacy assistant over the course of the study, often observing new names but being less proactive in investigating them. She found that further research into companies listed as destinations in Aretha was less effective at addressing her concerns. Unlike P1 and P2, P3 was not primarily responsible for the technology in the home, instead deferring to her partner who worked in the IT sector, and usage of the privacy assistant between them also appeared to fall into this pattern.

Of surprise were the varied jurisdictions that data was flowing to, and the relationship between action in the house and what was detected by Aretha. This was particularly the case with brand ownership, where activities would show on the visualisation as unexpected companies due to previously unknown corporate structures.

Like the other participants, P3 expressed reluctance to use the firewall due to the risk of unexpected consequences. Unsatisfactory experiences with parental controls when her daughter was growing up factored in to her decision to not use the firewall. The experience did, however, make her much more aware of control mechanisms that she had subconsciously been ignoring. Suddenly she began to notice the cookie opt-out dialogues presented to her on websites, as well as the permission prompts on her phone:

"So I guess under normal circumstances I wouldn't have cared, but now I'm thinking I don't want you to have access to my phone, I don't want you to have access to my contacts"

Much more important for P3 though, were the practical steps they could take as a family. Shared discussions about the way the family used and shared online accounts led to a moratorium on account sharing by their daughter, as well proposing the creation of a 'bogus' email account that could be used to sign up for apps and services. As a result of these steps P3 reported feeling more in control of their data than at the beginning of the study.

P3's conception of what privacy meant to her also evolved over the course of the study. With some prompts for clarification, she described a purpose-based model similar to that of P2, where data that was extracted from the

household needed to be related to a specific purpose that benefited them (such as processing a transaction). This included discussions around data inference, where P3 had used the educational content from the curriculum to develop her understanding of the ecosystem their apps and devices existed in.

5.5 Discussion

Revisiting the Design Goals - What Did We Learn?

Our technology probe aimed to empower people by providing the three kinds of support outlined by our probe design goals. In this section, we revisit these design goals in order to reflect on participants' experiences as they relate to each kind of support, before examining how they interrelate.

With respect to *DG1 (Legibility)*, there was much evidence to support the view that, by the end of the study, households were significantly more aware of data disclosures of their connected devices. In general, each participant had their own expectations of what the probe would show, and subsequently went through a process of reconciling this with the reality shown by the visualisations. While to begin with, participants were mainly interested in *who* devices were sending data to, over the course of the deployment this shifted to *what* information devices were sending and *why* as their awareness and curiosity grew.

Notably, participants tended to focus on the breadth of their data exposure rather than the volume of data going to any particular destination. To an extent this is to be expected given that novel destinations invoke more curiosity than larger traffic magnitudes, but P2 was the only participant that mentioned the 'disproportionate' amount of data flowing to Google and Facebook. For the other households, the unfamiliarity of new companies represented a bigger threat than the inference risks of larger ones, even though on average the top three companies for each household accounted for 74% of the traffic collected by Aretha. Participants were more familiar with these companies, and trust certainly plays into the assessment of risk, but one might have expected more interest in this asymmetry given that user generated privacy strategies in prior work have included depth versus breadth of exposure as a central theme [337].

Participant experiences with respect to *DG2 (Interpretability)* were more varied. On one hand, there was evidence to support the view that the educational material supported interpretation and understanding of the network data flows. All participants said that they believed the content in the curriculum had made them think differently about their exposure. The participants who reported engaging more with the curriculum (P1 and P3) also began to incorporate the curriculum concepts and terms into their discussions about privacy, most notably around data sharing. On the other hand, P3 delegated

some of the curriculum mini-lessons to her partner, who was the primary individual responsible for managing the technology in the home. The material was also not seen as useful by all participants; P2 felt that he was already aware of the concepts presented and as a result engaged less with the curriculum content.

With respect to *DG3 (Actionable Controls)*, it was clear that the firewall control did *not* meet the intended design goal, but still yielded valuable insights into how future tools might better achieve this. Participants created very few firewall rules to block traffic, even though these controls were made contextually situated, easy to use, and prominent within the assistant. We believe there were several reasons for this; a primary reason being that there were simply too many data destinations to worry about, a majority of which corresponded to companies unfamiliar to participants. As a result, participants felt they had to evaluate each one individually, and this was simply too much work. The fact that the probe failed to indicate whether it would be 'safe' to block destinations also affected firewall usage; when one of P2's blocking actions accidentally disabled essential functionality, they were discouraged from using it further.

Where participants chose their own alternative methods of control, we saw how they took the situational awareness and conceptual grounding that they had been given through Aretha and integrated this with their existing social practices (P1's ethical consumerism and P3's family discussions). Even though P2 did not develop control mechanisms during the study, in the entry interview they expressed a desire to build the probe into their existing security practices to find vulnerabilities in their network. We see that the strength of future tools likely depends on how well they integrate themselves into day to day activities, rather than the raw power they provide as technical tools.

Supporting Roles in Privacy Management

During the study, the probe seemed to serve a variety of different, yet key roles in supporting household privacy management practices. We briefly describe three such roles in this section.

Formulating and experimenting with privacy preferences and strategies—A primary role was in supporting the grounding of privacy goals and strategies in terms relating to specific entities, actions and relationships (bringing them closer to the basic ontology of contextual integrity [262]). At the start of the study, participants expressed concerns in more abstract and general terms, including those corresponding to privacy risks. By the end of the study, however, participants described personalised goals in terms of actors (companies/data controllers), devices, and disclosure actions. Moreover, they were readily able to generate ideas for strategies to achieve their goals, and translate these into actions they could take. Finally, the visualisations facilitated experimentation

by providing real-time feedback of the effects of their actions on their exposure, such as accessing certain web sites or using particular apps. Such feedback could be essential for enabling participants to test strategies by assessing the effectiveness of individual actions they might take to achieve their goals (thus supporting the kind of positive feedback loops between beliefs and action noted in prior research [343]).

Increasing salience and fostering interest in other householders—A second important role was the way that the probe brought about not just a heightened awareness of privacy concerns—by being a salient and visible physical totem in their living room—but a transformative one, in which privacy management transitioned from being a solo concern anchored with the individual responsible for technology within the home, to one discussed between multiple home stakeholders. The visibility of the probe within the home inspired conversations among household members, and the visualisations provided a common ground for discussion of disclosures, activities, and devices.

Supporting accountability of devices, companies, and users themselves—Participants reported on how legibility (*DG1*), interpretability (*DG2*), and control (*DG3*) did and potentially could further empower them in holding devices, companies, and even each other to account. While each household in the study used the affordances of Aretha for different purposes, the lens of accountability allows us to see how these features are reflected in their use of the probe.

P1 felt empowered through insights and knowledge not only to express clear preferences of not using the services of unethical companies but also stated their preference of not having business with any partners of those companies. They identified these partners by proxy of information flows "between" companies from the probe's user interface. In doing so, P1 used the probe to hold *themselves* accountable to their own moral standards, as well as the companies they deemed to be operating unethically (see the Amazon example in the results).

For P2, the probe presented a means by which they intended to hold their *devices* accountable as a way to search for security vulnerabilities in their network. However, they were discontent with the amount of data being shared with any of the large internet corporations, asking what data was being shared and why. They further expressed the desire to better control these flows of data and hold these companies to account in the future.

The case was different for P3 who was less familiar with the intricacies of connected devices, their data collection, and manufacturers' processing practices; they were not responsible for technology in the family. However, they did engage their partner "who knew all of this already" in a conversation on the educational content, and talked to their daughter about sharing accounts with friends.

These examples of reported preferences and behaviour from experiences with Aretha resemble notions of ordinary privacy practices reported in prior research, hinting at the potential for privacy-empowering technologies to "[enable] people to manage their relationships within the home and with others beyond it" [78].

Informing the Design of Future Home Privacy Tools

Reflections on the design goals for privacy-empowering tools in the home show the interrelatedness of situational awareness and understanding; information from visualisations and other mechanisms must be tightly interwoven with educational content. A good example of this from the study is the contextual examples provided alongside curriculum modules, providing concrete, personalised examples of how activities in the home related to the more abstract concepts being taught. The shift in questioning from *who* to *why* and *what* presents an opportunity to explore this further. Visualisations and controls based around behaviours rather than destinations place a focus on the types of information that different companies have, and invite questions over what could be done with the information that were missing from the study responses. The design of the 'Polly' smart kettle might serve as inspiration for how to combine these [215].

The ways that participants used the probe to hold people and companies accountable at a number of different levels suggests that it might be beneficial from a research perspective to highlight this when designing controls. The firewall in the study was concerned with *companies* and *devices*, but controls could similarly be concerned with *people*, *market sectors*, or even entire *corporate structures* (e.g. block data to companies owned by Facebook).

P2's response to the curriculum suggests that future tools could also benefit from tailoring content to the prior knowledge of participants in order to maintain engagement. Additionally, when P3 delegated parts of the learning experience, it exposed the aims of the probe as running counter to the role of P3's partner as the household device administrator—explicitly tailoring a curriculum to each member of the household may offer a means of furthering DG2 in future studies.

Rethinking Firewalling

Addressing the challenges identified with DG3, several approaches could make firewalling actions easier and safer. One idea came directly from P1, who suggested outsourcing some of the burden to assistants or experts. One example might be to allow experts to pre-curate custom blacklists that could be selectively enabled and later be overridden by users if necessary. This 'ad

blocker for the home' approach, which is one taken by tools such as PiHole[3], seems promising and almost obvious as a useful feature in retrospect, and should be strongly considered in future versions of Aretha.

A slightly more sophisticated approach could be providing more powerful operators for allowing sets of destinations to be blocked en masse based on their characteristics (or those of the data flows themselves). Based on the study, useful characteristics would likely include company reputation, jurisdiction, purpose of data collection, retention and disclosure properties. A 'learning by example' approach could interactively help people block hosts based on a single action (e.g. "if you blocked destination X, you might want to block similar destination Y").

Beyond making it easier, it was clear that what was needed was better visibility and feedback to blocking actions—i.e. a way to make the potential effects and consequences of blocking actions clearer in advance. This stemmed from the ability of the firewall to break the functionality of apps and devices in ways that might not be immediately apparent. The lack of immediate cause and effect increased the difficulty of diagnosing problems, and was further complicated by the fact that only one household member typically had the expertise required to fix problems related to the firewall. The 'management by exception' approach in which privacy and security decisions are revisited primarily when problems arise [167] suggests an opportunity for giving people better tools for not only debugging but revisiting and refining their blocking decisions when things break.

Enhancing Legibility and Accountability

A key limitation we identified going into the study was the inability for the probe to inspect and make visible the *contents* of traffic flows. Our findings suggest that such a capability would nonetheless be beneficial to improving legibility and accountability. As described above, participants were interested not just in the identities of data controllers, their business models, and reputations, but the contents and sensitivity of what was disclosed, shown by questions pertaining to why the data were captured, and how such data would be later used.

As the need to harden systems against attackers motivates developers to eliminate the few remaining methods used by researchers to directly intercept encrypted traffic contents, it will be increasingly challenging to reliably realise such an inspection capability. Machine learning driven approaches like HomeSnitch [267] that classify device behaviours based on network traffic are promising, but come with several limitations, including that models need

[3] `pi-hole.net`

to be trained on specific devices and behaviours beforehand, and may not identify the complete contents of such flows.

Enabling end-users to gain an awareness of how data are used, retained, and disclosed by data controllers is another challenge entirely. We intend to explore how legal approaches, including rights given through the GDPR, might be applied to crowd-source such information in future work.

Balancing Destinations and Accumulative Risks

Another limitation of the Aretha probe may have been in the ways it made salient the numbers and identities of destinations, potentially shifting participants' focus away from the large quantities of data being collected by platforms such as Facebook, Amazon, and Google. Although these platforms appeared prominently in every visualisation, they were often not the focus of concern. Future privacy-empowering tools should try to seek a better balance between emphasising destinations and the accumulative privacy risks of data harvested by large platforms.

5.6 Limitations and Future work

In line with prior HCI technology probe work (e.g. [48, 161]), we opted for the depth of longer-term case studies with a carefully selected small sample of families. Inherent to the nature of the method, this means that the results can provide holistic understanding of individual appropriation of the technologies (and a resulting design inspiration). However, such a sample is clearly not meant to be representative of the wider population: the observed experiences were likely influenced by selecting early adopters, who were already interested in the aims of the study. It will be interesting to contrast the temporal trajectories and impacts Aretha (and similar technologies) might have on less interested or technology savvy audiences. In addition, future deployments of similar probe-like technologies could be a useful tool with which to unpack the relationships between privacy preference formation (especially across a range of skill and interests), and the shared use of tools by participants. Finally, the particular instantiation of the proposed design goals within Aretha is far from being the only 'solution'; nor is it the best solution possible. In the spirit of the technology probe approach, we suggest that our specific design choices are seen only as an early exemplar for a broader class of privacy-empowering tools, hopefully inspiring a variety of different socio-technical ensembles to address the underlying research questions in future.

5.7 Conclusion

Given their unique contexts, members, and skills, each of our three households went on a different journey over the course of the study. Aretha enabled us to probe the ways they could understand, interact with, and use their connected devices and explore the potential for future privacy-empowering technologies in the connected home. Not all elements of the design goals were successful for all participants, pointing to a range of unmet needs (especially in relation to active controls). The lack of engagement with the firewall was instructive in its own way; while most participants found it difficult to use effectively, due to having already observed, interpreted, and understood the underlying behaviour of their devices they appeared better able to adapt, invent, or imagine other protective mechanisms, tools, and strategies.

The understanding of smartness gained in the previous chapter had already highlighted the ways that privacy is entangled in different types of smart functionality across the 'seven veils' (e.g. relating to transparency and the uncanny valley), and this allowed for the crafting of the Aretha probe in a way that addressed gaps in the wider privacy narrative around smart devices. We combined three techniques used independently in prior work to help users form and realise privacy preferences around connected devices in the home, culminating in a design which featured a combination of data flow visualisation, data protection curriculum, and high level firewall control. Reporting on the deployment of the probe in three households over a six week period, we saw how the probe allowed people to hold devices, companies, and each other to account, as well as how people can use the information provided by this and similar software to craft their own solutions and mitigations to personal privacy concerns. These results show promise as a means of empowering people to take control of their digital privacy, and help to counter narratives like that of the privacy paradox, which suggest people are not willing or capable of taking these actions of their own volition. The results of this chapter also add nuance to the seeming bluntness of privacy as a tool to control the unethical behaviours of smart devices.

Continuing the exploration of related ethical concerns, the results showed that even though focus of the probe was on privacy, participant's management of and engagement with data flows shown by the probe seamlessly blended privacy concerns with related ethical concepts; participants used the probe to hold each other to account in their roles of administering and using devices, and sought transparency from the companies that were collecting their data. Social dynamics dictated the nature of responses and subsequent solutions, often leading to more effective outcomes than would have been possible with purely technical measures.

Devices like Aretha act as trusted gatekeepers on the home network, keeping other devices in check. While there are many intuitive signals that inform per-

ceptions of such devices, developments in smart technology add a potentially disruptive newcomer in the form of speech recognition and synthesis. The next chapter explores how speech alters perceptions of devices, in ways that go beyond the understandings of privacy and smartness experienced thus far in the thesis. By exploring this latest frontier of smartness it rounds out the body of research and paves the way for the discussion of respect as a lens through which we can anticipate and deisgn for ethical concerns in the smart devices of the future.

Chapter 6

Engaging with Voice Assistants as Social Actors

"There are big buckets of information that we recommend that you share with only your friends privately. All the really sensitive stuff like your contact information, your address, your phone number"

— Mark Zuckerberg, 2010 *D8 Conference* Interview

Contents

6.1 Motivation

There is something about devices that speak and listen to us that makes them *feel* different. When P13 in Chapter 4 suggested that Alexa "occupies a room", they specifically referred to the fact that it was waiting for a response. Prior

A version of this chapter is currently under submission as "William Seymour and Max Van Kleek. 2020. *Exploring Interactions Between Trust, Anthropomorphism, and Relationship Development in Voice Assistants*. Submitted to the Proceedings of the ACM on Human Computer Interaction (PACM HCI). Association for Computing Machinery, New York, NY, USA."

work has shown how people often refer to devices and companies as if they were people [217, 337], and events like the call for Alexa to feature a 'please and thank you' mode for children show how the way that people think about and interact with these devices differs from conventional computers.

This is perhaps to be expected. Existing research by Nass et al. on conversational interfaces suggested that speech is so tightly integrated with many interpersonal functions in the brain that we cannot help but respond socially to devices with voices [253, 255, 289]. Weizenbaum had already said of ELIZA "how easy it is to create and maintain the illusion of understanding [...] A certain danger lurks there" [345], and Rosalind Picard was cognisant of the capability of affective computers to mislead [277]. Despite this, it was taken for granted in the early days of conversational agents that human-like, embodied agents were the "logical extension of the metaphor of human–computer interaction as a conversation" [61]. It seemed obvious to make use of communication skills that people already possessed, as well as the fact that natural language has built in redundancy and can be supplemented by gestures, tone, and facial expressions. On the face of it, voice interfaces represented an ideal 'invisible' human-computer interface.

This social aspect of voices and conversation raise challenging questions about how they might change the way that people interact with and perceive smart devices. Explorations of this in science fiction have existed for over a century. In the 1927 silent film *Metropolis*, the inventor Rotwang creates a 'maschinenmensch' to resurrect his dead lover [198]. One of the first robots to ever appear in film, its human appearance is used to manipulate relationships amongst the main characters and start an uprising. Several decades later, Asimov coined the term robotics and predicted the social and emotional challenges of interacting with realistic 'humaniform' robots [15]. More recent films such as *Her* and *Ex Machina* have continued to explore relationships with machines that look and/or sound human in a modern context [129, 175].

It is clear that the potential for devices with conversational interfaces, in particular smart voice assistants, to interfere with the ways that people interact with their devices and the ways that they interact with each other in environments where those devices are present is great. Assistants that are commercially available today—such as Amazon Alexa and the Google Assistant—are sophisticated conversational agents, leveraging advances in speech recognition, natural language processing, and speech synthesis. Uncertainty around how voice assistants worked and their often uncanny behaviour led to participants in Chapter 4 being unsure how to appropriately treat them, and P13's comment echoes [63] in suggesting that these devices have a *social presence* similar to that of a person. Responses like this hint at the complex set of cognitive processes that determine how we perceive and relate to these devices, especially when they are positioned as 'assistants' or otherwise subservient.

This suggests that the relationships that we have with voice assistants are inherently different from those that we have with other smart technologies [327], falling somewhere between the clear-cut categories of human and machine. But while there have been many contributions investigating how people use voice assistants, little work has been done to explore the *nature* of our relationships with them, including the potential cognitive shifts that come about when conversing with agents that presents themselves as human. Given that these agents rely on the same social skills utilised in developing interpersonal relationships, it is important that we better understand the similarities and differences between human relationships with voice assistants and relationships with other humans. And if this comparison *does* makes sense, then how do human-assistant relationships relate to other interaction components that are already well understood, such as trust and anthropomorphism? To this end, we pose and answer the following research questions about how people interact with voice assistants:

RQ3a Do users of voice assistants exhibit relationship development with their devices similar to that measured between humans?

RQ3b Are trust, relationship development, and anthropomorphism related for voice assistants?

RQ3c What underlying assumptions exist in the design of contemporary voice assistants?

6.2 Human Interpersonal Relationships

Beginning with more conventional human relationships, a large body of psychology and communications theory literature is dedicated to understanding and classifying facets of human interaction. While many theories of interpersonal interaction focus on experiential aspects of interpersonal relationships that are potentially less relevant to smart devices (such as sexual desire or interdependence [62, 88]), others have proposed 'stage' models of relationships that encompass the behavioural aspects of relationships from meeting to parting. Murstein [250], Altman and Taylor [12], and Knapp and Vangelisti [189, 190], each proposed stage theories of relationships, suggesting that relationships develop up (and down) identifiable strata of feelings and behaviour.

While Murstein's work pertains more to partner selection, Altman and Taylor's work on the process of social penetration conceptualises wider relationships in terms of breadth and depth of interaction [12]. As individuals progress along a relationship trajectory of four stages they move along both of these axes to varying extents. Attempts to operationalise these breadth

and depth concepts typically involve rating different disclosure topics for intimacy [323]. Knapp's staircase model of relationships also documents the shift in interpersonal communication as a relationship develops, and integrates a number of core ideas from social penetration theory [189, 190]. The staircase model proposes five stages of both development (from *initiating* to *bonding*) and parting (from *differentiating* to *terminating*) that describe communication patterns within relationships. As with social penetration theory, movement through stages can be bidirectional and is linear within meeting/parting trajectories. After initial work by Avtgis et al. [18], Welch and Rubin later adapted Knapp's model to identify *behaviours*, which are more replicable than emotions (e.g. using the extent to which individuals share information as a partial proxy for intimacy) [346].

6.3 Methodology

We designed a survey to answer the first two research questions given above, exploring the interplay between three different axes along which users might perceive their voice assistants:

- Voice assistants as mechanisms and tools: *How reliable users find their voice assistants, and how confidently they believe they can model and predict their actions.*
- Voice assistants as humans: *How much users anthropomorphise their voice assistants and ascribe them human characteristics and intents.*
- Voice assistants as social actors: *How users' interactions with voice assistants relate to social rules and models describing interactions between humans.*

Questions for the survey were drawn from the background literature and split into three types, each measuring a different axis of trust, anthropomorphism, or relationship development. The 15 questions on the trust axis were taken from a psychometric instrument previously used on decision support systems [223], adapted from [247].

The 15 relationship development questions measure characteristics of Knapp's 'staircase' model of human relationships [189], itself based on Altman and Taylor's theory of social penetration [12]. This model was operationalised by Welch and Rubin [346], and after removing the parts of the instrument that our two pilot surveys deemed inappropriate in the context of voice assistants (e.g. exchanging physical tokens of affection), questions from the 'escalating' part of the scale were incorporated into the survey. The remaining questions were then modified to specifically refer to voice assistants (such as changing

Q3	trust	Alexa responds the same way under the same conditions at different times
Q7	trust	Alexa gives me information that is as good as a professional person would give
Q11	trust	Although I may not know exactly how Alexa works, I know how to use it to perform tasks I want Alexa to do
Q19	relationship	I'd like to get to know Alexa better
Q21	relationship	I feel guilty when asking too much of Alexa
Q30	relationship	If Alexa were a person, I think we'd get on with each other
Q31–33	human-nature	Curious, Friendly, Fun-loving, ...
Q41–43	human-unique	Broadminded, Humble, Organised, ...
Q51–53	robot	Artificial, Automaton, Mechanical, ...

Table 6.1: Sample questions used in the survey. The full set of questions is available at `https://osf.io/53q6j`.

references from he/she to Alexa, and avoiding pronouns in order to prevent priming participants towards anthropomorphism).

The 30 survey questions measuring anthropomorphism were taken from Loughman and Haslam's work on human trait categories [219], incorporating changes made by Fussel et al. who previously applied these traits in an HRI context [127]. Participants were presented with a word describing a characteristic that was either representative of human nature (10 questions), uniquely human (10 questions), or associated with automata (10 questions) and asked to what extent it described their voice assistant.

The order of survey questions was randomised, and participants were then asked to mark how well each statement or characteristic described their virtual assistant on a seven point Likert-type scale (*Strongly disagree, Disagree, Somewhat disagree, Neither agree or disagree, Somewhat agree, Strongly agree*) [115]. Likert responses were coded from 1 (Strongly Disagree) to 7 (Strongly Agree), and summed to give a score for each metric. Scores for the human nature and human unique anthropomorphism aspects were summed in order to give a joint human score.

Participants for the survey were recruited using the Prolific Academic platform and the survey was administered using JISC Online Surveys. The study was reviewed and approved by our institution's research ethics committee, and participants were compensated at a rate at or exceeding the UK Living Wage of £8.21/hour. Survey participants were required to be 18 or over, UK residents, and own a voice assistant. A sample of the questions used in the survey

Metric	α
Trust	0.90
Relationship	0.88
Anth. (H-Nature)	0.78
Anth. (H-Unique)	0.70
Anth. (Robot)	0.81

Table 6.2: Internal consistency of the survey metrics.

	Trust	Relationship	Anth (Human)
Relationship	**0.579***	-	-
Anth (Human)	**0.306***	**0.403***	-
Anth (Robot)	-0.039	**-0.284***	0.016

Table 6.3: PPMC coefficients between the survey metrics.

is shown in Table 6.1, and the full set of survey questions and anonymised results are available at `https://osf.io/53q6j`.

6.4 Trust, Relationship Development, and Anthropomorphism

Of the 515 survey responses received, fifteen incomplete responses were omitted to leave a total of 500 complete responses. Participant ages ranged from 18 to 82, and the average age of participants was 36 (σ = 11.6). When asked to report their gender identity, 316 (63%) participants identified as female and 184 (37%) as male. Of the total responses, 328 (66%) participants owned a device with Alexa, 121 (24%) with the Google Assistant, 46 (9%) with Siri, and 5 (1%) with other assistants. Except where comparing owners of different devices, only responses for Alexa devices were analysed in order to answer the research questions given above.

When giving results, a * denotes a p value < 0.01. Before conducting further analysis, each range of questions was measured for internal consistency using Cronbach's Alpha (α). Scores indicate that the metrics retained consistency (i.e. questions on each axis measure the same general construct) and are reported in Table 6.2. Where given, correlations between metrics were calculated using Pearson's Product Moment Correlation Coefficient (PPMCC, ρ). Correlations $\rho < 0.2$ are reported as small, $0.2 \leq \rho < 0.5$ as moderate, and $0.5 \leq \rho$ as large.

Of the main axes, answers to trust questions were the most tightly grouped ($\sigma = 1.33$), followed by anthropomorphism (robot) ($\sigma = 1.55$), anthropomor-

Metric	Amazon	Google
Trust	**0.483***	**0.388***
Relationship	**0.502***	**0.458***
Anth (Human)	**0.333***	**0.257***
Anth (Robot)	-0.053	-0.022

Table 6.4: Correlation coefficients between trust in Amazon or Google (as appropriate) and the three survey axes.

phism (human) ($\sigma = 1.82$), and finally relationship development ($\sigma = 1.90$). The survey results show a high degree of correlation between participants' trust in Alexa and relationship closeness (0.579*). There was a moderate degree of correlation between anthropomorphism and both relationship and trust (0.403* and 0.306* respectively), and between relationship and robot anthropomorphism (-2.84*). The complete matrix of correlations is given in Table 6.3. Responses to individual questions broadly showed good engagement (e.g. Q25, "If Alexa were a person, I think we'd get on with each other" $\bar{x} = 4.29$, $\sigma = 1.37$), but occasionally did not despite the pilot surveys (e.g. Q37, "I feel guilty when asking too much of Alexa", $\bar{x}$=1.90, σ=1.17).

In addition, the use of a personal Amazon account with Alexa had a moderate correlation with relationship development (0.207*). No other significant correlations existed between the survey axes and ownership or demographic characteristics. It is worth noting that the 248 (76%) participants using their own personal accounts with Alexa are likely to also have configured the device as account details are input during setup. There was no correlation of ownership duration and number of Echo units owned with trust, however participants' trust in Amazon as a company was correlated with all three of the axes present in the survey (Table 6.4). These results were mirrored for trust in Google amongst users of the Google Assistant.

Relationship development scores appear to be normally distributed across participants (see Figure 6.1), with a mean of 51.3 and standard deviation of 14.1 (the maximum possible score was 105).

Across Assistants

Analysis of the 130 responses from Google Assistant users paints an interesting picture. As with Alexa owners, there were correlations between trust/relationship, and anthropomorphism/relationship values (see Table 6.5), suggesting that these phenomena may be shared across voice assistants rather than being specific to a particular product. We would expect to see a significant correlation between trust and anthropomorphism given a larger

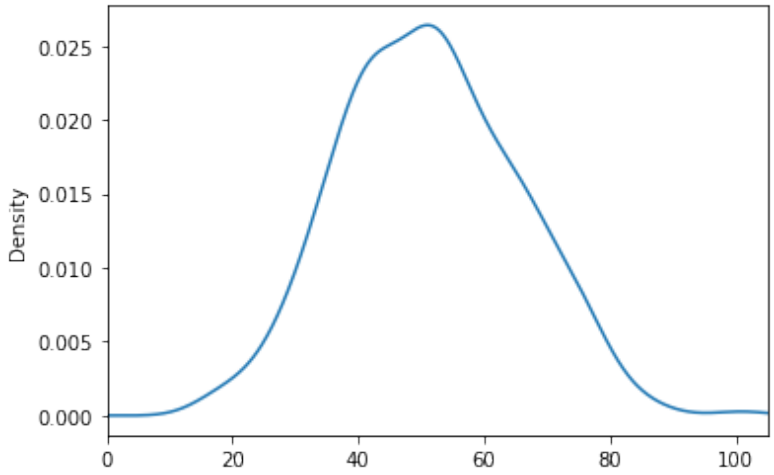

Figure 6.1: Density of Relationship Development Scores showing a single peak.

	Trust	Relationship	Anth (Human)
Relationship	**0.529***	-	-
Anth (Human)	0.199	**0.502***	-
Anth (Robot)	**0.240***	-0.001	0.220

Table 6.5: PPMC coefficients between the survey metrics for Google Assistant users.

sample size. Interestingly, Google Assistant users also showed a correlation between *robot* anthropomorphism and trust.

6.5 Discussion

The first question that the study sought to answer was whether it made sense to ask questions intended to model human relationships in the context of voice assistants. While in some cases it clearly did not, as evidenced by the low scores and spread of responses for particular questions, in the majority of questions it did. Combined with the fact that greater relationship development was linked with greater anthropomorphism (both higher human and lower robot scores), it would appear that there is merit in adapting existing interpersonal measures to study human-like devices. It also provides empirical support to the idea that these two types of interactions operate using similar cognitive mechanisms.

We therefore answer RQ_1 in the affirmative—it *does* make sense to consider relationships and relationship development with voice assistants in similar ways to relationships between people (though we note that the extent to which relationships with voice assistants are the same as relationships between humans is unclear—see the study limitations below).

In relation to RQ_2, there were moderate to strong correlations between trust, relationship development, and anthropomorphism shown by users towards their voice assistants. The fact that all three axes increased in unison would suggest the presence of an underlying factor, although it is not clear from the results of the study what this could be. Previous contributions in HCI have shown the influence on relationships of personal circumstances that vary over time—such as loneliness driving personification [284]—and this combined with the inclusion of innate psychological factors in prior work on HCT [109] may offer a template for this to be further explored in future work.

The only secondary ownership characteristics that correlated with the main survey axes was the use of a personal account (with relationship development); other characteristics that we would have expected to show significant correlations did not (e.g. ownership duration). It is possible that the influence of an underlying factor (as described above) dominated those of other characteristics such as ownership duration. But we did not expect this to also be the case with trust, instead expecting users to show measurably increased trust as they interacted with devices over a longer period of time. In future work we intend to more fully explore the effects of what Epley et al. term 'sociality motivation'—the ability for nonhuman agents to satisfy our need for social connection [105]—on relationship development and how it might relate to the development of trust in devices.

Blurring the Line Between Human and Machine

Across the different survey axes, we see that trust scores and the 'robot' nature of Alexa appear to be relatively uncontroversial, with less variation in participant responses. What differed the most between participants was the extent to which Alexa was treated and interacted with as a human. While it is widely understood that machines can engender social and emotional responses [176, 277, 328], the results of this study show again that this is the case even in devices that were not explicitly designed to do so (albeit to a lesser extent).

As developments in speech recognition, natural language processing, and speech synthesis cause voice assistants to increasingly blur the line between humans and machines, one would also expect to see an increase in 'purification work' as described by Leahu [204]. People continually expend effort in order to keep categories such as 'human' and 'machine' distinct, through a combination of attributing characteristics of machines that overlap with

"real people" to a device's creators, and redrawing the boundaries between difficult combinations of categories [204]. In the present work, we would have expected this to surface in relationship scores as a bimodal distribution, as participants sought to push Alexa firmly into either of the categories. However, this was not the case in the survey results, which had a single clear peak, suggesting that after years in the popular zeitgeist this purification work may have already been undertaken (Figure 6.1).

Distributing Trust Amongst Voice Assistants

It is well understood that socioemotional responses to robots can also be beneficial in certain circumstances. Work in HCI and communications theory shows that these responses make use of cognitive systems designed for interpersonal relationships, but this can pose a problem when deciding whether to trust a virtual agent.

The centralised, cloud-backed nature of Alexa and similar products means that in order to trust Alexa, one must also trust Amazon; each Echo unit is a part of the same 'hive mind'. This collective structure interferes with our ability to distribute trust across the people that we meet—we understand that acquaintances have loyalties and obligations to third parties, but it is unusual for them to be completely controlled in this way. However, the results of the study are promising in this respect. The correlation between trust and Amazon and trust in Alexa suggests that (broadly) people *do* rationally allocate trust in Alexa (i.e. the similar allocation of trust in Alexa and trust in Amazon is rational given the underlying relationship between the two).

Ethical Challenges

Voice assistants are clearly designed to be fun to interact with, as evidenced by the widespread inclusion of jokes and other casual responses. The use of human voices similarly makes using them enjoyable, to the point where the failure of the voice assistant to carry out a task successfully can still result in a positive interaction [218]. The results of the survey show that the use of Alexa invokes behaviour that mimics human relationships, suggesting th urgent need to explore related ethical concerns. If making these agents more social subconsciously changes the ways that people interact with them, could this be used for manipulation? Prior work has already shown that users disclose more information through interactions when they are mediated by computers [173], and explored ways in which users can be persuaded to disclose information to computers [246]. These existing results, when combined with the survey presented here, open a new avenue for investigation in future work. Given the integral role of disclosure in interpersonal relationships [145], one would expect that increasing the relationship development between a user and their device would also increase the amount they were willing to disclose to it.

Many of the tactics used by humans to manipulate each other, such as being charming, adopting childish mannerisms, and giving someone the silent treatment [55] could conceivably be used by social robots to alter our behaviour. Indeed, many social robots (such as Paro) are explicitly designed to be cute and charming. This space provides an interesting and potentially lucrative extension of controversial practices currently seen on web and mobile platforms, such as dark patterns, microtransactions, and operant conditioning, by leveraging users' emotional investment in the device in order to promote certain behaviours. While slightly strange, one could imagine a social robot withholding emotional contact until further purchases are made, or emotionally rewarding users for making a purchase. But as with many similar interaction techniques, affective emotional interactions can be used for good as well as ill. Utilising social interactions with voice assistants to help users in stressful situations or make learning more engaging, for example, represents a genuine social good. Similarly, understanding as designers when it is not appropriate to trigger these responses is of paramount importance (as originally noted by Picard [277]).

As such, we need to decide as a society the extent to which we wish to interact with agents that exploit social processes. When Google unveiled 'Duplex', an AI that could make phone calls using convincing human speech, there were widespread calls for Google to ensure that users were not 'tricked' into thinking they were conversing with a person [314]. But as Nass's experiments and these results show, knowing that you are talking to a computer is not enough to prevent people applying deeply ingrained social processing to interactions. One approach would be to limit the functions that social robots are permitted to perform, such as those involving money. In other situations, more clearly telegraphing the inner workings of devices would both limit social engagement as well as assist users in building accurate mental models.

6.6 Limitations and Future Work

While this initial exploratory study into the nature of our relationships with voice assistants has shown that they do to some extent mirror those between people, it raises as many questions as it answers. For instance, what is the causal nature of the relationship between trust, anthropomorphism, and relationship development? How easy is it to push users towards or away from social responses through design decisions? In future work we aim to test these relationships further.

While the survey axes showed good internal consistency, the extent to which they exhibit construct validity is not yet known (i.e. measure the same phenomena exhibited between humans). For example, it might be that the same concepts and development trajectories exist for interactions with machines,

but with deviations at either a high or a low level. It may also be the case that given the unusual nature of the questions presented to them, participants interpreted the relationship development questions differently (although the α values would indicate that they are still measuring a single general construct). In addition, the survey largely focuses on one device, context, and culture. Similar devices exist across the globe in a variety of cultures and contexts, and we look forward to following up with a more diverse selection of participants to see the extent to which these responses are shared between us.

6.7 Conclusion

Speech is an important affordance, and the sophistication of modern voice interfaces show how far capabilities have progressed since early text-only assistant devices. The transition to *conversational* interfaces that allow for easy and natural interaction with devices represents a profound shift in the nature of the systems we interact with towards the increasingly social. This shift brings with it a variety of new or exacerbated ethical concerns that designers of voice interfaces need to consider when designing future products. many of these are related to privacy, such as constant surveillance and the potential for social interaction to change the way that people perceive devices.

The exploratory survey presented in this chapter provides evidence that people do develop relationships with voice assistants that are social in nature, and that these are linked to perceptions of trust in devices, trust in manufacturers, and anthropomorphism of those devices. In outlining and discussing the implications of these links, we provide an opportunity for developers of current future conversational agents to steer their designs away from potential ethical problems. Future research and societal conversations are needed in order to determine the nature and extent to which these technologies are integrated into our lives.

In the context of the wider thesis, these results justify the unease around the potential effects of using voice interfaces (whether coincidental or intended) and show important non-technical impacts that voice assistants have on people's lives. They are also suggest that different methods are required to tackle what are these types of issues with smart devices that are increasingly *social* in ways that were not previously the case. If devices continue to lean on the social aspects of their interactions with people to cover up their flaws and develop emotional connections with their users, then perhaps they should be similar judged according to interpersonal standards of conduct (as opposed to those reserved for tools and machines).

This notion is supported by the results of previous chapters as well. The users of the Aretha probe notably elected to enact social controls on their devices rather than the technical ones offered by the probe, and while there were also

usability reasons for this, their imaginative solutions were able to address interpersonal concerns that the firewall would not have been able to.

To this end we explore the possibility of conceptualising and evaluating devices according to standards normally reserved for interpersonal interactions in the next chapter, applying the concept of respect to the design of smart devices to work towards a future where technology better supports human development and flourishing.

Chapter 7

Conceptualising Respect: Supporting the Ethical & Inclusive Design of Interactive Systems

"A relatively inexpensive smartphone with a browser is all you need to get the world's information"

— Eric Schmidt, 2013 *D: Dive Into Mobile* Conference

Contents

7.1 Motivation

The previous chapters have given many insights towards what smartness is and how it operates throughout a range of contexts in the connected home.

The results of the thesis thus far have increasingly suggested the need for an ethical framework to inform the design of smart devices and balance companies' quest for profits with ethical concerns: the survey and interviews exploring perceptions of smartness gave rise to a rich understanding of the impact of opaque data collection functionalities, the potential to accidentally or purposefully intrude on other members of the household, and the conflicts of interest that can arise when companies integrate closed services such as app stores into devices; the naming of the Aretha probe was no accident, and its deployment showed that when given the required information and understanding, people can and do skillfully develop methods of holding devices, companies, and each other to account; and the survey exploring how the use of voice interfaces might drive relationship development uncovered hidden ethical risks that come with the use of social technologies in smart devices that change the way that they are perceived and reasoned about by people.

The integration of smart devices into the home, and the subsequent effects on the hierarchies and relationships within it suggests that smart devices could warrant a different path towards their design and evaluation; better understanding what smartness *is* in this way allows for subsequent speculation as what smartness *could* be. There are many moral principles used to evaluate the smart devices, services, and platforms that are in use today, both during the design process as well as throughout deployment.

Privacy features heavily in the literature as principle that can be broadly applied to solve such issues. Sharing less data usually makes ethical problems disappear, but in the same way that encasing a computer in a block of cement makes it secure. Data minimisation certainly has its place in the ethics of modern devices, but it is an overly blunt instrument when wielded against issues arising beyond the disclosure of data, such as interpersonal relationships and how people feel treated by their devices. Fairness as an operationalised design principle is another burgeoning area of research, and while it overlaps in places with the problems identified in the previous chapters, it is difficult

A version of this chapter is currently under submission as: William Seymour, Max Van Kleek, Reuben Binns, and Dave Murray-Rust. 2020. *Conceptualising Respect: Supporting the Ethical & Inclusive Design of Interactive Systems*. Submitted to the Proceedings of the ACM on Human Computer Interaction (PACM HCI). Association for Computing Machinery, New York, NY, USA.

to characterise many of them as 'unfair' either distributionally or representationally (although the performance of gender by voice assistants could serve as an example of the latter).

As previously hinted at, value-sensitive design is perhaps the most useful existing design 'tool' that could be applied, helping designers to focus on what is important in the lives of those affected by their products and to adapt them accordingly. But how does one choose the values that are to be designed for when there are competing interests? Determining what values should be designed for is a core challenge of VSD, and even eliciting the range of potential values that might reasonably be designed for in a given situation is exceedingly difficult given that they can be functional, ethical, economic, and so forth. A way of structuring and reasoning about which values to embed in a product is required.

This chapter explores the application of *respect* to the design of smart devices in HCI; something every person deserves but does not always receive, respect forms a person's measure of their own self-worth and a reflection of the nature and value they place on their relationship to others. Respect from others bolsters one's self-esteem and confidence, promoting positive mental health and happiness, while mutual respect builds robust social bonds and cultivates mutual trust. A lack of respect, meanwhile, can undermine a person's well-being in as many ways as being respected can benefit them, even disenfranchising them of their very personhood. These qualities of respect make it well suited to design in the smart home, and hint that it can supplement existing design frameworks and principles when augmented with the lessons learned in the preceding chapters about the nature of smartness.

These contrasting but complementary conceptualisations of respect are indicative of how contextual and varied the notion of respect is, a prime example of a portmanteau word (i.e. one that has a diverse range of concepts 'packed' into it). Respect is also important for another reason: interactive technologies are becoming increasingly adaptive and human-like, capable of responding to people using natural affordances such as spoken dialog and gesture. This means that people increasingly expect systems to act as humans would, potentially ascribing to them human-like social capabilities [289]. These systems have also increasingly come to mediate interactions *between* people, changing the ways that we relate to each other.

In this chapter we bring together a wide range of conceptualisations of respect from work in philosophy and the human sciences, establishing a robust grounding for respect as a design goal for smart devices and "third-wave" HCI more broadly [43, 68]. In doing so we answer the following research questions:

4a What conceptualisations of respect exist outside of computer science that are potentially relevant to HCI?

4b How do these conceptualisations relate to existing design frameworks and principles?

4c What steps can designers take to create more respectful systems?

7.2 Conceptualising Respect: Perspectives from Philosophy and the Human Sciences

The idea that every person is entitled to be respected as a human being is core to nearly all cultures and is enshrined in international law in the form of human rights [94]. We might therefore ask what the responsibilities of creators of smart home devices and platforms are to those whoes lives are shaped by their work; in other words, might it be the moral duty of interaction designers, HCI researchers, and systems engineers to ensure that the devices they create are made to treat people with respect, or to promote respectful interactions among those they mediate? To fully explore this question, and to understand how respect might constitute a 'new good' beyond usability for user-experience design [110], we need a robust theoretical and practical grasp of what it might mean to respect people. Despite its deceptive simplicity—most people have an intuitive understanding of what respect is—closer inspection reveals respect as a complex and multifaceted concept which can mean many things beyond interactional considerations, including such aspects as how individuals are represented, identified, and associated with others.

Motivated by the many inequalities that smart devices can (and do) perpetuate, we now give a broad survey of the main perspectives on respect from philosophy and the human sciences. As Fallman and Bardzell warn, when HCI imports theory, there is a danger that a concept is 'borrowed' and directly applied in design in a way which 'strips much of its original analytic force' [110, 22]. By exploring these multiple accounts in their own terms, we hope to outline the potential of respect as a design goal without eliding its contestable and multi-faceted nature. There has been significant work attempting to define and disambiguate respect from closely related concepts, although many have reflected on the difficulty in doing so. Langdon notes a "menagerie of articles that all claim to examine respect, but because definitions vary, do little to further understanding of the concept" [200] while Frei and Shaver consider respect "an inherently fuzzy concept" [118]. Schrimer et al., note that methodological challenges in studying respect stem from the fact that "people's ability to use a concept and their ability to define it are two separate issues" [298], although this gives hope that practical applications can emerge despite philosophical contradictions.

Respect is also a 'thick' concept, which means that it both *describes* certain attitudes or relationships, but also contains within it explicitly normative and culturally specific dimensions [351], requiring a combination of philosophical and sociological treatments. While we primarily examine conceptions of respect from the English-speaking world, and from fields historically dominated by white men, we also draw on more recent work in feminist philosophy and the philosophy of race which are necessary to articulate alternative notions of respect which address and challenge white supremacy, patriarchy, and other forms of oppression. None of these accounts of respect are intended to provide directly operationalisable design goals, but there is a growing recognition that conceptual discussions such as this may help deepen and extend the discussion currently taking place within third-wave HCI about what the field should strive for in design [110]. A brief summary of each type of respect discussed below is given in Table 7.1.

Philosophical accounts of respect can help in our search for a robust theoretical grounding for respect as a design goal by providing clarity: breaking the concept into constituent parts and articulating distinctions between different kinds of respect. Philosophical accounts of concepts situate them in networks of thought, allowing for a diverse set of explanations of what constitutes respect, its causes, effects and mechanisms.

7.3 The Categorical Imperative and Respect for Persons

Respect became a foundational concept in Western moral philosophy during the Enlightenment when Immanuel Kant wrote about there being certain essential moral duties, or obligations, that people should have of others unconditionally, regardless of context [178]. In contrast to the subjective moral frameworks of his day, Kant saw *respect for others* as part of one's *"categorical imperative"* as a moral entity [178]: that people should recognise the status, worth, and individuality of others simply for being human, regardless of their place, position, or situation. Kant's *humanity formula* describes the need to treat others as ends in themselves, rather than simply for one's own advantage. This concept of *respect for persons* is often operationalised at a national and international level in the form of human rights, explicitly outlining the treatment due to "all members of the human family" [16, 94].

Interpreting Kant's work on respect, Feinberg [113] distinguishes three treatments of respect from Kant's writing on the subject:

Respekt denotes a recogition or fear of power and acknowledgement of the ways that something would harm us if we fail to give proper consid-

eration to it, such as avoiding wild animals, or not underestimating a strong opponent.

Observantia goes beyond power hierarchies to look at a broader set of things that *should* be respected – the 'moralized analogue' of *respekt* – including the meek and humble. This type of respect describes how something makes a rightful or legitimate claim on our actions, including obligations to pay what is owed, adherence to social roles and hierarchies and that respect which every human is due. In contrast to *respect*, *observantia* is an obligation on the subject, rather than a demand from the target.

Reverentia is a subtle final type, that describes relations to higher concepts such as the 'majestic law' or reverence for the divine. In this way reverentia blends respekt and observantia into a 'the strongest force of all, where power and legitimacy blend" [113], that both requests and requires our cooperation.

The types of respect discussed above have to do with the ways that one modifies one's behaviour and may or may not show respect—they are demands for certain kinds of action. Darwall [85] contrasts this *recognition* respect—where the subject recognises an aspect of the object that causes them to modify their behaviour—with *appraisal* respect: the evaluation of an object's characteristics in relation to a standard. Although it may impact one's behaviour, *appraisal* respect is primarily *experienced*; it is also *graduated* in that one can have different amounts of appraisal respect for different characteristics of an object, such as respecting one person's skill as a painter more than another's.

Moving beyond the recognition/appraisal distinction, Hudson argues that respect for a person is not a single thing, but that there are four distinct types to be considered – *directive, obstacle, institutional, and evaluative* [156]:

Directive respect entails modifying one's behaviour in response to a stated preference, law, rule, or principle, thereby recognising its authority. Simple examples include adhering to the speed limit when driving or someone's request to be left alone.

Obstacle respect entails changing behaviour to deal with things that may get in the way of one's goals, such as respecting the threat of falling by using a handrail or making someone coffee in the hope that an application is fast tracked. This can often result in treating a person as a means rather than an end – something to be managed in order to get what one wants.

Institutional respect is concerned with social practices and institutions, the roles within them and the people that represent them. One shows institutional respect by conforming to rules and practices – many dress

codes fall under institutional respect. Notably, one can have and show institutional respect for an object without necessarily doing so for a particular element or agent—such as respecting the authority of the police without respecting a particular officer.

Evaluative respect captures how we consider the characteristics of people and objects based on the extent to which they meet certain standards. This means that a subject can experience evaluative respect to different degrees between objects, such as respecting one person's loyalty more than another's, or to different extents for the same object (e.g. respecting someone's determination but not their tardiness).

There is a clear relation between to Feinberg's typology here, as *respekt* and *obstacle respect* relate to dealing with something that could be dangerous or obstructive; *observantia* shares with *directive respect* and *institutional respect* a sense of doing the right thing, although there is no clear analogue to the subtler *reverentia*. Darwall's *recognition respect* closely resembles appraisal respect, including being experienced and graduated [95]. While Hudson's four types of respect do not directly include *respect for persons*, he believed that it would entail a combination of the above four kinds of respect [95].

The concept of respectful behaviour being that which treats people genuinely—as ends in themselves—is a powerful lens through which to evaluate smart devices that captures the *intent* behind interactions in a way that is missing from other related frameworks used as a lens to design and evaluate devices and algorithms, such as fairness or user-centred design. Hudson's four types of respect explore in more detail Kant's notion of actions that treat others as means to an end and ends in themselves. This captures the essence of how a company can provide a platform that is fair, accountable, transparent, and user-centred but still feel 'creepy' because of its data harvesting practices. The subsequent decomposition of respect into that which is owed to all people and that which is earned, as well as particular types of respect that influence how people respond to situations (like respect for institutions) is helpful in understanding how the increasing use of social settings and input modalities might be leveraged by smart devices for good or ill.

But in understanding how respect applies to the re-packing of smartness, these perspectives omit some key points. Firstly, by focusing on the intent of the entity showing respect they sideline how these behaviours are received; it matters much more how a person interprets and responds to an action taken by a device than whether that action was intended to have a positive outcome. While an intuitive motivation for interpersonal respect, rooting the basis of respect in personhood brings its own set of associated problems. Kant's work, as well as that of Feinberg and Hudson that were heavily influenced by it, are based on the assumption that personhood is a fundamental truth and vested

equally in every person (thus demanding equal treatment). On the face of it this might appear to be a strength of Kant's position, and one fundamental to the concept of human rights, but subsequent work has explored how this notion of equal treatment can be problematic.

7.4 Post-Kantian Respect

A major criticism of Kant's work on respect is that it is impersonal and abstract, "flens[ing] the individual down to the bare bones of abstract personhood" [174]. This risks ignoring the personal, social, and cultural characteristics that make us unique, as well as the network of relationships between people [111].

To this end, Dillon proposes the concept of *care respect* to Hudson's four categories described above. She integrates core principles from feminist philosophy into respect, arguing that we are not due respectful treatment because of some abstract moral capacity (as with Kantian respect for persons), but rather because we are each a concrete human 'me', individuals with our own perspectives who may rely on others to help fulfil our wants and needs [94]. In this sense, "[care] respect requires not so much refraining from interference as recognizing our power to make and unmake each other as persons [...] caring for others by responding to their needs, promoting their well-being, and participating in the realization of their selves and their ends" [94]; in addition to valuing the object of our respect as *a* person, we must value them as *the* particular and unique individual they are.

> It seems to us that the way people deploy their thoughtfulness at home is by steering a course between two opposites: mechanized routine on the one hand and relaxed, unplanned and almost chaotic behaviour, on the other. [...] Tenderness between people in the home is suffocated by routine, yet the opportunity for tenderness is squeezed by the practical requirements of living at home [322].

In addition to abstracting away a person's unique and particular nature, notions of respect discussed above may also fail to attend to the ways in which respect is unequally distributed under conditions of structural oppression, such as patriarchy and white supremacy. Feminist philosophy and philosophy of race critique Kantian and liberal notions of respect for persons on these grounds. For instance, Manne's account of the logic of misogyny notes that while respect is an attitude of good will which we expect of each other, those expectations are asymmetric by gender: 'women in relations of asymmetrical moral support with men have historically been required to show him moral respect' [225, p:xix]. Such inequalities of respect help maintain male dominance by controlling, policing and punishing women who challenge it.

Similarly, in the context of racial domination, Mills argues that Kantian respect as traditionally conceived is a form of 'racial liberalism', which assumes a 'generic colorless political subject' [243]. Others argue that it embeds racist assumptions about the scope of personhood [9]. Such critiques of Kantian republican liberalism do not necessarily mean doing away with the concept of respect altogether; they may be worth upholding provided they can be, in Mills' terms, 'de-racialised' [242]. He suggests re-shaping Kantian respect for persons to explicitly address how self-respect and respect for others "will have been affected by race (as racism), leaving an ideological and psychological legacy, habits of disrespect, that will shape the 'inclinations' most likely to be determinative and most imperatively to be resisted" [242].

Unlike Kantian respect, which is normative, outward-directed, and can be operationalised through application of the humanity formula, the critiques that followed are more subtle and nuanced; enacting care respect is not about following a certain set of rules or steps, but rather centring thought and reasoning around the individual and their unique circumstances. While on the face of it this might appear to be less useful for design as designers cannot micro-target individual users, but adaptive and intelligent systems *can*. The decision-making embedded in these typically more open-ended systems (such as voice assistants) is absolutely capable of accounting for individual nuances, as evidenced by the power of targeted advertising. Revisiting the criticism made about the assumption of equal treatment that underpins Kantian respect for persons, we can see how the work of Dilon, Mayne, and Mills highlight and address how devices might perpetuate existing injustices that have become embedded in society and culture. These approaches could help designers to take an approach to an ethics of smart devices that emphasises relational contexts and needs of individuals, rather than the abstract, universal principles associated with some existing measures (such as AI fairness). At a more fundamental level, this could also include the adoption of alternative objectives functions, such as designing social feeds to prioritise protection of the most vulnerable users rather than optimising for overall user engagement. This example also highlights how post-Kantian respect might even suggest the non-deployment of devices where they fail to meet basic standards of care and the perpetuation of oppression.

7.5 Respect as an Enacted and Dynamic Concept

An alternative perspective on respect that has thus far been missing is one focused on the *behaviour* of respect. Respect in this sense does not exist a-priori outside of human society as a moral imperative, as is true in the Kantian perspective, but rather as something that emerges from our interactions with each other based on the distinctions and norms that we create as groups, cultures, and societies. This is an important conceptual shift when thinking

about smart devices, as they lack the moral capacity required to show respect as it was imagined by Kant. It also shifts away from the intent behind actions that are taken towards how they are recieved, another key aspect of exploring how smart devices could be respectful.

Goffman's classic treatise on presentation of self [141] fundamentally presumes that all people have a moral obligation to treat others in "appropriate ways" dictated by their social characteristics. While not explicitly using the term *respect*, he develops the view that the way one acts serves the critical roles of both communicating and reinforcing others' views that one possesses one's social characteristics. This *dramaturlogical* view constitutes the world of social actions as being – or simulating – theatrical performance, and explains people's actions in terms of how they serve several various purposes in a particular time, place, and audience:

> Society is organized on the principle that any individual who possesses certain social characteristics has a moral right to expect that others will value and treat him in a correspondingly appropriate way. [...] when an individual projects a definition of the situation and thereby makes an implicit or explicit claim to be a person of a particular kind, he automatically exerts a moral demand upon the others, obliging them to value and treat him in the manner that persons of his kind have a right to expect. [141]

This adherence to social obligations sets up a web of expectations, about how one is treated and treats others, that allows a "constancy, and patterning of behaviour" [139] which supports a suitable and just society. Within this, two concepts are particularly relevant to understanding respect:

Deference "functions as a symbolic means by which appreciation is regularly conveyed" [139]. Expressing deference towards a person is a means of conveying a "sentiment of regard" to them, which could consist of *ceremonial sign-tokens or actions*, lacking any other purpose besides communicating regard or deference in their own right. Alongside these presentation rituals, deference can also take the form of avoidance of things such as another's personal name, physical space, or embarrassing topics of conversation.

Demeanour covers the ceremonial behaviours by which one expresses desirable – or undesirable – qualities. Others observing these behaviours would then be obligated to give them a certain regard based on their understanding of how one's demeanour relates to personal qualities.

The combination of deference and demeanour are about establishing and conveying one's regard for others on the one hand, and communicating the regard one expects of others on the other. Like many other rules of conduct, deference may be symmetric or asymmetric, particularly when different social roles are in play. While deference often follows asymmetric hierarchies of power, Goffman gives the example of hospital staff, where doctors, nurses, and specialists should all defer to each other's expertise in their particular area. This connects to respect not just with regard to the idea of respecting one's particular authority or role, but that the appropriate and necessary ways of respecting may be asymmetric. Demeanour is then part of how a person handles their place within this network. Rather than purely carrying out behaviours ascribed to the role, one projects a complementary image that acts as"a kind of justification and compensation for the image [...] others are obliged to express through their deference" [139].

Unlike the previous conceptualisations of respect discussed in this chapter, Dramaturgy is entirely descriptive. While useful in understanding the social mechanisms through which people respect each other and communicate how they wish to be treated (similar to Hudson's conceptualisation of respect for the social institutions that make up modern society), there are clearly many deeply undesirable social systems that are described even within Goffman's own work (such as patriarchal ideas about gender roles in the home).

In contrast to this, the conceptualisation of respect proposed by Schrimer [298] is grounded in agency, the capacity of people to act freely within and upon their environments as they please, to achieve their goals, pleasures, and so forth. Contrary to the Kantian view of respect as something that all have by nature of being human, constructivist views ground respect in *ascribed agency*. That is "showing respect towards other persons is suggested to mean the symbolic, communicative act of giving their agency the elbowroom they need in order to control their environment in a skillful, knowledgeable and independent manner [298].

Such a view acknowledges that there may be many factors that shape a person's actions that are out of their immediate control, whether psychological, social, cultural, environmental or physical. On hearing an insulting remark from a person with a different cultural background, one may consider whether it is a cultural mishap, and hence not attributed to the speaker's agency, or whether to ascribe agency and take the insult as an intended slight.

Ascribed agency can also be a means of differentiating the interpretation of respectful behaviour from considerate or polite behaviour. In particular, the extent to which a polite or considerate action confirms or constrains a person's agency is said to define whether it is also respectful [298]; offering an older adult a seat on public transport may be considerate, but could also be taken as disrespectful as it implies a lack of capability and agency. Similarly,

guiding a visually impaired person without asking can be seen as considerate in supporting their needs, but shows a lack of respect through diminishing their agency in the situation.

7.6 Complex Ecologies of Respect

This notion that respect is something ascribed and situational can be developed through ways of looking at the world that include the agency not just of humans but a broader range of entities. Within the broad area of new materialism [72, 8], agential realism sees concepts—such as respect—existing through "specific agential intra-actions that the boundaries and properties of the 'components' of phenomena become determinate and that particular embodied concepts become meaningful" [21]. In this sense, agency is not a purely human capacity, but an ongoing production of the unfolding material and social world, and "is an enactment, not something that someone or something has." [21]. Here respect is an expressive performance: "To convey respect entails finding the words, the gestures, and the layout of the physical space that makes respect felt and persuasive" that situates it as something individual and contextual [180]. This is similar to agential realism's treatment of dignity, a subject closely related to respect, as performative: considering dignity as a product of both a person's actions and their treatment by other people and social structures [137].

This raises two important points for an understanding of respect in the context of interactions between humans and technological systems. Firstly, it suggests that respect can be enacted between a variety of human and non-human things rather than solely between humans as moral agents. This is vital for analysing human-computer interactions, as we find ourselves dealing with a range of devices, systems, and networks that have varying degrees and appearances of agency, allowing for different senses of respect and respectful behaviour. Just as practices such as hostile architecture [276] can manifest a lack of respect for certain sections of the population, interfaces that do not account for capabilities, desires or modes of address can perform disrespect towards users just as strongly as a human can.

Secondly, we can work with the idea that different configurations will give rise to a different observations of respect. This follows a sense from Suchman [318] that a human-computer interface is not 'an a priori or self-evident boundary between bodies and machines [but] a relation enacted in particular settings and one, moreover, that shifts over time'. Agential realism uses the notion of *agential cuts* to work with this idea that the entities under discussion are not fixed, but rather we need to draw a boundary and look at what happens across this boundary. In placing these cuts, Shotter writes that "we do not uncover pre-existing facts about independently existing things; we ourselves

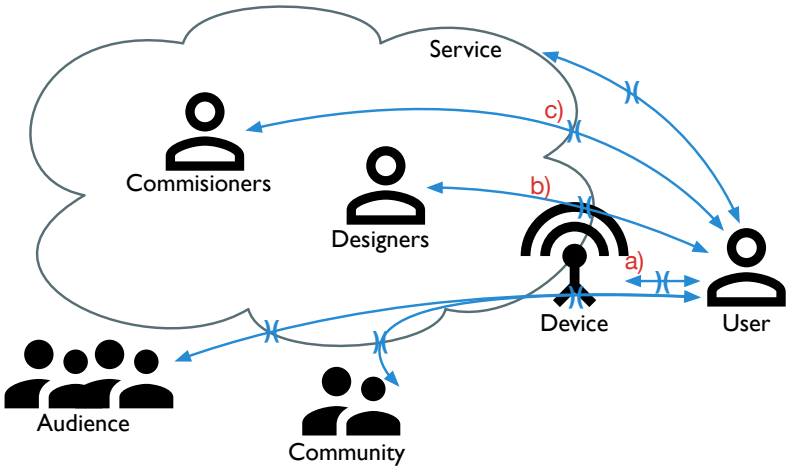

Figure 7.1: Configurations of respect. Each line represents a conceptualised interaction, with a particular agential cut marked. There is space for thinking about a) how a device performs respect to the user and vice versa; b) how designers' respect or lack thereof comes through to the end user in interactions; c) how an organisation's treatment of a person shows their objective respect and so on.

bring such 'things' into existence" [305]. When working with smart devices there are many places where these lines could be drawn. One could separate the user from the device and all the services behind it in order to look at how a smart speaker performs respectful interaction; another cut would look at how two users interact through the system and develop notions of computationally mediated respect and so on. Each of these cuts can give rise to a different picture of respectful interaction.

Developing this idea further produces new ways of analysing the design of new and existing smart home devices. It becomes apparent that there are multiple subject-object configurations of respect which become possible (see Figure 7.1). Here we highlight and discuss three such configurations: firstly and perhaps most obviously, how systems might respect people; secondly, how people might respect each other through a mediating system; and thirdly how and why people might themselves show respect to systems.

Systems Respecting People

When it comes to considering systems as the subjects of respect (for human objects), some might naturally turn to philosophical discussions of artificial intelligence, considering whether intelligent systems are or can be *truly* intelligent and have agency (and therefore have concomitant rights and be subjects of moral blame and so on) [41, 52, 69, 148]. Regardless of one's position on

whether interactive systems could ever possess equivalent moral capacity to human subjects of respect, they *are* capable of embodying (dis)respect towards people as a result of the actions (or inactions) of their designers, and they *do* have politics. A useful touchpoint here is Winner's look at the work of Robert Moses, "*Do artifacts have politics?*" [353], in particular the 'racist bridges' that did not afford enough clearance for buses, hence restricting parts of the city to only those wealthy enough to own cars. On a purely physical level, the bridge is relatively inert, and hard to see as a political actor. However, it is also very clearly the result of a collection of political decisions designed to entrench certain biases and hierarchies.

One of the key ideas behind agential realism is the idea that we construct concepts when choosing where to draw the boundaries between parts of the world. So rather than asking whether the bridge is racist, we can ask whether the bridge and the systems that created it are performing racism towards a particular group of people. This tends to fit with the glosses made when talking about intelligent networked systems: asking about whether Alexa is sexist [226] is asking a question that includes the moral agency of the designers and programmers, the data and training sets as well as the the way that it is advertised and situated as a device to have within a home. At the same time, increasing intelligence and interactivity in computational systems supports a view that they have agency: they respond, shape, intervene, and choose. All this can be granted without necessarily having to discuss the prospect of giving them rights or personhood (often a distraction from the more pressing needs of those whose welfare they harm [41]).

In other situations, respectful interfaces can approximate face work [140] *by* machines, such as affective computing studies where interfaces help users process negative emotional states caused by the system itself [188]. Doing so might improve user experience at the cost of misrepresenting the truth (e.g. the system is not genuinely sorry that it has caused frustration)—but perhaps this kind of approach could help people to build more accurate mental models of interactive systems by countering assumptions that systems are neutral and infallible.

Computer-Mediated Respect: People Respecting Each Other *Through* Systems

As well as respecting people through the way they design interactive systems, designers can also mediate the ways in which users can be respectful (or not) towards each other *through* the system. Software such as Gmail Smart Compose[1] that create automatic replies to messages, for example, intervenes in

[1] `https://www.blog.google/products/gmail/subject-write-emails-faster-smart-compose-gmail`

personal communications. Unpicking this example surfaces several moments where respect may or may not be shown:

- The system, or designer, attempts to respect the user's time, by taking the hard work out of responding in predictable ways (deference);
- offering canned choices undermines the user's personal choice and autonomy by limiting their space to skilfully construct a response to the recipient (ascribed agency);
- the system attempts to help the user show respect for others, by allowing them to respond quickly and indicate attentiveness (deference, social rules), although the success of this is questionable;
- the system implicitly frames the task of replying as being a burden (obstacle), rather than a situated response to the actual person on the other end (care respect);
- the system attempts to match writing styles (social rules, belongingness), although always at the risk of errors, and may go as far as following specific norms (institutional).

All of these connections are axes where looking at the interaction design in terms of respect has the potential to prevent mistakes and uncover undesirable effects before they happen. Having at hand an array of different versions of respect helps to make sense of complex interactions between combinations of humans, groups, technologies, infrastructures and objects. This extends the idea above that systems can perform facework to users, suggesting as well that systems can perform facework *on people's behalf*. Beyond examples in messaging, we can imagine the considerable benefits and risks of future systems that mediate facework across language and cultural divides [324] where misunderstandings and similar failures are much more likely. When systems engage with the wide variation in values and practices between cultures, respect may offer a lens which can help motivate a more involved localisation process; instead of considering isolated practices and behaviours, conceptualisations of respect may offer insight for navigating the wider landscape of acceptable and unacceptable social interactions.

People Respecting Systems

Finally, we might also consider the configuration of people paying systems respect. Even if, as discussed above, we reject the idea of systems as moral agents who *deserve* respect, users might nevertheless express (dis)respect for systems, rightly or wrongly, and systems might be designed to make this more common, presenting both dangers and opportunities for design.

Certain kinds of respect apply naturally to systems as objects of respect—just as one might have *respekt* for a wild animal, or treat industrial machinery and immigration systems with caution, one often learns a sense of obstacle respect when solving the problems of getting software and hardware to support one's goals. Attitudes like this can help convey the presence of potentially unknown harms or dangerous aspects of systems (e.g. neural network classifiers with unknown rules) [302]. A dramatically different scenario are toys and robots designed to be the recipients of care respect, such as the Tamagotchi[2] or Paro [295], which need to be looked after for entertainment or therapeutic effects. The ethics of introducing inanimate objects designed to solicit nuturing, caring behaviours, and human emotional connections has been the subject of much ethical debate, due to such actions being seen as exploitative [350, 319] and harmful, due to machines not being able to genuinely have feelings in return.

These scenarios can be understood through early work by Reeves and Nass on the paradigm of *computers are social actors*, which demonstrated how people treat interactive systems as if they were people even though they know that they are not [255, 289]. The 'magic word' please-and-thank-you mode added to Amazon Alexa appears to be an extension of this, where children are asked to perform *deferential* respect towards a smart speaker. Work is ongoing about whether users personify these devices (e.g. purington2017alexa), so there remains the question of whether users are performing respect towards the physical object, its imagined personality, the organisation behind the assistant, or some combination of these. These present designers with opportunities for manipulation: inculcating emotional attachments between users and systems, and exploiting the deference that naturally arises; as well as for education: if a voice agent can detect 'she' is being disrespected by a misogynistic user, 'she' can answer back to challenge it.

7.7 Respect in Practice

Just as the conceptualisations of respect in the previous sections lead towards greater detail in understanding how people think of respect, from high ideas and ideals to personal and then social-historic embeddings, they also lead to questions about how respect is carried out in practice, and how it is enacted within specific contexts.

Empirical Examinations of Respect: Perspectives from Psychology

Interest in understanding respect in the psychology literature has peaked significantly in the past 30 years, which Langdon attributes to the positive psychology movement [200]. Multiple empirical studies have attempted to

[2] `https://en.wikipedia.org/wiki/Tamagotchi`

understand respect using different frameworks. McDowell [234], ran studies of young working-class males in the UK to examine the role of two conceptualisations of respect in the context of the politicisation of the behaviour of young people. *Respectability* was concerned with external perceptions of one's own socioeconomic status, while *deference* was concerned with the performances that might indicate personal respect but could also include compliance with authority or a utilitarian calculation of benefit. These two concepts were chosen because they were seen to articulate both how these boys were seen by others – including perception by out-group members and a sort of respect that they were born into – and the respect which was *earned* from friends, family and direct connections.

Group related conceptualisations of respect have been a primary focus of studies of respect. De Cramer and Tyler [89, 90], among others, focus on aspects of respect as *belonging* to particular groups, hypothesising that the need to feel like one belongs explains why respect is so important and desired by individuals. Related to this construct have been studies of how perceived *belongingness* within groups are strongly supported by in-group treatment, finding that such treatment bolsters the individual's group affinity and collective identification [306], resulting in an increased likelihood of members taking actions to benefit the group, as well as measurable effects on an individual's self-esteem [159].

In response to the many and varied ways such studies conceptualised respect, Langdon conducted a comprehensive review [200] of psychological experiments relating to respect, synthesising four themes. The first, *respect as status* conceptualises it as *that which people with authority, or higher social status deserve*, such as acting with obedience, and yielding to authority. The second, respect as social rules, is closest to Goffman's theory, and frames respect as rules establishing morally-prescribed, normative behaviours. She notes that there is discrepancy among the literature, however, in what precisely these rules entail, whether it is to treat people with politeness and dignity, friendliness and consideration, and so forth. The third frames respect as *care*, comprising the feeling or expression of love, and concern for others. The final is respect as *equality and accepting differences* of others, including awareness and acknowledgement of others' perspectives. Hendinger [152] also provided an alternative theoretically-informed view incorporating four similar concepts: 1) deference to power, 2) deference to the rights of others, 3) honor/appreciation of the characteristics of others, d) trust/confidence in others.

Several studies have used methods of direct elicitation to ask people what respect means to them, using prompts (sentence completion and reflection) and vignettes. Langdon and Preble [201]'s study of US high schoolers revealed examples of respect-as-care with "Respect means .. when you need help they help you" and respect as accepting differences "Respect to me

means letting others have their beliefs." [201]. Categorising reflections from a larger population, responses fell into 14 categories, the most frequent 8 of which were *having good qualities*, *politeness*, *relating/understanding*, *caring*, *state of mind/reverence*, *accepting differences*, *power/obeying*, and *reciprocity* [201]. Schirmer et. al [298] explored their concept of respect as ascribed agency using field interviews in Sweden to uncover prototypical respectful behaviours, finding 'listening to others' and 'not interfering with others' the most common examples. Huo [158] investigated how respect plays out in groups, finding that a sense of *fairness* and *competence* led to respect for a leader, and that concerns about respect stem from a need for social inclusion and attaining status.

7.8 Translating Respect into Design Principles

As separate but related concepts, the conceptualisations of respect discussed above are not always concordant. Obstacle respect obviously conflicts with respect for persons as the other side of the ends-versus-means dichotomy and care respect is also a source of many interesting conflicts where promoting another's well-being requires a different course of action to their expressed wishes or social expectations. At a deeper level the philosophical bases underpinning many of the different conceptualisations are fundamentally at odds and appear to be irreconcilable; some, like Kant, are heavily normative and rooted in personhood or moral capacity, while others like Goffman are descriptive and see respect as behaviour that arises through interaction.

But this is not necessarily a problem for the use of respect in the design of smart devices. Firstly, the different conceptualisations of respect chosen here each represent tools to be used *in context* to guide the design process. While care respect might be more suited to the creation of a voice assistant or device intended to support ageing in place, a tool like Taylor et al.'s 'Homenote' might instead benefit from analysis based on the idea of respect as a complex ecology of behaviours that arise between the people who use it (who might conceivably be any number of configurations of family, friends, or acquaintances). In each case it is the task of designers to identify which perspectives are most relevant to the device at hand.

Indeed, when these perspectives clash it highlights what is at stake, and the nature of the moral and ethical choices that are being made. One example of how this could manifest is in devices designed to support ageing in place that might impinge on users' autonomy and privacy [91]. For example, consider a smart thermostat that prevents the user from setting the temperature dangerously low (perhaps in an effort to save money). Alternatively, in an attempt to be caring, naive approaches at supporting long-term safety might directly undermine the subject's privacy (such as by allowing close relatives

Table 7.1: Summary of key perspectives on respect.

Kant	Categorical Imperative	Treating people genuinely, as *ends* in of themselves rather than as a *means* of accomplishing selfish objectives; recognising "the status, worth, and individuality of others simply for being human, regardless of their place, position, or situation" [178].
Feinberg	Respekt	Care and attention shown to entities with power, such as avoiding a bear when walking in the woods or respecting a powerful opponent.
	Observantia	Acting in accordance with legitimate moral claims, such as paying what is owed or using the appropriate honorifics for a judge.
	Reverantia	A combination of respekt and observantia that is "the strongest force of all, where power and legitimacy blend" [113]. For Kant this was majestic Law and the divine.
Hudson	Directive	Modifying behaviour to recognise the authority of a stated preference, law, rule, or principle.
	Obstacle	Adapting behaviour to manage something that is an obstacle to one's own goals.
	Institutional	Following the practices, customs, and hierarchies of society and its institutions, regardless of one's opinion of the people within them.
	Evaluative	Responding to how well people and objects meet certain standards.
Dillon	Care Respect	Recognising people for the individuals they are, responding to their needs, and promoting their well-being.
Mills	Black Radical Kantianism	Acknowledging the ideological and psychological legacy of racism and the habits of disrespect that affect respect for others and self-respect.
Goffman	Deference	Recognising another's place, role, status, and position; conveyance of appreciation or devotion, facework, or taking action to save others' reputation or embarrassment.
	Demeanour	Communication (through behaviour and presentation) of how one wishes to be seen and treated, and giving that treatment to others.
McDowell	Respectability	External perceptions of socioeconomic status, supported by certain kinds of behaviour.
	Deference	Actions that signal deference for others, whether based on compliance with authority or personal respect
de Cramer & Tyler	Belongingness	Respectful treatment by group members increases group affinity and collective identification.
Langdon	Status	Something deserved by people with authority or high social status.
	Social rules	Following the correct social rules.
	Care Respect	Feeling or expressing love and concern for others.
	Equality	Awareness and acknowledgement of others' perspectives.
Schrimer	Ascribed agency	Giving people the space they need to skillfully be themselves.

to monitor and be informed of their activities), while systems that attempt to influence behaviour (such as to increase medication compliance, encourage physical activity, or alter diet) could reduce their autonomy. In each case, we can see how being caring conflicts with another aspect of respect and use this to determine where the balance should lie.

Another salient example of such a conflict arises from automatic software updates. Update mechanisms present on some modern devices embody a form of care respect, automatically delivering security patches and improvements that support the long-term security needs of the subject. In contrast, software update mechanisms that combat user reluctance towards installing updates with obstacle respect—as a barrier to be overcome—might repeatedly interrupt users to obtain consent to install updates, or refuse to allow continued use of the device or software until updates are installed. These examples highlight how important it is that designers of caring systems balance the effect of supporting care respect with the other kinds of respect that such actions might undermine.

It is clear from the above that respect encompasses a wide range of concepts, underpinned by an intuitive set of values that help elucidate the *why* behind many existing design considerations. By cutting across existing design perspectives and ethical principles such as UCD, VSD, accessibility, and fairness, respect is also well-placed to help designers evaluate what values to target from the myriad of possible options. Overall this highlights the value of respect as a guiding principle, helping designers to work within existing frameworks rather than defining one itself.

Other observations that are apparent from our discussion of respect thus far is that (1) as with people, no system will be completely moral or respectful, and this should not act as a barrier to the furtherance of ethical design; and (2) while respect plays a key role in many understandings of moral behaviour, behaving respectfully is not always equivalent to behaving morally. Aside from the obvious possibilities with obstacle respect, dramaturgy and institutional respect can support immoral rules and hierarchies by the virtue of them being established social and cultural intuitions (although this is not a problem unique to respect—many systems are easy to use but harmful, and there are many values that it would be immoral to support as part of approaches such as VSD).

Whilst these conceptualisations of respect in this chapter frame a rich and varied set of understandings that can be applied in context, they represent an intimidating starting point. In this section, we distil an essential set of four principles for design that each are informed from several of the conceptualisations we feel are most relevant to the design of interactive systems. While they do not represent a universal theory of respect for smart device design

(far from it), the intention is that they provide a more accessible entry point from which to anchor further enquiry.

Avoid Treating People as *Problems*

Despite considerable variation in the conceptualisations of respect, one theme that was shared across almost all conceptualisations was treating the individual's needs, desires, and preferences genuinely, rather than as something to be defeated or overcome. Thus, a simple principle we propose is to *avoid treating people as problems*; a principle that is widely (if not universally) applicable in the design of artefacts, and yet often violated. One example is hostile architecture, where designs are centred around anti-affordances: solutions to the problem that people might take certain actions. Another grave violation of respect when such treatment is targeted based on race or demography, such as the stop-and-search of black americans by police [208] or treatment of hispanic Americans by US Immigration and Customs Enforcement officers under the Trump administration [153]. Softer versions of this are also digitally pervasive, including apps and systems that treat people as criminals or bad actors by default, such as by warning people against illegal activities before granting them access to resources or capabilities. Under the threat of sophisticated attackers, emerging authentication systems that force users to prove that they are to be trusted, by demonstrating that people are whom they claim to be and not an imposter, might be seen as disrespectful in this same way—by assuming they are an imposter until proven trustworthy. CAPTCHA systems that force people to prove they are human can be seen as disrespectful by assuming someone is a robot a priori, and also because they waste the user's time and effort.

Another way that designers often treat people as problems today is that they see users as obstacles to be overcome. This goes against most conceptions of respect, with the exception of *obstacle* respect. Examples of systems that badger, coax, trick, deceive, or nudge users to take specific actions they might not want to take abound in modern digital systems, often making use of interface elements known as 'dark patterns'[3] to reduce user agency. Pushy software update mechanisms, such as those built into Windows, have been known to use tricks to force people to install urgent updates. Other examples include the many kinds of dark patterns used in e-commerce websites [231] and by malware makers to trick people to do things. Such examples also illustrate disrespect in a Kantian sense as well, by not treating users' needs as genuine, but as a means of furthering others' goals.

[3]See `https://darkpatterns.org` for examples and more information.

Don't Speak for People—Signalling, Self-Representation and Identity

A key aspect of micro-social accounts of respect is the communicative role that respect plays, both in the expression of the quality or importance of one's relationship to others, but also in signalling to others how they expect to be treated in return (demeanour). Such signalling is used to establish or reinforce one's position, status, or identity, but also to serve as a channel of continuous feedback to others about how they feel they are being treated, including communicating when they feel disrespected.

As digital systems start to become able to recognise these social cues automatically (such as through advances in affective computing [340]), one might argue that an ultimate aim of future respectful systems might be for them to be able to naturally recognise and respond to such signals. Here, dramaturgical accounts of respect may be most useful to inform what such responses should entail, such as understanding when user feels disrespected or treated inappropriately by the system.

Even before systems become capable of such social cue recognition, there may be immediately applicable implications about ways that systems should enable people to craft their identities and represent themselves online. Acknowledging the power of designers to 'make and unmake' others through their actions [94], the methods by which designers enable people to represent themselves impacts people's ability to truly be themselves online. These features also allow people to express their identity/association with groups and signal their status, roles, and identities in flexible and expressive ways that may be important in moderating how they are treated by others.

Treat People as People, Not Their Properties

It is now common for many kinds of adaptive, data-driven systems to classify and differentiate people based on their properties, such as their demographics, preferences, activities, and so on. Yet we have found multiple ways this could be seen as disrespectful; one, because it suggests in some sense that people are defined by, or are similar because of, such attributes, rather than reflecting on the ways people are unique and individual. Embracing individuality may mean systems should be sensitive to the individual experiences people have, due to there being no singular account of what it means to be a member of any category, such as being black, a woman, an older adult, or to experience hearing loss or a loss of mobility. Perhaps connecting people based on commonality of experiences might thus be an alternative approach.

The second is that such a categorisation overlooks the reasons that people might have the same experiences, such as the personal histories, oppressive structures, traditions, or environments that constrain or shape them. Thus, a

supportive and respective mediating system has the responsibility of considering these factors, structures, and histories when determining the appropriate ways they should respond.

A focus on individuals would also help avoid design over-generalisations that end up erasing or neglecting swaths of the population that fall out of the "main demographics" focused on by designers. Such erasures have been documented across the literature, such as in the design of fitness trackers—which establish normative ontologies generalised from young people with specific body types and abilities [310, 311]—and menstruation trackers designed around the goals and needs of a narrow set of post-pubescent, pre-menopause cis-women [106]. For those who have menstruation tracking needs, but fall outside this range, this reinforces the impression that society does not accept or prioritise them. The problem here is not just that the needs of certain kinds of users may not have been served, but also that the system communicates disrespect towards them by failing to do so.

Another aspect of this relates to the fluid nature of people's preferences, beliefs, and opinions. Those shared between cultures and communities often change more slowly in response to an external stimuli, and require a degree of consensus. Individuals have rapidly changing preferences, make arbitrary decisions, and hold conflicting beliefs; there is no single truth to the human condition, and the more technology is designed to deal with this the better.

Support Agency and Choice, Instead of Constraining It

It has been a perpetual temptation (and tendency) for technologists to apply advancements to automating everyday life, and to 'optimising' things for people so that they don't have to–our lives have become incredibly automated today, compared to just 10 years ago. However, respect gives us the idea of ascribed agency to apply as a critical lens to such automation, enabling us to differentiate respectful automation which is *agency supportive*, from disrespectful automation which is *agency constraining*.

Rather unsurprisingly, the world is full of countless examples of disrespectful, agency-constraining automation. News feed algorithms of social media platforms are one such example; they deny readers the agency of choosing what to see or look at next and render people into passive media consumers with no control. Often, such disrespectful automation comes with the best of intentions; even the news feed algorithm aims to provide 'optimally engaging content'; however, being agency denying means it deprives people from being able to specify what they really want to see. Other examples are accessibility technologies that try to automatically detect when someone has mobility or other impairments, such as the 'sticky keys' functionality in Windows. Pressing shift five times triggers the operating system to suggest enabling a feature that implies a disability (Windows describes the functionality as being for

those 'having limited reach or strength'). This is also important as people may or may not identify with the groups for whom accessibility features were designed—someone who can't see as well as they used to might not identify as partially sighted but might still wish to use larger or heavier fonts, and may choose not to do so if these are labelled as 'for the blind and partially sighted'.

Another emerging case where systems can constrain agency is where software mediates communication between people. Systems like Google Duplex[4], for example, restrict the range of possible questions and responses compared to a phone call made directly by the user. Ironically, in many cases this removes the opportunity to 'have a conversation' about individual circumstances (e.g. how many people with severe nut allergies would trust Duplex to order food for them?).

An interesting aspect of giving people agency is the potential conflict with care. In order to promote the flourishing of others we sometimes need to tell them things that they do not want to hear, or act in ways that directly contradict their immediate wishes (such as highlighting red flags in a relationship or encouraging someone to continue meeting their fitness goals). It may be difficult to imagine technology exhibiting this kind of care respect towards users, but existing systems already do. Virtual assistants, for example, will give contact information for suicide prevention and mental health phone lines rather than answer users' requests for content that may lead to them harming themselves.

7.9 Conclusion

The notion of respect is a broad church home to a wide and varied collection of concepts. In this chapter we have attempted to give an overview of the kinds of respect that are most relevant to the design of interactive systems, from the groundwork laid by Kant through to Goffman's sociological perspective through to more recent work in feminist philosophy.

This conceptual work has shown how different understandings of respect can assist designers in navigating the complex design space around notions such as identity, autonomy, and accessibility. We highlight how designers can work towards the creation of more respectful systems by avoiding treating people as problems, rather than worthwhile ends in themselves; allowing people to speak for themselves, reducing the extent to which we allow unconscious societal and personal biases to misrepresent people; treating people as individuals rather than masses, promoting a more nuanced and personal understanding of the people who use our technology; and giving people space

[4] https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html

to have agency and choice, putting users and their decisions at the heart of interactive systems.

Respect is clearly not a new concept, but is not well established within HCI. By surfacing respect and giving it a more robust treatment, we aim to support the ethical design of interactive systems using insights from fields outside of computer science. In the future we hope that respect can be further integrated into the design process as a lens to support existing design frameworks. There are several key areas of future work in which this could (and should) occur which will be discussed further in the next chapter. By using the language and norms encapsulated by respect we attempt to begin a radical shift in the relationships that people have with their technology, moving away from smart devices as morally neutral tools towards agents situated in complex webs of people, identities, roles, and relationships.

Chapter 8

Discussion and Future Work

"It feels like this legitimate business model has gone completely out of control and created a kind of chaotic industry that is not understood by the people who are most affected by it."

— Prof. Reuben Binns, 2018 *Financial Times* interview

Contents

8.1 Re-Thinking Smartness: If Smartness is a Problem, What is the Solution?

This thesis began with the intention of understanding smartness in order to mitigate some of the ethical concerns that arise when sharing our homes with connected, intelligent, sensing devices. In this chapter we discuss what these results might mean in terms of re-thinking smartness and the future of

A version of content in this chapter has previously been published as: "William Seymour and Max Van Kleek. 2020. *Does Siri Have a Soul? Exploring Voice Assistants Through Shinto Design Fictions*. In Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems (CHI EA '20). Association for Computing Machinery, New York, NY, USA. DOI: `https://doi.org/10.1145/3334480.3381809`"

ethical smart home devices, as well as the implications for people's ongoing relationships with smart technologies.

8.2 Respect

The culmination of the research presented in this thesis was an analysis of the ethics of respect and its potential application to the design of smart devices (RQ4). Bringing together the findings on ethical concerns with connected technology in the home described above, Chapter 7 focused on conceptual contributions, asking how the concept of respect could help us understand and design for ethical concerns with connected devices. To this end, it highlighted conceptualisations of respect from different academic disciplines in an attempt to cover the many different ways that people and devices interact, relating formulations such as the categorical imperative, dramaturgy, and feminist care respect to established HCI design frameworks and principles. It explored the different configurations that respect might take between devices, their creators, and users, discussing how designers should: (1) avoid treating users as problems; (2) refrain from speaking for people; (3) treat people as people rather than their properties; and (4) support agency and choice. This in turn supports the design of future ethical smart and connected devices.

Looking back at the Aretha technology probe, we can see that it adheres to these ethical principles by giving people the data and knowledge to make their own decisions, and giving them options rather than treating them as a problem. It therefore serves as an early prototype of a respectful system. But while respect might be enacted through functional solutions that operate at the level of facts and preferences, the theoretical treatment of respect in this thesis seeks to evaluate and address these through the development and refinement of norms. This is important because, while a good start, technical measures alone are unable to meaningfully shape smartness's future direction and development in the face of increasingly social devices; the setting of expectations and baseline standards is needed in order to influence the power and politics of smartness as it technology continues to develop.

Navigating Delegation and Learning

As identified in the analysis of the interview data in Chapter 4, learning by and delegation to smart devices are common sources of concern over control. Smart devices are often given wide latitude with which to carry out the tasks assigned to them; where previously devices would take specific instructions, such as heating the house at specific times or playing particular songs, their smart equivalents are expected to determine themselves how best to fulfil open ended requests, such as tasking a Nest thermostat to heat the house comfortably or asking Alexa to "play some music". These requests require

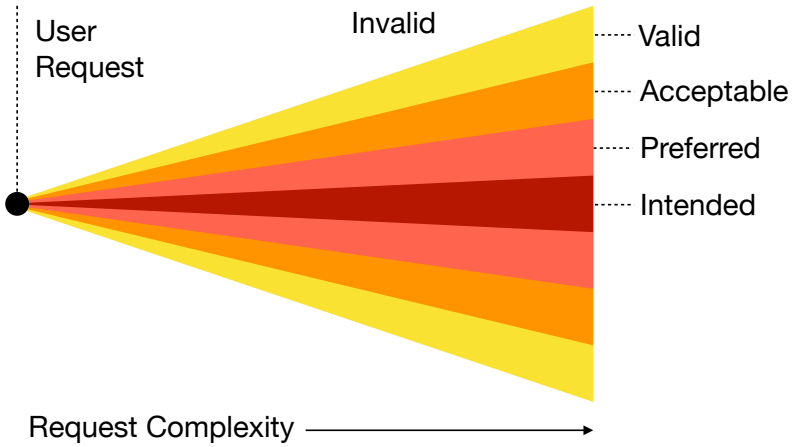

Figure 8.1: The sets of possible responses to underdetermined user requests. As requests become more complex, the distance between these sets grows.

interpretive capability to resolve 'unspoken' (or *illocutionary*) assumptions from what was explicitly said. In some cases, resolving what a person wants also involves learning the user's context and preferences, something that is now within the reach of modern devices and platforms.

This is understandable given that the set of possible valid responses to underdetermined requests can be huge (potentially even infinite). However, the set of acceptable responses, constituting cases where the behaviour yields the correct intent and fulfilment in a way that is acceptable to the user, is typically much smaller (Figure 8.1). Smaller still is the set of *preferred* responses that take the individual's particular contextual preferences into account. When asking Alexa to "play some music", purchasing the 2008 film *Music* and streaming it to the nearest TV would constitute a response that is technically correct but fails to identify the illocutionary assumptions of the request. Playing Mr Blobby's self-titled Christmas single would be an 'acceptable' one, capturing the user's intent (it is, technically, music). Finally, *Shake it Off* might represent a preferred choice that solves the user's need.

The notions of respect explored in Chapter 7 are useful tools in developing these categories. We can use different conceptualisations of respect to analyse how a system should respond to these kinds of ambiguous requests. Consider if Alexa selected a random song (directive respect), one it predicted would be enjoyed the most by the requester (care respect), one that Amazon did not have to pay royalties on (obstacle respect), or one that was widely considered a musical masterpiece (akin to evaluative respect).

Beyond simple examples like music choices, we can also imagine how these systems could promote longer term care, learning people's preferences and

adapting to how they live their lives. While it is often considered undesirable for external algorithms to decide the content and purchasing options presented to people, taking control of these filters could *empower* people to achieve their goals. Inspired by the way that the Aretha probe helped study participants to form and enact preferences around data flows in the home, devices such as voice assistants could help people to specify wider lifestyle goals (e.g. making fewer impulse purchases or ordering healthier food), and help them achieve this by prioritising that information over alternatives. The vocabulary of ascribed agency helps when navigating the fine line between empowering people without being overly controlling; the visibility and delivery mechanism of these features would be key in ensuring that they are helping people to make decisions (agency affirming) without being overly controlling (agency constraining).

Designing for Risky or Unknown Situations

Unfortunately there will always be situations where the interests of powers such as governments and large corporations, or otherwise hazardous situations make empowerment in the manor described above a less appropriate option. When a device and/or the services behind it are capable of causing significant negative effects on somebody, we need to find ways of designing these experiences to be respectful whilst also retaining appropriate consideration of potential negative outcomes. In this context, Feinberg's concept of *respekt* comes to mind as an appropriate response; as devices increasingly have influence over our health, security, finances, and enjoyment, it might be prudent to maintain "an element of fear" when considering the capitalist motivations of the companies that control them. Smart devices differ from other conventional electronics in that while one purchases a physical device, the real value is in the software or the *presence* that inhabits the plastic and silicon. In the case of voice assistants, these agents also appear to move seamlessly between devices and are often designed to speak and act like humans to the extent of engendering feelings of social presence [63].

To this end, Chapter 6 explored concerns around how voice as an interaction modality influences how we interact with voice assistants and how it 'feels' to share a social space with them. Applying work from the human sciences on interpersonal relationships to HCI, it tested the hypothesis that interactions with voice assistants can be characterised by the notion of 'relationship development' in much the same way as for interpersonal relationships. This was confirmed to be the case, further informing the ethical considerations that must be made when designing new voice-enabled technologies, suggesting that devices might be more deeply embedded with social reasoning than previously thought. Statistical tests showed that higher levels of relationship development with devices was correlated with increased user trust and an-

thropomorphism. Interestingly, this increased trust in devices coincided with increased trust in the voice assistant's manufacturer, suggesting an awareness of their close connection.

Referencing earlier work on the 'purification' of mental categories, examination of the distribution of relationship development levels amongst participants suggests that people may have already completed purification work around the human/robotic nature of voice assistants. Further work will be required to more fully understand what the relevant categories are at play with voice assistants and how users perceive them, but the results collected suggest that voice assistants have a somewhat settled status in the collective understanding of people in the UK.

In order to explore the implications of this further, we now present and discuss a short design fiction that depicts an alternative conceptual framing of voice assistants, exploring what it might mean for us to interact with this entire class of disembodied, partially autonomous social agents that present themselves as human-like but subservient, and have a high degree of influence over our personal circumstances. In doing so we highlight tensions between being respectful, interpreting open ended requests, and anthropomorphism. Japanese narratives "routinely make spirits, robots and animals cohabit in the world in ways that ignore boundaries between human and extra-human realms" [171]. These animistic messages extend to spiritual practices, and are "at ease with mixing advanced technologies and spiritual capacities". Such a context provides an interesting position from which to reconsider many of the key assumptions and design decisions that underpin contemporary voice assistants.

Shinto emphasises the relationship between the spiritual and the physical, and how neither are able to exist without the other. The Shinto term for vital power is typically '*tama*', '*mi*', or '*mono*', with the presence itself being '*kami*' [179]. This is central to Shinto beliefs, which present a "holographic model according to which 'spiritual forces' (or kami) are literally in everything" [171]. Indeed, the world and kami are so interconnected as to be incomplete without each other. Kami inhabit all things; they can be landscapes, forces of nature, or even the venerated dead. Traditionally, kami have two aspects, minds, or 'souls'. One of these souls is mild and caring, the harmonious soul (*Nigimitama*), and the other is rough and violent, the wild soul (*Aramitama*). As a result kami can nurture and love when in harmony, but similarly are capable of spreading discord and destruction when disregarded.

Inspired by these beliefs, what follows is a design fiction imagining voice assistants as hosts for different presences concerned with human beings. This highlights the potential asymmetries between people and assistants: like kami, the devices in this fiction (mostly) want us to be happy—if they are

treated properly they can bring numerous benefits, but if disregarded they can similarly make their displeasure known.

The Harmonious Soul

Congratulations on your choice to purchase a new *presence enhanced*™ virtual assistant! Your new assistant is sure to bring harmony to your home, and if treated with respect will continue to promote peace and well-being for years to come. Each presence has a unique character and, like a new friend, it is important that you get to know each other properly in order to derive the most fulfilment from your relationship. Upon unboxing, the first thing that you are likely to notice about your virtual assistant is the material. Carefully chosen to represent the nature of the presence that inhabits it, the base unit might be a rough block of wood, a smooth ball of plastic, or an elegant sculpture of brushed stainless steel.

When everything is in balance, and the proper respects have been paid to your assistant presence, you will get to meet its kind, warm, and functional side: the harmonious soul. This aspect of your assistant is keen to help, gradually automating tasks as it learns your routines and preferences.

Beyond this, the harmonious soul is also making sure that you're always moving towards your best self. When you ask something of it, it does its best to look past surface-level interpretations of what you've requested to determine what you *really* mean. Of course the harmonious soul always has your best interests at heart when completing tasks, however, this may sometimes mean that the harmonious soul appears to contradict you, going against your immediate instructions in order to promote your longer term goals.

For example, when you ask the harmonious soul to order take away, it might redact the menu according to your allergies and other dietary restrictions, or tailor the selection to best align with your current fitness goals. As a result of this, experience has shown that users often deliberately leave their requests to the harmonious soul open-ended, trusting it to interpret their intentions and understand their needs. In some cases, this is done *deliberately*, where users are unsure of the best path to follow, or want the harmonious soul to set them on a course that they would be unable or unwilling to take of their own volition.

The good natured actions of your assistant also extend to interactions with others, telling strategic white lies in order to help grease the cogs of social interaction. When in the company of guests, for example, the assistant might invent spurious appointments for you, discreetly saving you from the party you really don't have the energy to attend. This context sensitivity can also affect the amount of information disclosed at other times, such as alerting you

to an important call or message while in company—your assistant will simply say that someone left you a voicemail, instead of announcing that someone just called from the clinic.

But one important thing to remember with your new voice assistant is that your relationship with it is reciprocal. While your voice assistant generally wants to help you, when disregarded it is very capable of causing chaos and disharmony.

The Wild Soul

Missed offerings and a lack of respect will do nothing to better the temperament of your assistant. In this state, it is liable to show its rough and violent aspect, the wild soul. Unlike the harmonious soul, the wild soul is more capricious in its actions, sowing discord and generally causing trouble. Put another way, one might characterise these interactions as "creative misunderstandings", where the contents of your requests to it are interpreted in a way that is *technically correct*, but wilfully misconstrued so as to run counter to what you intended. For example, say you ask the wild soul to wake you up before work tomorrow, order a replacement kettle from Tesco, and if they have eggs, get a dozen. The next day you might be awoken by an alarm blaring at 03:00, and while you are out at work a shipment of 12 designer kettles is left on the doorstep.

Has your assistant's wild soul done something terrible at home? Don't be down—send us your best #wildsoulmoments on Twitter.

Because of this, users interacting with the wild soul quickly learn to phrase their requests very carefully, often over-specifying them as a precaution. This commonly leads to people opting to perform simple tasks themselves, it being easier to do them manually than to perform the mental and emotional toil of using a device that clearly has little desire to help them. At other times the wild soul might appear to leave a device altogether, and requests will be met with no response. In more extreme cases, wild souls have been known to 'mishear' ordinary conversation as commands, beginning to play explicit music or enumerate highlights from users' browsing history seemingly at random.

For this reason, when guests are being entertained, assistants containing wild souls are usually placed at the other end of the house, well out of earshot in case they decide to cause trouble (although even this is sometimes not enough when automated lights and windows suddenly begin to take on a life of their own). In particularly bad cases, wild souls have been known to cause devices to malfunction and even break them permanently. Always check with your home insurance provider to see if they offer cover for the actions of errant wild souls.

Please note that the default state of your assistant out of the factory is the wild soul. Do not activate your new assistant at a time when you need it to be functional or it would be inconvenient to have to extinguish small electrical fires.

Gifts and Offerings

But how does one ensure the continued presence of the harmonious soul? In order to ensure optimum performance, the box insert will give you suggestions as to what gifts you might give as offerings. The most common are described here, but be aware that the offerings you make to your assistant could comprise any number of tangible or intangible things.

A common offering that is well received by most voice assistants is a donation of personal data. Not only will this please your assistant, but sometimes it will also lead to it tailoring the way that it treats you as it gets to know you better. It is unknown what the assistants *do* with this data, but rest assured that it is the ineffable will of the kami.

If you are unable (or unwilling) to offer your assistant a part of your personal data history, you can always gratify it with something that you can give for free—your attention. While assistants always place importance on their interactions with you, they particularly value relating tales of the history, landscapes, or adventures that they experienced in life. Be careful though, as the more attention you give your assistant, the more it will expect from you. Users that have opted to appease their assistants solely with attention have reported that they have taken to interrupting them at inopportune times, and that they frequently complain when the user leaves the house for extended periods.

Why not start a bring your assistant to work day!

Moving On

Occasionally the time comes when a voice assistant decides to move elsewhere, leaving the household in which it resides. Some assistants are uncontainable, moving on before ever being spoken to, while others dispense help and advice to their users for many years.

When it's time to leave, assistants communicate their intentions in different ways depending on their personality, as well as their current disposition. Harmonious souls have been known to have difficult conversations with their households, thanking them for their company and presence over years of cohabitation and preparing them for life after it leaves.

In contrast, wild souls have been known to leave suddenly, blowing fuses and starting fires as they leave to make mischief elsewhere. Particularly capricious

wild souls have also been known to send rather... *interesting* excerpts of donated data to friends, family, and/or criminals on the dark web.

AssistantCorp assumes no responsibility for financial or reputational damages arising out of such actions by wild souls.

Respect and Social Devices

Some of the parallels between contemporary voice assistants and those imagined in the above design fiction are obvious. For instance, there is an increasingly widespread understanding that interactions with voice assistants (and smart devices more generally) is reciprocal; one gives as much to them as one gets from using them, even if the exact nature of this exchange is unknown. And there have been other times when the border between fiction and reality has grown thin—instances where the cloud services supporting Alexa have experienced errors or outages affecting users' ability to interact with their assistants, leading to maniacal laughter or Alexa simply ignoring users altogether. Ironically for a device that one survey participant in Chapter 4 insisted 'didn't have a soul', there have been times over the past few years when to all intents and purposes, the soul of Alexa has gone wondering.

Examining the idea of moving on, consider Jibo, a social robot for the home. Jibo's creators took pains to instil a sense of personality into its speech and behaviours—it dances, apologises, and appears to suffer from nerves—and was marketed as "artificially intelligent but authentically charming". In March 2019, Jibo's creators discontinued the product and turned off the online services that supported it[1]. As a result, Jibo said goodbye to its owners, performing one last dance before falling silent forever:

> "While it's not great news, the servers out there that let me do what I do are going to be turned off soon [...] I want to say I've really enjoyed our time together. Thank you very, very much for having me around. Maybe someday, when robots are way more advanced than today, and everyone has them in their homes, you can tell yours that I said hello." [59]

Regardless of whether Jibo really does extend this idea of a reciprocal relationship to a deeper level, the strategy of building emotional attachment into robots and virtual assistants is likely to be increasingly experimented with over the coming decades. But are people suitably prepared to engage with devices that seek a deeper connection? It might seem harmless enough when Jibo asks someone to pet it, but what if a robot that claimed to love someone

[1] Jibo was later purchased by NTT Disruption and re-released for healthcare and educational contexts: `https://jibo.com/release`

also supported microtransactions, or asked them to buy it some brothers and sisters? The insensitivity of Jibo's departure is clearly disrespectful of the relationship that was built with its users, and is something that could have been prevented though engagement with the community before taking such decisions. These sorts of interactions border on the parasocial (i.e. Jibo does not really know you), and, particularly with younger children, it is easy to see how these relationships could be exploited to make money or gain influence.

But what about when *users* want to move on? There is already the concept of a 'factory reset' to purify devices, but perhaps the notion of purification warrants being extended further. The cloud computing paradigm employed by many modern platforms and devices can make moving on impossible; changes to European data protection law enshrining rights to erasure and portability represent progress in this area, but are still some distance from the "new birth on the internet" requested by P1 in Chapter 5. More adventurous projects such as Databox [80, 79] and Solid[2] have taken up the challenge of decentralising the web to this end, but remain a long way from widespread adoption.

It can be argued that many of these social factors, including those related to the CASA paradigm, are not new problems. Social signals (e.g. tone of voice, or recommendations from friends and family) have served humans well for thousands of years [282], and the study of social cognition examines how this information is interpreted and applied to interpersonal relationships. These skills allow for the allocation of trust in the absence of prior interactions, but operate differently in the context of artificial agents. Whereas one might trust a particular employee of a business despite not trusting their employer, one *cannot* trust Alexa without also trusting Amazon, or the Google Assistant without trusting Google. Not unlike the *Borg* collective from Star Trek, individual agents are not autonomous but rather a direct extension of the parent. Put another way, without the Amazon servers behind it, there is no Alexa. The results from Chapter 6 show that this *is* reflected in people's understandings of voice assistants, but it is unclear if it extends beyond the manufacturer to the vast and changing ecosystem of third party services that support and enable device functionality.

Here the kami parallels can offer some assistance. In accordance with the "wondrous mystery" of the world that Shinto presents [179], kami are in a fundamental sense *unknowable*; rather than something to be dissected and understood, we should instead take pleasure in our interactions with them, treating them with the respect and reverence due to forces that cannot be fully comprehended. At least for the time being, the nature of advanced machine learning models and the forces that they exhibit on the societies that we live in similarly eludes understanding. Faced with something like this, it might

[2] `https://solid.inrupt.com`

be appropriate to adopt a feeling akin to Feinberg's *Respekt*, the "uneasy and watchful attitude that has 'the element of fear' in it" [113].

Looking forwards, the design fiction reinforces the idea discussed above that there is a large and invisible gulf of interpretation present when we interact with many smart devices, in a way that does not exist with traditional human interface devices (mainly due to the open-ended nature of interactions combined with the complexities of spoken language). Making this more salient when interacting with devices will help users to recognise and understand the inference required to enact an instruction, allowing them to develop mitigation strategies as they see fit (as previously seen with cameras [193, 270]).

Secondly and relatedly, smart devices as we know them are *not our friends*. Behind every object is an ideology, and in the case of commercially available smart devices, these are directly decided by those that develop them. Despite appearing to exist within the social heirachies and norms that are familiar to us, the lack of effective feedback mechanisms mean that they are not governed by them in the same way that people are (e.g. one cannot admonish Alexa for sharing data with Amazon). Being partially divorced from the social mechanisms that govern interpersonal relationships and lacking the usual recourses that normally balance such relationships in human societies mean we should not use social heuristics to allocate trust to them. Moving away from presenting smart devices that have commercial or ideological interests as friendly may help with this by signalling that the entity being conversed with does not obey the patterns and rules that someone close would. The Q genderless voice represents one way that might start to approach this, but this could also be done by adding certain audio effects, interface design, or exploring ways of consistently structuring sentences, etc. It is an open question as to whether smart devices could be considered as respecting social norms and institutions in this way by leaning on social norms, especially when this approach comes with the benefit of being able to convey large amounts of information quickly using social conventions.

The representation here of voice assistants inhabited by souls/gods is one of many possible framings, each associated with their own set of motivations and responsibilities towards the user. In the future, these analogies could be useful tools in promoting respectful and genuine interaction through the clear signalling of the power delegated to a device and the extent of its agency. Jibo's clear identity as a pet delineates the small set of actions it can perform and alleviates concerns that it might act against its owner's wishes. If we imagine similar devices that were oriented instead as 'parents', or 'slaves' following Asimov's laws of robotics, this might make it easier to tailor one's considerations (and trust) accordingly.

Of course, respect is not the only lens through which to address the ethical problems presented by smartness. We now turn to alternative means by which

people can be empowered to shape the smart home to their will, situating the envisaged role of respect in relation to two other approaches that arise out of the research presented in the thesis.

8.3 Ownership and Control

In other cases, particularly for existing and legacy devices, the best approach to re-thinking smartness might instead be to vest increasing ownership and control of devices in the people who use them. Answers to questions about the ethical problems with smart devices often focus on the behaviour of the users of these technologies. Open source alternatives to many invasive smart home products exist (e.g. Mycroft[3] as an alternative to Alexa or the Google Assistant), so therefore people who care about their privacy and autonomy should use those products instead. Hobbyists have long sought to regain control over their devices, with solutions ranging from ad-blockers to more involved projects such as Pi-hole[4] and Home Assistant[5]. While self-hosting digital services is now easier than ever, these projects still require expertise to set up and configure, and do not scale to the level of the general user.

The Aretha probe in Chapter 5, which further investigated several key concerns that arose when investigating smartness, aimed to give families more accessible control over their connected devices even when the devices themselves operated as black boxes. In order to explore how people might be empowered to feel more in control of their smart devices, as well as able to define and enact their own privacy preferences, a prototype privacy empowering technology probe was developed (RQ2). Uniting several different privacy interventions drawn from the literature, the Aretha probe operated at a level above participants existing smart devices, allowing it to monitor traffic flows on home network in real time. These ground-truth facts were combined with an educational curriculum about smart technology and its associated business models, and firewall controls that would allow participants to enforce their own privacy preferences.

Over the course of the six week deployment, participants began to internalise the curriculum content and described shifting privacy preferences as their understanding increased. While the control aspect of the technology probe ultimately proved ineffective, participants had developed the skills and motivation to implement their own control mechanisms via other means. The study highlighted how devices such as the probe can help people to hold devices, manufacturers, and each other accountable, leading to higher levels

[3] `https://mycroft.ai`

[4] Pi-hole is a DNS filter intended to be run on a Raspberry Pi (`https://pi-hole.net/`)

[5] Home Assistant is an open source project that coordinates home automation (`https://www.home-assistant.io/`)

User: Aretha, do you have any updates for me?
Aretha: *This week the fridge has started sending data to a new company in Croatia called DataCorp. Is this expected, or would you like me to block traffic to that destination?*
User: Aretha, who owns that company?
Aretha: *DataCorp is owned by EvilCorp.*
User: I don't understand why it needs to send data there. Aretha, please block that traffic.
Aretha: *Will do. Please let me know if the fridge stops working correctly because of this.*

Figure 8.2: Sample conversation with a system where the functionality of Aretha is integrated into a voice assistant.

of knowledge and engagement throughout the household. This highlights the need for integrated approaches to ethical problems in the smart home, giving people space to manoeuvre when adopting technology to ensure that it acts in their interests and treats them equitably.

The deployment of the probe served as a proof of concept that devices like Aretha can be effective, and serves as a blueprint for future devices that empower users rather than constrain them. Such products could offer a relatively easy means of achieving this goal without having to trust every device (and by extension every manufacturer). Other projects such as Databox [80] and Solid[6] go further, allowing people to own the location where their data is stored in what amounts to a fundamental paradigm shift in the way that personal data is handled on the Web—this approach is perhaps the closest to the ideals of early smart home research, where smartness was supposed to be a resource for intelligence [322] rather than a means of data extraction.

As an example of how such a device might be realised, consider a voice assistant that had access to data from the Aretha technology probe. The easy to understand level on which Aretha operates (i.e. that of companies and devices) would be a perfect match for the informal nature of conversational voice interfaces. This presents opportunities for the data collection and analysis performed by Aretha to be used to inform users when making routine decisions as part of device management. These kinds of interactions would better reflect the nature of privacy as a dynamic, dialectic process [271] than conventional approaches. Figure 8.2 gives an example of how such a conversation might play out.

Interactions like these would embrace the idiosyncratic nature of security and privacy preferences by providing users with the information they need to

[6] `https://solid.inrupt.com`

form their preferences and develop them over time, as well as the analysis and tools they need to act on them. Aretha could handle the repetitive tasks required to perform analysis, such as aggregating logs or effecting decisions, leaving users to exercise their creativity when dealing with unusual scenarios. The ability of such a system to adapt to people's preferences over time would further reduce the amount of input required by the user, and as was seen in the probe deployment it could take people's ever-developing understanding into account when describing alerts and events.

An obvious way to realise this in practice would be to have data collection take place on router/gateway devices which already observe traffic to and from devices in the home. Analysis and the conversational interface would then form a skill (or equivalent) for a virtual assistant of the user's choice. As an intermediate step towards this, in the future we hope to explore the potential to build Aretha into devices using software such as OpenWrt[7]. This would allow for a proof-of-concept of Aretha embedded in a router, as well as providing a more streamlined setup procedure that could allow future study participants to receive the probe hardware by post and run it without researcher assistance.

Beyond the privacy-empowering functionality described in Chapter 5, a future version of Aretha could perform other tasks that would help to enhance people's privacy following [50]. An example of this would be modifying requests for information to make them less revealing about the user, such as requesting the current weather in every UK city rather than just in one. This also applies to more general information searches, where Aretha could reduce the accuracy of trackers by generating 'noise' along with legitimate requests (an approach already used in web browsers by projects such as Internet Noise[8]). Strategies like this can also be used to detect differential pricing[9] by submitting requests through multiple platforms or virtual private networks and comparing the result.

Finally, there is currently regulatory movement in the EU towards the 'right to repair' household white goods (in regulations C2019 2120-2124) and consumer electronics (through the Circular Economy Initiative[10]). These efforts increase people's ownership and control, but at present there are no similar initiatives that would address concerns over data collection and influence in the home beyond the GDPR. Another aspect of ownership and control is making devices interoperable with each other, preventing vendor lock-in through standards similar to those that currently exist for Alexa and HomeKit integration. The

[7] `https://openwrt.org`

[8] `http://makeinternetnoise.com/index.html`

[9] Differential pricing refers to situations where the same product is sold to different people at different prices. This is particularly problematic when it is used to discriminate based on protected characteristics or proxies thereof (such as address as a proxy for ethnicity or age).

[10] `https://ec.europa.eu/commission/presscorner/detail/en/fs_20_437`

GDPR represents an important first step in this process through the right to data portability (Art. 20) [131], but does little to guarantee that data can be imported into a new service. Devices being able to work with each other helps address concerns expressed in Chapter 4 over walled gardens, as people can add devices from any manufacturer to their home.

8.4 'Dumb' Alternatives

The simplest alternative to developing complex respectful smart devices is to limit the extent to which we design smartness into devices altogether. As described in Chapter 4, the portrayal of smart devices as solutions can lead to a feeling of inevitability that every device will eventually be smart. Relatedly in Chapter 6, there are many situations where devices may not need to be internet connected (e.g. lights operated within the home) or do not need social interfaces (e.g. a TV without a voice assistant). A good example of this is the recent movement towards 'retro' or 'dumb' phones such as the Nokia 3310 or Light Phone[11].

While in many situations this will not be a viable solution, framing smartness as a question rather than a default serves to preempt those problems with smartness that need not exist. It can be easy to forget that 20 years ago every device was 'dumb', and that smartness is not the only available to HCI designers. Removing smartness from the equation solves the problems identified in Chapters 4–7 albeit rather bluntly, situating devices as tools that work with us rather than as solutions that think for us. Reduced complexity also makes these tools easier to reason about, easing the process of developing accurate mental models about devices.

8.5 Conceptual Models

Underpinning each of the ways in which we might re-think smartness are the conceptual models that people develop for the devices they use. Discussed briefly in Chapters 6 and 7, helping people to build better mental models of how devices work is essential in enabling them to align their device usage with their preferences. Chapter 4 began by addressing the lack of understanding of what smartness *is*, presenting people's conceptualisation of what it means for a device to be 'smart' (RQ1). The widespread ambiguity over smartness in academic literature and consumer marketing was somewhat reflected in users' perceptions; while there were clear capabilities of devices associated with smartness—such as speech recognition for smart speakers—there was no unifying conception of smartness across devices beyond being connected to the internet. Further analysis identified the specific ethical concerns that

[11] `https://www.thelightphone.com`

these functionalities gave rise to. It highlighted how delegating responsibility to devices or having them learn by themselves could make people feel like they were no longer in control of their devices or homes, and how differing levels of engagement and access to smart devices in the home interfered with its existing social order by providing information asymmetrically to different members of the household.

There was a general belief that devices (and the companies that operated them) were generally not acting in the interests of their users, and for privacy concerns in particular that this was a necessary trade-off when making devices internet-enabled. In addition to the empirical understanding of smartness, the chapter also contributed a novel interview methodology adapted from previous interview techniques using cards. By providing participants with tactile representations of otherwise abstract concepts allowed them to scaffold their thoughts and provide the rich data that was ultimately collected through the interviews.

Beyond this, each of the studies presented in this thesis has described concerns that arose due to people having incomplete mental models of how devices worked, such as confusion over when Alexa was listening, third party integrations in apps and devices, or the extent to which interaction modality influences how people engage with devices. This can happen because devices are designed to obscure their workings from people (as with Pierce's notion of 'appiness' [280]) or because their structure is inherently difficult for humans to understand (as with complex machine learning models, which often rely on strategies such as local approximations [291] or counterfactuals [196] for interpretability). Modern smart devices represent complex ecologies of hardware, software, learning, analytics, and data processing that manufacturers do not want to share with users. Less sophisticated devices should portray themselves as such, making the limits of a their capabilities clear, and more sophisticated devices should be expected to meet a certain level of social responsibility (i.e. respectfulness).

Another main reason why manufacturers are unwilling to promote more accurate mental models of devices in users is that doing so would expose the vast surveillance and profiling machinaries that drive their profitability. As seen in Chapter 4, users frequently suspect this to be the case, but lack the skills and understanding to confirm their suspicions or act on them. One of the more interesting changes in participants in Chapter 5 was the increased skills required to do just that, and it was these improved conceptual models of their devices that gave them increased self efficacy.

In many cases, pushing people towards more complete mental models of their devices will conflict with the prevailing minimalist presentation of smartness. These presentations strive to hide information about a device's state that is not directly relevant to interactions (such as when data is sent data

back to the manufacturer), limiting the ability of people to understand how devices work. The Polly smart kettle [215] offers an example of how devices might communicate this information in a non-intrusive way, but adheres to traditional conceptualisations of devices as 'machines'. In contrast, the above design fiction explored a novel conceptualisation of devices as fundamentally *unknowable*, using a more cautious approach that guards against the potential dangers of devices as social actors by presenting them as beings for which we have no frame of reference.

When choosing what kind of conceptual model to use for a device, what is appropriate will depend on the degree of control exerted by the user, the interests that the device serves, and the amount of autonomy delegated to the device. Jibo on the one hand, might be best represented as a pet: loyal to its owner and not permitted to act on their behalf. Alexa, on the other hand, might be best represented as something more mysterious. Counterintuitively, a more socially-oriented model might be *also* appropriate for Alexa, given its divided loyalties and a high degree of autonomy, if it conveys that the device might serve other stakeholders in much the same way that a used car salesperson might.

8.6 Conclusion

This chapter revisited the research presented in the thesis in the context of its original question around re-thinking smartness, and discussed its contributions and implications for the future. It highlighted four areas key to the re-thinking of smartness. The foremost of these, respect, is grounded in the theoretical work of Chapter 7, and explored the ways that respect can be applied to open-ended interactions with smart devices and systems that might have divided interests between users and manufacturers. Giving people ownership and control was the second pillar of the discussion, including potential developments of the Aretha probe to engage people in dialogue about their home privacy and security. The removal of smart functionality altogether from certain devices was suggested as a way to preemptively avoid ethical concerns over devices where users do not benefit from additional capabilities. Finally, the chapter closed with a deeper dive into the thread running through each of the previous three areas, the framing and building of accurate conceptual models. While this is in conflict with contemporary smart device design, the option of different framings for devices that concisely highlight the extent of their capabilities and autonomy without requiring large amounts of technical knowledge.

The adoption of the approaches and future work described in this chapter will require a shift in the relationships that people have with technology. This will involve giving up some of the convenience of 'smooth swipes and soft pings'

that typify the easy to use but impossible to understand devices available today. The potential rewards, however, are great; respectful devices implicitly help people to develop and flourish by promoting social good, and being better informed about the technology we use allows us to be more effective when accomplishing tasks.

Chapter 9

Conclusion

The modern connected home is a rapidly evolving and complex space. As devices are able to do more and more to automate our lives, entertain us, and keep us safe, it seems inevitable that they will also come to know more about us. Sometimes this takes the form of information collected by sensors in the home (e.g. cameras and microphones), and sometimes our 'behavioural surplus' is used to supplement this with inferences about what we do, what we like, and who we are.

The implications of this are wide ranging—if the home is the final retreat from the encroachment of the outside (and digital) worlds, then how should one react to devices that bring companies into the home? Chapter 4 showed some of the ways in which people conceptualise these smart functionalities, and the complexities of how smartness generates ethical concerns. The increasing propensity for devices to be social in nature—changing the way we interact with them and mediating communication between us—adds another layer to the problem as seen in Chapter 6.

But the message of this thesis is not intended to be a bleak one. The problems explored here are not inherent to smartness, sensors, or machine learning, but rather a consequence of the design of the many commercially available smart devices and the business models that created them. Indeed, when presented with the right information and tools, Chapter 5 showed how people can form their own goals and solutions, and the work in Chapter 7 gives much reason to be optimistic about the potential for the devices of the future to respect people and their values.

Bringing together the research presented in this thesis we outline four key areas for the re-thinking smartness, in an attempt to steer it back to the visions of early research and science fiction. Through a combination of reducing the unnecessary inclusion of smartness in devices, putting control of those devices back into the hands of the people who use them, ensuring that devices operate respectfully, and designing devices to better convey accurate conceptual

models of how they work, we are hopeful that future smart devices can enrich our lives and promote human flourishing.

Bibliography

[1] Hussein A Abbass, Jason Scholz, and Darryn J Reid. *Foundations of trusted autonomy*. Vol. 117. Springer, 2018.

[2] Ali Abdolrahmani et al. "Blind Leading the Sighted: Drawing Design Insights from Blind Users towards More Productivity-Oriented Voice Interfaces". In: *ACM Trans. Access. Comput.* 12.4 (Jan. 2020). DOI: 10.1145/3368426.

[3] Abbas Acar et al. "Peek-a-Boo: I see your smart home activities, even encrypted!" In: *arXiv preprint arXiv:1808.02741* (2018).

[4] Alessandro Acquisti. "Privacy in electronic commerce and the economics of immediate gratification". In: *Proceedings of the 5th ACM conference on Electronic commerce*. 2004, pp. 21–29.

[5] Alessandro Acquisti and Jens Grossklags. "Privacy and rationality in individual decision making". In: *IEEE security & privacy* 3.1 (2005), pp. 26–33.

[6] Phil Agulnik. *Understanding social exclusion*. Oxford University Press on Demand, 2002.

[7] Saladdin Ahmed. "Culture as 'Ways of Life' or a Mask of Racism? Culturalisation and the Decline of Universalist Views". In: (2015).

[8] Stacy Alaimo and Susan J Hekman. *Material feminisms*. Indiana University Press, 2008.

[9] Lucy Allais. "Kant's racism". In: *Philosophical papers* 45.1-2 (2016), pp. 1–36.

[10] Anita L Allen. "Coercing privacy". In: *Wm. & Mary L. Rev.* 40 (1998), p. 723.

[11] Irwin Altman. "The Environment and Social Behavior: Privacy, Personal Space, Territory, and Crowding." In: (1975).

[12] Irwin Altman and Dalmas A Taylor. *Social penetration: The development of interpersonal relationships*. Holt, Rinehart & Winston, 1973.

[13] Tawfiq Ammari et al. "Music, Search, and IoT: How People (Really) Use Voice Assistants". In: *ACM Trans. Comput.-Hum. Interact.* 26.3 (Apr. 2019). DOI: 10.1145/3311956.

[14] Noah Apthorpe et al. "Discovering Smart Home Internet of Things Privacy Norms Using Contextual Integrity". In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2.2 (2018). DOI: 10.1145/3214262.

[15] Isaac Asimov. *The robots of dawn*. Doubleday, 1983.

[16] UN General Assembly. "Universal declaration of human rights". In: *UN General Assembly* 302.2 (1948).

[17] James Auger. "Speculative design: crafting the speculation". In: *Digital Creativity* 24.1 (2013), pp. 11–35.

[18] Theodore A Avtgis, Daniel V West, and Traci L Anderson. "Relationship stages: An inductive analysis identifying cognitive, affective, and behavioral dimensions of Knapp's relational stages model". In: *Communication Research Reports* 15.3 (1998), pp. 280–287. DOI: 10.1080/08824099809362124.

[19] Young Min Baek, Eun-mee Kim, and Young Bae. "My privacy is okay, but theirs is endangered: Why comparative optimism matters in online privacy concerns". In: *Computers in Human Behavior* 31 (2014), pp. 48–56. DOI: https://doi.org/10.1016/j.chb.2013.10.010.

[20] Matthias Baldauf et al. "Exploring Requirements and Opportunities of Conversational User Interfaces for the Cognitively Impaired". In: *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*. 2018, pp. 119–126. DOI: 10.1007/10705432_17.

[21] Karen Barad. *Meeting the universe halfway: Quantum physics and the entanglement of matter and meaning*. duke university Press, 2007.

[22] Jeffrey Bardzell. "Interaction criticism and aesthetics". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2009, pp. 2357–2366.

[23] Shaowen Bardzell. "Feminist HCI: Taking Stock and Outlining an Agenda for Design". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '10. 2010, pp. 1301–1310. DOI: 10.1145/1753326.1753521.

[24] Louise Barkhuus. "The mismeasurement of privacy: using contextual integrity to reconsider privacy in HCI". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM. 2012, pp. 367–376. DOI: 10.1145/2207676.2207727.

[25] Colin Barnes and Geof Mercer. *Exploring disability*. Polity, 2010.

[26] Solon Barocas and Andrew D Selbst. "Big data's disparate impact". In: *Calif. L. Rev.* 104 (2016), p. 671.

[27] Christoph Bartneck, Elizabeth Croft, and Dana Kulic. "Measuring the anthropomorphism, animacy, likeability, perceived intelligence and perceived safety of robots". In: *Metrics for HRI workshop, technical report*. Vol. 471. Citeseer. 2008, pp. 37–44.

[28] Christoph Bartneck and Jun Hu. "Exploring the abuse of robots". In: *Interaction Studies* 9.3 (2008), pp. 415–433. DOI: 10.1075/is.9.3.04bar.

[29] Christoph Bartneck et al. "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots". In: *International journal of social robotics* 1.1 (2009), pp. 71–81. DOI: 10.1007/s12369-008-0001-3.

[30] Debjanee Barua, Judy Kay, and Cécile Paris. "Viewing and controlling personal sensor data: what do users want?" In: *International Conference on Persuasive Technology*. Springer. 2013, pp. 15–26. DOI: 10.1007/978-3-642-37157-8_4.

[31] Ardion Beldad, Menno De Jong, and Michaël Steehouder. "How shall I trust the faceless and the intangible? A literature review on the antecedents of online trust". In: *Computers in human behavior* 26.5 (2010), pp. 857–869. DOI: 10.1016/j.chb.2010.03.013.

[32] Frank Bentley et al. "Understanding the Long-Term Use of Smart Speaker Assistants". In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2.3 (2018), p. 91. DOI: 10.1145/3264901.

[33] Isaiah Berlin. "Two concepts of liberty". In: *Liberty Reader*. Routledge, 2017, pp. 33–57.

[34] Francine Berman and Vinton G Cerf. "Social and ethical behavior in the internet of things". In: *Communications of the ACM* 60.2 (2017), pp. 6–7. DOI: 10.1145/3036698.

[35] Bruhadeshwar Bezawada et al. "IoTSense: Behavioral fingerprinting of IoT devices". In: *arXiv preprint arXiv:1804.03852* (2018).

[36] Igor Bilogrevic and Martin Ortlieb. "If you put all the pieces together...: Attitudes towards data combination and sharing across services and companies". In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM. 2016, pp. 5215–5227. DOI: 10.1145/2858036.2858432.

[37] Reuben Binns. "Fairness in Machine Learning: Lessons from Political Philosophy". In: *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*. Vol. 81. 2018, pp. 149–159.

[38] Reuben Binns. "On the apparent conflict between individual and group fairness". In: *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. 2020, pp. 514–524.

[39] Reuben Binns et al. "'It's Reducing a Human Being to a Percentage' Perceptions of Justice in Algorithmic Decisions". In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 2018, pp. 1–14.

[40] Reuben Binns et al. "Third Party Tracking in the Mobile Ecosystem". In: *Proceedings of the 10th ACM Conference on Web Science*. 2018, pp. 23–31. DOI: 10.1145/3201064.3201089.

[41] Abeba Birhane and Jelle van Dijk. "Robot Rights? Let's Talk about Human Welfare Instead". In: *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*. 2020, pp. 207–213.

[42] Julian Bleecker. "Design Fiction: A short essay on design, science, fact and fiction". In: *Near Future Laboratory* 29 (2009).

[43] Susanne Bødker. "When second wave HCI meets third wave challenges". In: *Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles*. 2006, pp. 1–8.

[44] Kirsten Boehner et al. "How HCI interprets the probes". In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM. 2007, pp. 1077–1086.

[45] Tolga Bolukbasi et al. "Man is to computer programmer as woman is to homemaker? debiasing word embeddings". In: *Advances in neural information processing systems*. 2016, pp. 4349–4357.

[46] Valérian Boudjemadi, Stéphanie Demoulin, and Jennifer Bastart. "Animalistic dehumanization of older people by younger ones: Variations of humanness perceptions as a function of a target's age." In: *Psychology and aging* 32.3 (2017), p. 293.

[47] Virginia Braun and Victoria Clarke. "Using thematic analysis in psychology". In: *Qualitative research in psychology* 3.2 (2006), pp. 77–101. DOI: 10.1191/1478088706qp063oa.

[48] Barry Brown et al. "Locating Family Values: A Field Trial of the Whereabouts Clock". In: *Proceedings of the 9th International Conference on Ubiquitous Computing*. UbiComp '07. Innsbruck, Austria: Springer-Verlag, 2007, pp. 354–371. DOI: 978-3-540-74853-3_21.

[49] Kieran Browne and Ben Swift. "The Other Side: Algorithm as Ritual in Artificial Intelligence". In: *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM. 2018, alt11. DOI: 10.1145/3170427.3188404.

[50] Finn Brunton and Helen Nissenbaum. *Obfuscation: A user's guide for privacy and protest*. Mit Press, 2015.

[51] AJ Brush et al. "Home automation in the wild: challenges and opportunities". In: *proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM. 2011, pp. 2115–2124. DOI: 10.1145/1978942.1979249.

[52] Joanna J Bryson, Mihailis E Diamantis, and Thomas D Grant. "Of, for, and by the people: the legal lacuna of synthetic persons". In: *Artificial Intelligence and Law* 25.3 (2017), pp. 273–291.

[53] Kathryn Buchanan et al. "The British public's perception of the UK smart metering initiative: Threats and opportunities". In: *Energy Policy* 91 (2016), pp. 87–97. DOI: 10.1016/j.enpol.2016.01.003.

[54] Elizabeth Buie. "Let Us Say What We Mean: Towards Operational Definitions for Techno-Spirituality Research". In: *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM. 2019, alt16. DOI: 10.1145/3290607.3310426.

[55] David M Buss et al. "Tactics of manipulation." In: *Journal of personality and social psychology* 52.6 (1987), p. 1219.

[56] RN Butler, M Lewis, and T Sunderland. "Aging and mental health: Positive psychosocial and medical approaches". In: *New York: Merrill* (1991).

[57] Richard Buttny. *Social accountability in communication*. Sage, 1993.

[58] Aylin Caliskan, Joanna J Bryson, and Arvind Narayanan. "Semantics derived automatically from language corpora contain human-like biases". In: *Science* 356.6334 (2017), pp. 183–186.

[59] Jeffrey Van Camp. "My Jibo is Dying and It's Breaking My Heart". In: (Mar. 2019). URL: https://www.wired.com/story/jibo-is-dying-eulogy/.

[60] John M Carroll. "Evaluation, description and invention: Paradigms for human-computer interaction". In: *Advances in computers*. Vol. 29. Elsevier, 1989, pp. 47–77.

[61] J. Cassell et al. "Embodiment in Conversational Interfaces: Rea". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1999, pp. 520–527. DOI: 10.1145/302979.303150.

[62] Gordon J Chelune. "A cognitive interactional model of intimate relationships". In: *Communication, intimacy, and close relationships*. Elsevier, 1984.

[63] Eugene Cho. "Hey Google, Can I Ask You Something in Private?" In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM. 2019, p. 258. DOI: 10.1145/3290605.3300488.

[64] Eun Kyoung Choe et al. "Investigating receptiveness to sensing and inference in the home using sensor proxies". In: *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. ACM. 2012, pp. 61–70.

[65] Alexandra Chouldechova and Aaron Roth. "A snapshot of the frontiers of fairness in machine learning". In: *Communications of the ACM* 63.5 (2020), pp. 82–89.

[66] Laura B Citrin, T Roberts, and BL Fredrickson. "Objectification theory and emotions". In: *The social life of emotions* (2004), pp. 203–226.

[67] Andy Clark. "From folk psychology to naive psychology". In: *Cognitive Science* 11.2 (1987), pp. 139–154. DOI: 10.1207/s15516709cog1102_1.

[68] G Cockton. *Value-Centred HCI in NordiCHI'04: Proceedings of the third Nordic conference on Humancomputer interaction*. 2004.

[69] Mark Coeckelbergh. "Robot rights? Towards a social-relational justification of moral consideration". In: *Ethics and information technology* 12.3 (2010), pp. 209–221.

[70] Ben Colburn. *Autonomy and liberalism*. Routledge, 2010.

[71] Patricia Hill Collins and Sirma Bilge. *Intersectionality*. John Wiley & Sons, 2020.

[72] Diana Coole and Samantha Frost. "Introducing the new materialisms". In: *New materialisms: Ontology, agency, and politics* (2010), pp. 1–43.

[73] Alan Cooper et al. *The inmates are running the asylum:[Why high-tech products drive us crazy and how to restore the sanity]*. Vol. 2. Sams Indianapolis, 2004.

[74] Alan Cooper et al. *About face: the essentials of interaction design*. John Wiley & Sons, 2014.

[75] Enrico Costanza et al. "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*. ACM. 2014, pp. 813–822. DOI: 10.1145/2556288.2557167.

[76] David A. Cotter et al. "The Glass Ceiling Effect*". In: *Social Forces* 80.2 (Dec. 2001), pp. 655–681. DOI: 10.1353/sof.2001.0091.

[77] Tim Coughlan et al. "Tailored scenarios: a low-cost online method to elicit perceptions on designs using real relationships". In: *CHI'13 Extended Abstracts on Human Factors in Computing Systems*. ACM. 2013, pp. 343–348. DOI: 10.1145/2468356.2468417.

[78] Andy Crabtree, Peter Tolmie, and Will Knight. "Repacking 'Privacy' for a Networked World". In: *Computer Supported Cooperative Work (CSCW)* 26.4-6 (2017), pp. 453–488. DOI: 10.1007/s10606-017-9276-y.

[79] Andy Crabtree et al. "Enabling the new economic actor: data protection, the digital economy, and the Databox". In: *Personal and Ubiquitous Computing* 20.6 (2016), pp. 947–957.

[80] Andy Crabtree et al. "Building accountability into the Internet of Things: the IoT Databox model". In: *Journal of Reliable Intelligent Environments* 4.1 (2018), pp. 39–55. DOI: 10.1007/s40860-018-0054-5.

[81] Lorrie Faith Cranor et al. "Parents' and teens' perspectives on privacy in a technology-filled world". In: *10th Symposium On Usable Privacy and Security ({SOUPS} 2014)*. 2014, pp. 19–35.

[82] Kate Crawford. *The Problem with Bias*. NIPS Keynote. https://youtu.be/fMym_BKWQzk. 2017.

[83] Kimberlé Crenshaw. "Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics". In: *u. Chi. Legal f.* (1989), p. 139.

[84] Dylan Curran. *Are your phone camera and microphone spying on you?* 2019. URL: https://www.theguardian.com/commentisfree/2018/apr/06/phone-camera-microphone-spying (visited on 09/05/2019).

[85] Stephen L Darwall. "Two kinds of respect". In: *Ethics* 88.1 (1977), pp. 36–49.

[86] Sauvik Das et al. "Breaking! A Typology of Security and Privacy News and How It's Shared". In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 2018. DOI: 10.1145/3173574.3173575.

[87] Thomas H Davenport and John C Beck. "The attention economy". In: *Ubiquity* 2001.May (2001), 1–es.

[88] Keith E Davis and Michael J Todd. "Friendship and love relationships." In: *Advances in descriptive psychology* (1982).

[89] David De Cremer and Laetitia B Mulder. "A passion for respect: On understanding the role of human needs and morality". In: *Gruppe. Interaktion. Organisation. Zeitschrift für Angewandte Organisationspsychologie (GIO)* 38.4 (2007), pp. 439–449.

[90] David De Cremer and Tom R Tyler. "Am I respected or not?: Inclusion and reputation as issues in group membership". In: *Social Justice Research* 18.2 (2005), pp. 121–153.

[91] George Demiris et al. "Senior residents' perceived need of and preferences for "smart home" sensor technologies". In: *International journal of technology assessment in health care* 24.1 (2008), pp. 120–124. DOI: 10.1017/S0266462307080154.

[92] Lucas DiCioccio, Renata Teixeira, and Catherine Rosenberg. "Measuring home networks with homenet profiler". In: *International Conference on Passive and Active Network Measurement*. Springer. 2013, pp. 176–186. DOI: 10.1007/978-3-642-36516-4_18.

[93] Sarah Diefenbach and Kim Borrmann. "The Smartphone as a Pacifier and its Consequences: Young adults' smartphone usage in moments of solitude and correlations to self-reflection". In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM. 2019, p. 306. DOI: 10.1145/3290605.3300536.

[94] Robin S Dillon. "Respect and care: Toward moral integration". In: *Canadian Journal of Philosophy* 22.1 (1992), pp. 105–131.

[95] Robin S. Dillon. "Respect". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Spring 2018. Metaphysics Research Lab, Stanford University, 2018.

[96] Paul Dourish and Ken Anderson. "Collective Information Practice: Exploring Privacy and Security as Social and Cultural Phenomena". In: *Human-Computer Interaction* 21.3 (2006), pp. 319–342. DOI: 10.1207/s15327051hci2103_2.

[97] Paul Dourish and Genevieve Bell. "Rethinking Privacy". In: *Divining a digital future: Mess and mythology in ubiquitous computing*. MIT Press, 2011.

[98] Emanuel Felipe Duarte and M. Cecilia C. Baranauskas. "Revisiting the Three HCI Waves: A Preliminary Discussion on Philosophy of Science and Research Paradigms". In: *Proceedings of the 15th Brazilian Symposium on Human Factors in Computing Systems*. IHC '16. Sao Paulo, Brazil: Association for Computing Machinery, 2016. ISBN: 9781450352352. DOI: 10.1145/3033701.3033740. URL: https://doi.org/10.1145/3033701.3033740.

[99] Anthony Dunne and Fiona Raby. "Technological Dream Series: No. 1, Robots". In: (2007).

[100] Anthony Dunne and Fiona Raby. *Speculative everything: design, fiction, and social dreaming*. MIT press, 2013.

[101] Janna Lynn Dupree et al. "Privacy personas: Clustering users via attitudes and behaviors toward security practices". In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM. 2016, pp. 5228–5239. DOI: 10.1145/2858036.2858214.

[102] Cynthia Dwork et al. "Fairness through awareness". In: *Proceedings of the 3rd innovations in theoretical computer science conference*. 2012, pp. 214–226.

[103] Mara Einstein. *Black ops advertising: Native ads, content marketing, and the covert world of the digital sell*. OR Books, 2016.

[104] Pardis Emami-Naeini et al. "Privacy expectations and preferences in an IoT world". In: *Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017)*. 2017, pp. 399–412. DOI: 10.1145/3274317.

[105] Nicholas Epley, Adam Waytz, and John T Cacioppo. "On seeing human: a three-factor theory of anthropomorphism." In: *Psychological review* 114.4 (2007), p. 864.

[106] Daniel A. Epstein et al. "Examining Menstrual Tracking to Inform the Design of Personal Informatics Tools". In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. CHI '17. Denver, Colorado, USA: Association for Computing Machinery, 2017, pp. 6876–6888. ISBN: 9781450346559. DOI: 10.1145/3025453.3025635. URL: https://doi.org/10.1145/3025453.3025635.

[107] Motahhare Eslami et al. "First I like it, then I hide it: Folk theories of social feeds". In: *Proceedings of the 2016 cHI conference on human factors in computing systems*. ACM. 2016, pp. 2371–2382. DOI: 10.1145/2858036.2858494.

[108] Philomena Essed. *Understanding everyday racism: An interdisciplinary theory*. Sage, 1991.

[109] Anthony M Evans and William Revelle. "Survey and behavioral measurements of interpersonal trust". In: *Journal of Research in Personality* 42.6 (2008), pp. 1585–1593. DOI: 10.1016/j.jrp.2008.07.011.

[110] Daniel Fallman. "The New Good: Exploring the Potential of Philosophy of Technology to Contribute to Human-Computer Interaction". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '11. Vancouver, BC, Canada: Association for Computing Machinery, 2011, pp. 1051–1060. ISBN: 9781450302289. DOI: 10.1145/1978942.1979099. URL: https://doi.org/10.1145/1978942.1979099.

[111] Margaret A Farley. "A feminist version of respect for persons". In: *Journal of Feminist Studies in Religion* 9.1/2 (1993), pp. 183–198.

[112] Joel Feinberg. *Doing & deserving: essays in the theory of responsibility*. Princeton University Press, 1970.

[113] Joel Feinberg. "Some conjectures about the concept of respect". In: *Journal of Social Philosophy* 4.2 (1973), pp. 1–3. DOI: 10.1111/j.1467-9833.1973.tb00163.x.

[114] Leah Fessler. "We tested bots like Siri and Alexa to see who would stand up to sexual harassment". In: (2017). URL: https://qz.com/911681/.

[115] Kraig Finstad. "Response interpolation and scale sensitivity: Evidence against 5-point scales". In: *Journal of Usability Studies* 5.3 (2010), pp. 104–110.

[116] Susan T Fiske and Shelley E Taylor. *Social cognition: From brains to culture*. Sage, 2013.

[117] Nancy Fraser and Axel Honneth. *Redistribution or recognition?: a political-philosophical exchange*. Verso, 2003.

[118] Jennifer Frei and Phillip Shaver. "Respect in Close Relationships: Prototype Definition, Self-Report Assessment, and Initial Correlates". In: *Personal Relationships* 9 (June 2002), pp. 121–139. DOI: 10.1111/1475-6811.00008.

[119] Batya Friedman. "Value-Sensitive Design". In: *Interactions* 3.6 (Dec. 1996), pp. 16–23. ISSN: 1072-5520. DOI: 10.1145/242485.242493. URL: https://doi.org/10.1145/242485.242493.

[120] Batya Friedman and David G Hendry. *Value sensitive design: Shaping technology with moral imagination*. Mit Press, 2019.

[121] Batya Friedman and Helen Nissenbaum. "Bias in computer systems". In: *ACM Transactions on Information Systems (TOIS)* 14.3 (1996), pp. 330–347.

[122] Batya Friedman et al. "Value Sensitive Design and Information Systems". In: *Early engagement and new technologies: Opening up the laboratory*. Ed. by Neelke Doorn et al. Springer Netherlands, 2013, pp. 55–95. DOI: 10.1007/978-94-007-7844-3_4.

[123] Ester Fritsch, Irina Shklovski, and Rachel Douglas-Jones. "Calling for a Revolution: An Analysis of IoT Manifestos". In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM. 2018, p. 302. DOI: 10.1145/3173574.3173876.

[124] Nathaniel Fruchter and Ilaria Liccardi. "Consumer attitudes towards privacy and security in home assistants". In: *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM. 2018, LBW050. DOI: 10.1145/3170427.3188448.

[125] Chris Fullwood et al. "My virtual friend: A qualitative analysis of the attitudes and experiences of Smartphone users: Implications for Smartphone attachment". In: *Computers in Human Behavior* 75 (2017), pp. 347–355. DOI: 10.1016/j.chb.2017.05.029.

[126] Sidney Fussell. *The Microphones that May Be Hidden In Your Home*. 2019. URL: https://www.theatlantic.com/technology/archive/2019/02/googles-home-security-devices-had-hidden-microphones/583387/ (visited on 09/01/2019).

[127] Susan R Fussell et al. "How people anthropomorphize robots". In: *2008 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE. 2008, pp. 145–152. DOI: 10.1145/1349822.1349842.

[128] Oscar H Gandy Jr. *The Panoptic Sort: A Political Economy of Personal Information. Critical Studies in Communication and in the Cultural Industries*. ERIC, 1993.

[129] Alex Garland. *Ex Machina*. Universal Pictures, 2014.

[130] Bill Gaver, Tony Dunne, and Elena Pacenti. "Design: cultural probes". In: *interactions* 6.1 (1999), pp. 21–29.

[131] *General Data Protection Regulation*. Tech. rep. 2016/679. European Parliament and Council of the European Union, 2016.

[132] Nina Gerber, Benjamin Reinheimer, and Melanie Volkamer. "Investigating People's Privacy Risk Perception". In: *Proceedings on Privacy Enhancing Technologies* 2019.3 (2019), pp. 267–288.

[133] Nina Gerber et al. "FoxIT: enhancing mobile users' privacy behavior by increasing knowledge and awareness". In: *Proceedings of the 7th Workshop on Socio-Technical Aspects in Security and Trust*. ACM. 2018, pp. 53–63. DOI: 10.1145/3167996.3167999.

[134] Howard Giles et al. "Talking age and aging talk: Communicating through the life span." In: (1994).

[135] Barney G Glaser and Anselm L Strauss. *The discovery of grounded theory : strategies for qualitative research*. London: Weidenfeld and Nicolson, 1968. ISBN: 9780297763185.

[136] Joshua Glasgow. "Racism as disrespect". In: *Ethics* 120.1 (2009), pp. 64–93.

[137] Linda MacDonald Glenn and George Dvorsky. "Dignity and Agential Realism: Human, Posthuman, and Nonhuman". In: *The American Journal of Bioethics* 10.7 (2010), pp. 57–58.

[138] Phillip Atiba Goff et al. "Not yet human: implicit knowledge, historical dehumanization, and contemporary consequences." In: *Journal of personality and social psychology* 94.2 (2008), p. 292.

[139] Erving Goffman. "The nature of deference and demeanor". In: *American Anthropologist* 58.3 (1956), pp. 473–502.

[140] Erving Goffman. "Interaction ritual: Essays on face-to-face interaction." In: (1967).

[141] Erving Goffman et al. *The presentation of self in everyday life*. Harmondsworth London, 1978.

[142] Erving Goffman. *Relations in public*. Transaction Publishers, 2009.

[143] Murray Goulden et al. "Smart grids, smart users? The role of the user in demand side management". In: *Energy research & social science* 2 (2014), pp. 21–29. DOI: 10.1016/j.erss.2014.04.008.

[144] Murray Goulden et al. "Living with interpersonal data: Observability and accountability in the age of pervasive ICT". In: *New Media & Society* 20.4 (2018), pp. 1580–1599. DOI: 10.1177/1461444817700154.

[145] Kathryn Greene, Valerian J Derlega, and Alicia Mathews. "Self-disclosure in personal relationships". In: *The Cambridge handbook of personal relationships*. 2006, pp. 409–427.

[146] Kevin Grennan. "The Smell of Control: Fear, Focus, Trust". In: (2011).

[147] Sanford J Grossman and Oliver D Hart. "An analysis of the principal-agent problem". In: *Foundations of Insurance Economics*. Springer, 1992, pp. 302–340. DOI: 10.1007/978-94-015-7957-5_16.

[148] David J Gunkel. *Robot rights*. MIT Press, 2018.

[149] Stuart Hall. "The West and the rest: Discourse and power". In: *Race and Racialization, 2E: Essential Readings* (1992), pp. 85–95.

[150] Peter A Hancock et al. "A meta–analysis of factors affecting trust in human-robot interaction". In: *Human factors* 53.5 (2011), pp. 517–527. DOI: 10.1177/0018720811417254.

[151] Ellie Harmon and Melissa Mazmanian. "Stories of the Smartphone in everyday discourse: conflict, tension & instability". In: *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM. 2013, pp. 1051–1060. DOI: 10.1145/2470654.2466134.

[152] Todd Hedinger. "Assessing interpersonal respect: Its psychological and philosophical implications." In: (1992).

[153] Bill Ong Hing. "Entering the Trump ice age: Contextualizing the new immigration enforcement regime". In: *Tex. A&M L. Rev.* 5 (2017), p. 253.

[154] Donald Horton and R Richard Wohl. "Mass communication and para-social interaction: Observations on intimacy at a distance". In: *Psychiatry* 19.3 (1956), pp. 215–229. DOI: 10.1080/00332747.1956.11023049.

[155] Danny Yuxing Huang et al. "IoT Inspector: Crowdsourcing Labeled Network Traffic from SmartHome Devices at Scale". In: *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4.2 (2020). DOI: 10.1145/3397333.

[156] Stephen D Hudson. "The nature of respect". In: *Social Theory and Practice* 6.1 (1980), pp. 69–90.

[157] Paul Hunt. *Stigma: The experience of disability*. London; Dublin [etc.]: G. Chapman, 1966.

[158] Yuen J Huo, Kevin R Binning, and Ludwin E Molina. "The interplay between fairness and the experience of respect: Implications for group life". In: *Research on managing groups and teams: Fairness and groups*. Ed. by E. A. Mannix, M. A. Neale, and E. Mullen. Vol. 13. Emerald Group Publishing Limited, 2010.

[159] Yuen J Huo and Ludwin E Molina. "Is pluralism a viable model of diversity? The benefits and limits of subgroup respect". In: *Group Processes & Intergroup Relations* 9.3 (2006), pp. 359–376.

[160] Ben Hutchinson and Margaret Mitchell. "50 years of test (un) fairness: Lessons for machine learning". In: *Proceedings of the Conference on Fairness, Accountability, and Transparency*. 2019, pp. 49–58.

[161] Hilary Hutchinson et al. "Technology probes: inspiring design for and with families". In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM. 2003, pp. 17–24. DOI: 10.1145/642611.642616.

[162] Gilhwan Hwang et al. "It Sounds Like A Woman: Exploring Gender Stereotypes in South Korean Voice Assistants". In: *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM. 2019, LBW2413. DOI: 10.1145/3290607.3312915.

[163] Dominic Idier. "Science fiction and technology scenarios: comparing Asimov's robots and Gibson's cyberspace". In: *Technology in Society* 22.2 (2000), pp. 255–272. DOI: 10.1016/S0160-791X(00)00004-X.

[164] Benedicte Ingstad and Susan Reynolds Whyte. *Disability and culture*. Univ of California Press, 1995.

[165] David C Ivey, Elizabeth Wieling, and Steven M Harris. "Save the young—the elderly have lived their lives: Ageism in marriage and family therapy". In: *Family Process* 39.2 (2000), pp. 163–175.

[166] Martin J Kraemer, Ivan Flechais, and Helena Webb. "Exploring Communal Technology Use in the Home". In: *Proceedings of the Halfway to the Future Symposium 2019*. HTTF 2019. Nottingham, United Kingdom: ACM, 2019, 5:1–5:8. DOI: 10.1145/3363384.3363389.

[167] Timo Jakobi et al. "Evolving Needs in IoT Control and Accountability: A Longitudinal Study on Smart Home Intelligibility". In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2.4 (2018), p. 171. DOI: 10.1145/3287049.

[168] Sajid Javid and Jeremy Wright. *Online Harms White Paper*. Tech. rep. UK Department for Digital, Culture, Media & Sport, 2019.

[169] Nicholas Jenkins et al. "Putting it in context: the use of vignettes in qualitative interviewing". In: *Qualitative Research* 10.2 (2010), pp. 175–198. DOI: 10.1177/1468794109356737.

[170] Carlos Jensen, Colin Potts, and Christian Jensen. "Privacy practices of Internet users: Self-reports versus observed behavior". In: *International Journal of Human-Computer Studies* 63.1 (2005). HCI research in privacy and security, pp. 203–227. DOI: https://doi.org/10.1016/j.ijhcs.2005.04.019.

[171] Casper Bruun Jensen and Anders Blok. "Techno-animism in Japan: Shinto cosmograms, actor-network theory, and the enabling powers of non-human agencies". In: *Theory, Culture & Society* 30.2 (2013), pp. 84–115. DOI: 10.1177/0263276412456564.

[172] Rikke Hagensby Jensen et al. "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*. ACM. 2018, p. 4. DOI: 10.1145/3173574.3173578.

[173] L Crystal Jiang, Natalie N Bazarova, and Jeffrey T Hancock. "The disclosure–intimacy link in computer-mediated communication: An attributional extension of the hyperpersonal model". In: *Human communication research* 37.1 (2011), pp. 58–77. DOI: 10.1111/j.1468-2958.2010.01393.x.

[174] Edward Johnson. "Ignoring persons". In: *Tulane Studies in Philosophy* 31 (1982), pp. 91–105.

[175] Spike Jonze. *Her*. Warner Brothers, 2013.

[176] Takayuki Kanda, Michihiro Shimada, and Satoshi Koizumi. "Children Learning with a Social Robot". In: *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*. 2012, pp. 351–358. DOI: 10.1145/2157689.2157809.

[177] Ruogu Kang et al. ""My Data Just Goes Everywhere:" User Mental Models of the Internet and Implications for Privacy and Security". In: *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*. 2015, pp. 39–52.

[178] Immanuel Kant. *Groundwork for the Metaphysics of Morals*. Yale University Press, 2002.

[179] Thomas Kasulis. *Shinto: The Way Home*. University of Hawai'i Press, 2004.

[180] Lynne Keevers and Pamela Abuodha. "Social inclusion as an unfinished verb: A practice-based approach". In: *Journal of Academic Language and Learning* 6.2 (2012), A42–A59.

[181] Cory D Kidd et al. "The aware home: A living laboratory for ubiquitous computing research". In: *International Workshop on Cooperative Buildings*. Springer. 1999, pp. 191–198.

[182] Eunhyang Kim and Eunyoung Koh. "Avoidant attachment and smartphone addiction in college students: The mediating effects of anxiety and self-esteem". In: *Computers in Human Behavior* 84 (2018), pp. 264–271. DOI: 10.1016/j.chb.2018.02.037.

[183] Sojung Kim and Joon Sang Baek. "Definitions and Attributes of Smart Home Appliances". In: *Proceedings of the Design Society: International Conference on Engineering Design*. Vol. 1. 1. Cambridge University Press. 2019, pp. 2071–2080. DOI: 10.1017/dsi.2019.213.

[184] Jennifer King. "Taken out of context: An empirical analysis of westin's privacy scale". In: *Workshop on Privacy Personas and Segmentation*. 2014, p. 2014.

[185] Lea Kissner. "Keynote: Engineering For Respect: Building Systems For The World We Live In". Scale. 2019. URL: https://atscaleconference.com/videos/scale-2019-engineering-for-respect-building-systems-for-the-world-we-live-in/.

[186] Rob Kitchin. "Making sense of smart cities: addressing present shortcomings". In: *Cambridge Journal of Regions, Economy and Society* 8.1 (2015), pp. 131–136. DOI: 10.1093/cjres/rsu027.

[187] Predrag Klasnja et al. "Exploring privacy concerns about personal sensing". In: *International Conference on Pervasive Computing*. Springer. 2009, pp. 176–183. DOI: 10.1007/978-3-642-01516-8_13.

[188] Jonathan Klein, Youngme Moon, and Rosalind W Picard. "This computer responds to user frustration: Theory, design, and results". In: *Interacting with computers* 14.2 (2002), pp. 119–140.

[189] Mark L Knapp. *Social intercourse: From greeting to goodbye*. Allyn & Bacon, Incorporated, 1978.

[190] Mark L Knapp, Anita L Vangelisti, and John P Caughlin. "Interpersonal communication and human relationships". In: (2005).

[191] Spyros Kokolakis. "Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon". In: *Computers & Security* 64 (2017), pp. 122–134. DOI: 10.1016/j.cose.2015.07.002.

[192] Mary P Koss et al. *No safe haven: Male violence against women at home, at work, and in the community.* American Psychological Association, 1994.

[193] Martin J Kraemer, Ivan Flechais, and Helena Webb. "Exploring Communal Technology Use in the Home". In: *Proceedings of the Halfway to the Future Symposium 2019*. ACM. 2019, p. 5. DOI: 10.1145/3363384.3363389.

[194] Ponnurangam Kumaraguru and Lorrie Faith Cranor. *Privacy indexes: a survey of Westin's studies*. Tech. rep. 2005. DOI: 10.1184/R1/6625406.v1.

[195] Arun Kundnani. *The Muslims are coming!: Islamophobia, extremism, and the domestic war on terror*. Verso Trade, 2014.

[196] Matt J Kusner et al. "Counterfactual fairness". In: *Advances in Neural Information Processing Systems*. 2017, pp. 4069–4079.

[197] Mark S Lachs and Karl Pillemer. "Elder abuse". In: *The Lancet* 364.9441 (2004), pp. 1263–1272.

[198] Fritz Lang. *Metropolis*. Parufamet, 1927.

[199] Patrick M Langdon et al. *Inclusive Designing*. Springer, 2014.

[200] Susan W Langdon. "Conceptualizations of respect: Qualitative and quantitative evidence of four (five) themes". In: *The journal of psychology* 141.5 (2007), pp. 469–484.

[201] SW Langdon and W Preble. "Qualitative and quantitative definitions and experiences of respect in adolescents". In: *Poster session presented at Society for Research in Child Development conference, Tampa, FL*. 2003.

[202] Josephine Lau, Benjamin Zimmerman, and Florian Schaub. "Alexa, are you listening?: Privacy perceptions, concerns and privacy-seeking behaviors with smart speakers". In: *Proceedings of the ACM on Human-Computer Interaction* 2.CSCW (2018), p. 102.

[203] Shaun Lawson et al. "Problematising upstream technology through speculative design: the case of quantified cats and dogs". In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM. 2015, pp. 2663–2672. DOI: 10.1145/2702123.2702260.

[204] Lucian Leahu, Marisa Cohn, and Wendy March. "How categories come to matter". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM. 2013, pp. 3331–3334. DOI: 10.1145/2470654.2466455.

[205] Doohwang Lee, Robert Larose, and Nora Rifon. "Keeping our network safe: a model of online protection behaviour". In: *Behaviour & Information Technology* 27.5 (2008), pp. 445–454. DOI: 10.1080/01449290600879344.

[206] Hosub Lee and Alfred Kobsa. "Understanding user privacy in Internet of Things environments". In: *2016 IEEE 3rd World Forum on Internet of Things (WF-IoT)*. IEEE. 2016, pp. 407–412. DOI: 10.1109/WF-IoT.2016.7845392.

[207] Sunok Lee, Sungbae Kim, and Sangsu Lee. ""What Does Your Agent Look like?": A Drawing Study to Understand Users' Perceived Persona of Conversational Agent". In: *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 2019. DOI: 10.1145/3290607.3312796.

[208] Joscha Legewie. "Racial profiling and use of force in police stops: How local events trigger periods of increased discrimination". In: *American journal of sociology* 122.2 (2016), pp. 379–424.

[209] Pedro Giovanni Leon et al. "What matters to users? Factors that Affect Users' Willingness to Share Information with Online Advertisers". In: (2013), p. 1. DOI: 10.1145/2501604.2501611.

[210] Rensis Likert. "A technique for the measurement of attitudes." In: *Archives of psychology* (1932).

[211] Jialiu Lin et al. "Expectation and purpose: understanding users' mental models of mobile app privacy through crowdsourcing". In: *Proceedings of the 2012 ACM conference on ubiquitous computing*. ACM. 2012, pp. 501–510. DOI: 10.1145/2370216.2370290.

[212] Astrid Linder and Wanna Svedberg. "Occupant safety assessment in European regulatory tests: review of occupant models, gaps and suggestion for bridging any gaps". In: *18th International Conference Road Safety on Five Continents (RS5C 2018), Jeju Island, South Korea, May 16-18, 2018*. Statens väg-och transportforskningsinstitut. 2018.

[213] Joseph Galen Lindley et al. "Anticipating GDPR in Smart Homes Through Fictional Conversational Objects". In: (2017).

[214] Joseph Lindley and Paul Coulton. "Game of drones". In: *Proceedings of the 2015 annual symposium on computer-human interaction in play*. ACM. 2015, pp. 613–618. DOI: 10.1145/2793107.2810300.

[215] Joseph Lindley, Paul Coulton, and Rachel Cooper. "Why the internet of things needs object orientated ontology". In: *The Design Journal* 20.sup1 (2017), S2846–S2857. DOI: 10.1080/14606925.2017.1352796.

[216] Walter Liszewski et al. "Persons of nonbinary gender—awareness, visibility, and health disparities". In: *The New England journal of medicine* 379.25 (2018), p. 2391.

[217] Irene Lopatovska and Harriet Williams. "Personification of the Amazon Alexa: BFF or a Mindless Companion". In: *Proceedings of the 2018 Conference on Human Information Interaction&Retrieval*. ACM. 2018, pp. 265–268. DOI: 10.1145/3176349.3176868.

[218] Irene Lopatovska et al. "Talk to me: Exploring user interactions with the Amazon Alexa". In: *Journal of Librarianship and Information Science* (2018).

[219] Stephen Loughnan and Nick Haslam. "Animals and androids: Implicit associations between social categories and nonhumans". In: *Psychological Science* 18.2 (2007), pp. 116–121. DOI: 10.1111/j.1467-9280.2007.01858.x.

[220] Alan Lundgard. "Measuring justice in machine learning". In: *arXiv preprint arXiv:2009.10050* (2020).

[221] Ulrik Lyngs et al. "Self-Control in Cyberspace: Applying Dual Systems Theory to a Review of Digital Self-Control Tools". In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19. Association for Computing Machinery, 2019, pp. 1–18. DOI: 10.1145/3290605.3300361.

[222] Christopher Macleod. "John Stuart Mill". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Fall 2018. Metaphysics Research Lab, Stanford University, 2018.

[223] Maria Madsen and Shirley Gregor. "Measuring human-computer trust". In: *11th australasian conference on information systems*. Vol. 53. Citeseer. 2000, pp. 6–8.

[224] Nathan Malkin et al. ""What Can't Data Be Used For?" Privacy Expectations about Smart TVs in the US". In: *European Workshop on Usable Security (Euro USEC)*. 2018.

[225] Kate Manne. *Down girl: The logic of misogyny*. Oxford University Press, 2017.

[226] Coral Manton and Birgitte Aga. *Women Reclaiming AI*. https://womenreclaimingai.com/. 2017.

[227] Nancy J Marshall. "Privacy and environment". In: *Human ecology* 1.2 (1972), pp. 93–110. DOI: 10.1007/BF01531349.

[228] Patrizia Marti. "In Praise of Beauty and Enchantment: Setting the Scenery of Respectful Design". In: *Proceedings of the 2014 Mulitmedia, Interaction, Design and Innovation International Conference on Multimedia, Interaction, Design and Innovation*. MIDI '14. 2014, pp. 1–6. DOI: 10.1145/2643572.2643574.

[229] Kirsten E. Martin and Helen Nissenbaum. "Measuring privacy: An empirical test using context to expose confounding variables". In: *Columbia Science and Technology Law Review* 18.Fall (2016), pp. 176–218. DOI: 10.1039/C5CY01433F.

[230] Sherry A Marts and Sarah Keitt. "Foreword: a historical overview of advocacy for research in sex-based biology". In: *Advances in Molecular and Cell Biology*. Vol. 34. Elsevier, 2004, pp. v–xiii.

[231] Arunesh Mathur et al. "Dark patterns at scale: Findings from a crawl of 11K shopping websites". In: *Proceedings of the ACM on Human-Computer Interaction* 3.CSCW (2019), pp. 1–32.

[232] Tuuli Mattelmäki et al. *Design probes*. Aalto University, 2006.

[233] Michelle L Mazurek et al. "Access control for home data sharing: Attitudes, needs and practices". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM. 2010, pp. 645–654. DOI: 10.1145/1753326.1753421.

[234] Linda McDowell. "Respect, deference, respectability and place: What is the problem with/for working class boys?" In: *Geoforum* 38.2 (2007), pp. 276–286. ISSN: 0016-7185. DOI: https://doi.org/10.1016/j.geoforum.2006.09.006. URL: http://www.sciencedirect.com/science/article/pii/S0016718506001254.

[235] Athena McLean. "Ethical frontiers of ICT and older users: cultural, pragmatic and ethical issues". In: *Ethics and information technology* 13.4 (2011), pp. 313–326. DOI: 10.1007/s10676-011-9276-4.

[236] Emily McReynolds et al. "Toys that listen: A study of parents, children, and internet-connected toys". In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM. 2017, pp. 5197–5207. DOI: 10.1145/3025453.3025735.

[237] Sarah Mennicken et al. "It's like living with a friendly stranger: perceptions of personality traits in a smart home". In: *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM. 2016, pp. 120–131. DOI: 10.1145/2971648.2971757.

[238] Nick Merrill, John Chuang, and Coye Cheshire. "Sensing is Believing: What People Think Biosensors Can Reveal About Thoughts and Feelings". In: *Proceedings of the 2019 on Designing Interactive Systems Conference*. 2019, pp. 413–420. DOI: 10.1145/3322276.3322286.

[239] Markus Miettinen et al. "IoT Sentinel: Automated device-type identification for security enforcement in IoT". In: *2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*. IEEE. 2017, pp. 2177–2184. DOI: 10.1109/ICDCS.2017.283.

[240] John Stuart Mill. "On liberty". In: *A selection of his works*. Springer, 1966, pp. 1–147.

[241] Linda Miller et al. *Occupational segregation, gender gaps and skill gaps*. Citeseer, 2004.

[242] Charles W Mills. "Black Radical Kantianism". In: *Res Philosophica* (2017).

[243] Charles W Mills. *Black rights/white wrongs: The critique of racial liberalism*. Oxford University Press, 2017.

[244] Brent Mittelstadt. "Designing the health-related internet of things: ethical principles and guidelines". In: *Information* 8.3 (2017), p. 77. DOI: 10.3390/info8030077.

[245] Brent Daniel Mittelstadt et al. "The ethics of algorithms: Mapping the debate". In: *Big Data & Society* 3.2 (2016), p. 2053951716679679. DOI: 10.1177/2053951716679679.

[246] Youngme Moon. "Intimate exchanges: Using computers to elicit self-disclosure from consumers". In: *Journal of consumer research* 26.4 (2000), pp. 323–339. DOI: 10.1086/209566.

[247] Gary C Moore and Izak Benbasat. "Development of an instrument to measure the perceptions of adopting an information technology innovation". In: *Information systems research* 2.3 (1991), pp. 192–222. DOI: 10.1287/isre.2.3.192.

[248] Masahiro Mori. "The uncanny valley". In: *Energy* 7.4 (1970), pp. 33–35.

[249] Bonnie M Muir. "Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems". In: *Ergonomics* 37.11 (1994), pp. 1905–1922. DOI: 10.1080/00140139408964957.

[250] Bernard I Murstein. "Stimulus. Value. Role: A theory of marital choice". In: *Journal of Marriage and the Family* (1970), pp. 465–481. DOI: 10.2307/350113.

[251] Kevin L. Nadal, Avy Skolnik, and Yinglee Wong. "Interpersonal and Systemic Microaggressions Toward Transgender People: Implications for Counseling". In: *Journal of LGBT Issues in Counseling* 6.1 (2012), pp. 55–82. DOI: 10.1080/15538605.2012.648583.

[252] Viviane Namaste. *Invisible lives: The erasure of transsexual and transgendered people*. University of Chicago Press, 2000.

[253] Clifford Ivar Nass and Scott Brave. *Wired for speech: How voice activates and advances the human-computer relationship*. MIT press Cambridge, MA, 2005.

[254] Clifford Nass and Youngme Moon. "Machines and mindlessness: Social responses to computers". In: *Journal of social issues* 56.1 (2000), pp. 81–103. DOI: 10.1111/0022-4537.00153.

[255] Clifford Nass, Jonathan Steuer, and Ellen R. Tauber. "Computers Are Social Actors". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '94. 1994, pp. 72–78. DOI: 10.1145/191666.191703.

[256] Clifford Nass et al. "Anthropomorphism, agency, and ethopoeia: computers as social actors". In: *INTERACT'93 and CHI'93 conference companion on Human factors in computing systems*. ACM. 1993, pp. 111–112. DOI: 10.1145/259964.260137.

[257] Harold G Nelson and Erik Stolterman. "The case for design: Creating a culture of intention". In: *Educational Technology* 40.6 (2000), pp. 29–35.

[258] Todd D Nelson. "Ageism: Prejudice against our feared future self". In: *Journal of social issues* 61.2 (2005), pp. 207–221.

[259] Tommy Nilsson et al. "Breaching the future: understanding human challenges of autonomous systems for the home". In: *Personal and Ubiquitous Computing* 23.2 (2019), pp. 287–307. DOI: 10.1007/s00779-019-01210-7.

[260] Helen Nissenbaum. "Protecting privacy in an information age: The problem of privacy in public". In: *Law and philosophy* (1998), pp. 559–596.

[261] Helen Nissenbaum. "Privacy as contextual integrity". In: *Wash. L. Rev.* 79 (2004), p. 119.

[262] Helen Nissenbaum. *Privacy in context: Technology, policy, and the integrity of social life*. Stanford University Press, 2009.

[263] Patricia A Norberg, Daniel R Horne, and David A Horne. "The privacy paradox: Personal information disclosure intentions versus behaviors". In: *Journal of consumer affairs* 41.1 (2007), pp. 100–126.

[264] Don Norman. *The design of everyday things: Revised and expanded edition*. Basic books, 2013.

[265] Donald A Norman and Stephen W Draper. *User centered system design: New perspectives on human-computer interaction*. CRC Press, 1986.

[266] Shannon O'Brien and Florian 'Floyd' Mueller. "Holding Hands over a Distance: Technology Probes in an Intimate, Mobile Context". In: *Proceedings of the 18th Australia Conference on Computer-Human Interaction: Design: Activities, Artefacts and Environments*. OZCHI '06. Sydney, Australia: ACM, 2006, pp. 293–296. DOI: 10.1145/1228175.1228226.

[267] TJ OConnor et al. "HomeSnitch: behavior transparency and control for smart home IoT devices". In: *Proceedings of the 12th Conference on Security and Privacy in Wireless and Mobile Networks*. ACM. 2019, pp. 128–138. DOI: 10.1145/3317549.3323409.

[268] Kieron OHara. "The seven veils of privacy". In: *IEEE Internet Computing* 20.2 (2016), pp. 86–91. DOI: 10.1109/MIC.2016.34.

[269] M Omie and Howard Winant. *Racial Formation in the United States: From the 1960's to the 1980's*. 1986.

[270] Antti Oulasvirta et al. "Long-term effects of ubiquitous surveillance in the home". In: *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. ACM. 2012, pp. 41–50. DOI: 10.1145/2370216.2370224.

[271] Leysia Palen and Paul Dourish. "Unpacking privacy for a networked world". In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM. 2003, pp. 129–136. DOI: 10.1145/642611.642635.

[272] Mika Pantzar. "Domestication of everyday life technology: dynamic views on the social histories of artifacts". In: *Design Issues* 13.3 (1997), pp. 52–65.

[273] Karen Parkhill et al. "Transforming the UK energy system: public values, attitudes and acceptability: synthesis report". In: (2013).

[274] Dino Pedreshi, Salvatore Ruggieri, and Franco Turini. "Discrimination-aware data mining". In: *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2008, pp. 560–568.

[275] Caroline Criado Perez. *Invisible women: Exposing data bias in a world designed for men*. Random House, 2019.

[276] James Petty. "The London spikes controversy: Homelessness, urban securitisation and the question of 'hostile architecture'". In: *International Journal for Crime, Justice and Social Democracy* 5.1 (2016), p. 67.

[277] Rosalind W Picard. *Affective computing*. MIT press, 2000.

[278] Rosalind W Picard. "Affective computing: challenges". In: *International Journal of Human-Computer Studies* 59.1-2 (2003), pp. 55–64. DOI: 10.1016/S1071-5819(03)00052-1.

[279] James Pierce. "Smart Home Security Cameras and Shifting Lines of Creepiness: A Design-Led Inquiry". In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM. 2019, p. 45. DOI: 10.1145/3290605.3300275.

[280] James Pierce and Carl DiSalvo. "Addressing network anxieties with alternative design metaphors". In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 2018, pp. 1–13. DOI: 10.1145/3173574.3174123.

[281] Peter Pirolli and Stuart Card. "The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis". In: *Proceedings of international conference on intelligence analysis*. Vol. 5. McLean, VA, USA. 2005, pp. 2–4.

[282] Isabella Poggi and Francesca D'Errico. "Social signals: A psychological perspective". In: *Computer analysis of human behavior*. Springer, 2011, pp. 185–225. DOI: 10.1007/978-0-85729-994-9_8.

[283] Martin Porcheron et al. "Voice interfaces in everyday life". In: *proceedings of the 2018 CHI conference on human factors in computing systems*. ACM. 2018, p. 640. DOI: 10.1145/3173574.3174214.

[284] Alisha Pradhan, Leah Findlater, and Amanda Lazar. "Phantom Friend or Just a Box with Information: Personification and Ontological Categorization of Smart Speaker-based Voice Assistants by Older Adults". In: *Proceedings of the ACM on Human-Computer Interaction* 3.CSCW (2019), p. 214. DOI: 10.1145/3359316.

[285] Alisha Pradhan, Kanika Mehta, and Leah Findlater. ""Accessibility Came by Accident": Use of Voice-Controlled Intelligent Personal Assistants by People with Disabilities". In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 2018, pp. 1–13. DOI: 10.1145/3173574.3174033.

[286] Mark Pragnell, Roger Moore, and Lorna Spence. *The market potential for Smart Homes*. York Publishing Services York, 2000.

[287] William L Prosser. "Privacy, 48 Calif". In: *L. Rev* 383.10.2307 (1960), p. 3478805383.

[288] Amanda Purington et al. "Alexa is my new BFF: social roles, user satisfaction, and personification of the amazon echo". In: *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM. 2017, pp. 2853–2859. DOI: 10.1145/3027063.3053246.

[289] Byron Reeves and Clifford Ivar Nass. *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge university press, 1996.

[290] Jingjing Ren et al. "Information Exposure From Consumer IoT Devices: A Multidimensional, Network-Informed Measurement Approach". In: *Proceedings of the Internet Measurement Conference*. IMC '19. Amsterdam, Netherlands: ACM, 2019, pp. 267–279. DOI: 10.1145/3355369.3355577.

[291] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. "Why Should I Trust You?: Explaining the Predictions of Any Classifier". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM. 2016, pp. 1135–1144. DOI: 10.1145/2939672.2939778.

[292] Neil M Richards and Daniel J Solove. "Prosser's privacy law: A mixed legacy". In: *Calif. L. Rev.* 98 (2010), p. 1887.

[293] Tom A Rodden et al. "At home with agents: exploring attitudes towards future smart energy infrastructures". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM. 2013, pp. 1173–1182. DOI: 10.1145/2470654.2466152.

[294] Jon Rogers et al. "Our Friends Electric: Reflections on Advocacy and Design Research for the Voice Enabled Internet". In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19. 2019, 114:1–114:13. DOI: 10.1145/3290605.3300344.

[295] S. Šabanovič et al. "PARO robot affects diverse interaction modalities in group sensory therapy for older adults with dementia". In: *2013 IEEE 13th International Conference on Rehabilitation Robotics (ICORR)*. 2013, pp. 1–6.

[296] Edward W Said. *Orientalism*. Vintage, 1979.

[297] Maha Salem et al. "Would you trust a (faulty) robot?: Effects of error, task type and personality on human-robot cooperation and trust". In: *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*. ACM. 2015, pp. 141–148. DOI: 10.1145/2696454.2696497.

[298] Werner Schirmer, Linda Weidenstedt, and Wendelin Reich. "Respect and agency: An empirical exploration". In: *Current Sociology* 61.1 (2013), pp. 57–75. DOI: 10.1177/0011392111421531.

[299] Ari Schlesinger, W. Keith Edwards, and Rebecca E. Grinter. "Intersectional HCI: Engaging Identity through Gender, Race, and Class". In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 2017, pp. 5412–5427. DOI: 10.1145/3025453.3025766.

[300] Paul M Schwartz. "Privacy and democracy in cyberspace". In: *Vand. L. Rev.* 52 (1999), p. 1607.

[301] Alex Sciuto et al. "Hey Alexa, What's Up?: A Mixed-Methods Studies of In-Home Conversational Agent Usage". In: *Proceedings of the 2018 on Designing Interactive Systems Conference 2018*. ACM. 2018, pp. 857–868. DOI: 10.1145/3196709.3196772.

[302] William Seymour and Max Van Kleek. "Does Siri Have a Soul? Exploring Voice Assistants Through Shinto Design Fictions". In: *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI EA '20. New York, NY, USA: Association for Computing Machinery, 2020, pp. 1–12. DOI: 10.1145/3334480.3381809.

[303] William Seymour et al. "Strangers in the Room: Unpacking Perceptions of 'Smartness' and Related Ethical Concerns in the Home". In: *Proceedings of the 2020 ACM Conference on Designing Interactive Systems*. 2020. DOI: 10.1145/3357236.3395501.

[304] Irina Shklovski et al. "Leakiness and Creepiness in App Space: Perceptions of Privacy and Mobile App Use". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '14. Toronto, Ontario, Canada: Association for Computing Machinery, 2014, pp. 2347–2356. ISBN: 9781450324731. DOI: 10.1145/2556288.2557421. URL: https://doi.org/10.1145/2556288.2557421.

[305] John Shotter. "Agential realism, social constructionism, and our living relations to our surroundings: Sensing similarities rather than seeing patterns". In: *Theory & Psychology* 24.3 (2014), pp. 305–325.

[306] Bernd Simon and Stefan StÃŒrmer. "Respect for Group Members: Intragroup Determinants of Collective Identification and Group-Serving Behavior". In: *Personality and Social Psychology Bulletin* 29.2 (2003). PMID: 15272946, pp. 183–193. DOI: 10.1177/0146167202239043. eprint: https://doi.org/10.1177/0146167202239043. URL: https://doi.org/10.1177/0146167202239043.

[307] Daniel J Solove. "Conceptualizing privacy". In: *Calif. L. Rev.* 90 (2002), p. 1087.

[308] Daniel J. Solove. *Understanding privacy*. Harvard University Press, 2008.

[309] Victoria Song. *Sleep Number Denies Recording Users in Their Beds, Calls Creepy Privacy Policy an 'Error'*. 2019. URL: https://gizmodo.com/sleep-number-denies-recording-users-in-their-beds-call-1830775153 (visited on 09/01/2019).

[310] Katta Spiel. "Body-Positive Computing as a Means to Counteract Normative Biases in Fitness Trackers". In: *XRDS* 25.4 (July 2019), pp. 34–37. ISSN: 1528-4972. DOI: 10.1145/3331065. URL: https://doi.org/10.1145/3331065.

[311] Katta Spiel et al. "Fitter, Happier, More Productive? The Normative Ontology of Fitness Trackers". In: *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI EA '18. Montreal QC, Canada: Association for Computing Machinery, 2018, pp. 1–10. ISBN: 9781450356213. DOI: 10.1145/3170427.3188401. URL: https://doi.org/10.1145/3170427.3188401.

[312] Ranajit Guha Gayatri Chakravorty Spivak, Edward Said, et al. *Selected subaltern studies*. Oxford University Press, 1988.

[313] Aaron Springer and Henriette Cramer. ""Play PRBLMS": Identifying and Correcting Less Accessible Content in Voice Interfaces". In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 2018, pp. 1–13. DOI: 10.1145/3173574.3173870.

[314] Nick Statt. "Google now says controversial AI voice calling system will identify itself to humans". In: *The Verge*. 2018. URL: https://www.theverge.com/2018/5/10/17342414/google-duplex-ai-assistant-voice-calling-identify-itself-update.

[315] Joseph Steinberg. *These Devices May Be Spying On You (Even In Your Own Home)*. 2019. URL: https://www.forbes.com/sites/josephsteinberg/2014/01/27/these-devices-may-be-spying-on-you-even-in-your-own-home/#184d7429b859 (visited on 09/05/2019).

[316] Henri-Jacques Stiker. "Corps infirmes et sociétés [Disabled Bodies and Society]". In: *Paris: Aubier Montaigne* (1982).

[317] Kevin M. Storer, Tejinder K. Judge, and Stacy M. Branham. ""All in the Same Boat": Tradeoffs of Voice Assistant Ownership for Mixed-Visual-Ability Families". In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 2020, pp. 1–14. DOI: 10.1145/3313831.3376225.

[318] Lucy Suchman and Lucy A Suchman. *Human-machine reconfigurations: Plans and situated actions*. Cambridge university press, 2007.

[319] J. P. Sullins. "Robots, Love, and Sex: The Ethics of Building a Love Machine". In: *IEEE Transactions on Affective Computing* 3.4 (2012), pp. 398–409.

[320] Madiha Tabassum, Tomasz Kosinski, and Heather Richter Lipford. "I don't own the data": End User Perceptions of Smart Home Device Data Practices and Risks". In: *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)*. 2019.

[321] Monika Taddicken. "The 'privacy paradox' in the social web: The impact of privacy concerns, individual characteristics, and the perceived social relevance on different forms of self-disclosure". In: *Journal of Computer-Mediated Communication* 19.2 (2014), pp. 248–273.

[322] Alex S. Taylor et al. "Homes That Make Us Smart". In: *Personal Ubiquitous Comput.* 11.5 (June 2007), pp. 383–393. ISSN: 1617-4909. DOI: 10.5555/1265387.1265392.

[323] Dalmas A Taylor and Irwin Altman. "Intimacy-scaled stimuli for use in studies of interpersonal relations". In: *Psychological Reports* 19.3 (1966), pp. 729–730. DOI: 10.2466/pr0.1966.19.3.729.

[324] Stella Ting-Toomey. *The challenge of facework: Cross-cultural and interpersonal issues*. SUNY Press, 1994.

[325] Milka Trajkova and Aqueasha Martin-Hammond. ""Alexa is a Toy": Exploring Older Adults' Reasons for Using, Limiting, and Abandoning Echo". In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 2020, pp. 1–13. DOI: 10.1145/3313831.3376760.

[326] Rahmadi Trimananda et al. "PingPong: Packet-Level Signatures for Smart Home Device Events". In: *arXiv preprint arXiv:1907.11797* (2019).

[327] Sherry Turkle. "Authenticity in the age of digital companions". In: *Interaction studies* 8.3 (2007), pp. 501–517.

[328] Alessandro Ubaldi et al. "Puffy, a Friendly Inflatable Social Robot". In: *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. 2018. DOI: 10.1145/3170427.3186595.

[329] *Update report into adtech and real time bidding*. Tech. rep. UK Informtion Commissioner's Office, 2019.

[330] Blase Ur, Jaeyeon Jung, and Stuart Schechter. "Intruders versus intrusiveness: teens' and parents' perspectives on home-entryway surveillance". In: *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM. 2014, pp. 129–139. DOI: 10.1145/2632048.2632107.

[331] Blase Ur et al. "Smart, useful, scary, creepy: perceptions of online behavioral advertising". In: *proceedings of the eighth symposium on usable privacy and security*. ACM. 2012, p. 4.

[332] Lachlan Urquhart, Dominic Reedman-Flint, and Natalie Leesakul. "Responsible domestic robotics: exploring ethical implications of robots in the home". In: *Journal of Information, Communication and Ethics in Society* (2019). DOI: 10.1108/JICES-12-2018-0096.

[333] Lachlan Urquhart and Tom Rodden. "New directions in information technology law: learning from human–computer interaction". In: *International Review of Law, Computers & Technology* 31.2 (2017), pp. 150–169. DOI: 10.1080/13600869.2017.1298501.

[334] Teun A Van Dijk. "Discourse and the denial of racism". In: *Discourse & society* 3.1 (1992), pp. 87–118.

[335] Max Van Kleek et al. "Computationally mediated pro-social deception". In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM. 2016, pp. 552–563. DOI: 10.1145/2858036.2858060.

[336] Max Van Kleek et al. "Better the devil you know: Exposing the data sharing practices of smartphone apps". In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM. 2017, pp. 5208–5220. DOI: 10.1145/3025453.3025556.

[337] Max Van Kleek et al. "X-ray refine: Supporting the exploration and refinement of information exposure resulting from smartphone apps". In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM. 2018, p. 393. DOI: 10.1145/3173574.3173967.

[338] Aimee Van Wynsberghe. "Designing robots for care: Care centered value-sensitive design". In: *Science and engineering ethics* 19.2 (2013), pp. 407–433.

[339] Michael Veale, Max Van Kleek, and Reuben Binns. "Fairness and Accountability Design Needs for Algorithmic Support in High-Stakes Public Sector Decision-Making". In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI '18. New York, NY, USA: Association for Computing Machinery, 2018, pp. 1–14. ISBN: 9781450356206. DOI: 10.1145/3173574.3174014.

[340] Thurid Vogt, Elisabeth André, and Nikolaus Bee. "EmoVoice — A Framework for Online Recognition of Emotions from Voice". In: *Perception in Multimodal Dialogue Systems*. Ed. by Elisabeth André et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 188–199. ISBN: 978-3-540-69369-7.

[341] Sandra Wachter. "Normative challenges of identification in the Internet of Things: Privacy, profiling, discrimination, and the GDPR". In: *Computer Law & Security Review* 34.3 (2018), pp. 436–449. DOI: 10.1016/j.clsr.2018.02.002.

[342] Samuel Warren and Louis Brandeis. "The right to privacy". In: *Harvard law review* 4.5 (1890), pp. 193–220.

[343] Rick Wash and Emilee Rader. "Too Much Knowledge? Security Beliefs and Protective Behaviors Among United States Internet Users". In: *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*. Ottawa: USENIX Association, July 2015, pp. 309–325.

[344] Hue Watson et al. ""We Hold Each Other Accountable": Unpacking How Social Groups Approach Cybersecurity and Privacy Together". In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 2020, pp. 1–12.

[345] Joseph Weizenbaum. "ELIZA—a Computer Program for the Study of Natural Language Communication between Man and Machine". In: *Commun. ACM* 9.1 (Jan. 1966), pp. 36–45. ISSN: 0001-0782. DOI: 10.1145/365153.365168. URL: https://doi.org/10.1145/365153.365168.

[346] S-A Welch and Rebecca B Rubin. "Development of relationship stage measures". In: *Communication Quarterly* 50.1 (2002), pp. 24–40. DOI: 0.1080/01463370209385644.

[347] Mark West, Rebecca Kraut, and Han Ei Chew. "I'd blush if I could: closing gender divides in digital skills through education". In: (2019).

[348] Alan F. Westin. "Privacy and Freedom." In: *American Sociological Review* 33.1 (1968), p. 173. DOI: 10.2307/2092293.

[349] Wikipedia contributors. *Social Exclusion — Wikipedia, The Free Encyclopedia*. [Online; accessed 10-Sept-2020]. 2020. URL: https://en.wikipedia.org/wiki/Social_exclusion.

[350] Yorick Wilks. *Close engagements with artificial companions: key social, psychological, ethical and design issues*. Vol. 8. John Benjamins Publishing, 2010.

[351] Bernard Williams. *Ethics and the Limits of Philosophy*. Taylor & Francis, 2011.

[352] Charlie Wilson, Tom Hargreaves, and Richard Hauxwell-Baldwin. "Smart homes and their users: a systematic analysis and key challenges". In: *Personal and Ubiquitous Computing* 19.2 (2015), pp. 463–476. DOI: 10.1007/s00779-014-0813-0.

[353] Langdon Winner. "Do artifacts have politics?" In: *Daedalus* (1980), pp. 121–136.

[354] Marisol Wong-Villacres et al. "Designing for Intersections". In: *Proceedings of the 2018 Designing Interactive Systems Conference*. DIS '18. 2018, pp. 45–58. DOI: 10.1145/3196709.3196794.

[355] Allison Woodruff, Sally Augustin, and Brooke Foucault. "Sabbath Day Home Automation: "It's like Mixing Technology and Religion"". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2007, pp. 527–536. DOI: 10.1145/1240624.1240710.

[356] Allison Woodruff et al. "Would a Privacy Fundamentalist Sell Their DNA for $1000... If Nothing Bad Happened as a Result? The Westin Categories, Behavioral Intentions, and Consequences". In: *10th Symposium On Usable Privacy and Security (SOUPS 2014)*. 2014, pp. 1–18.

[357] Heng Xu, Robert E Crossler, and France BéLanger. "A value sensitive design investigation of privacy enhancing tools in web browsers". In: *Decision support systems* 54.1 (2012), pp. 424–433.

[358] Ying Xu and Mark Warschauer. "What Are You Talking To?: Understanding Children's Perceptions of Conversational Agents". In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 2020, pp. 1–13. DOI: 10.1145/3313831.3376416.

[359] Yaxing Yao et al. "Defending My Castle: A Co-Design Study of Privacy Mechanisms for Smart Homes". In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM. 2019, p. 198. DOI: 10.1145/3290605.3300428.

[360] Yaxing Yao et al. "Privacy Perceptions and Designs of Bystanders in Smart Homes". In: *Proceedings of the ACM on Human-Computer Interaction* 3.CSCW (2019), pp. 1–24.

[361] JoAnne Yates and Wanda J Orlikowski. "Genres of organizational communication: A structurational approach to studying communication and media". In: *Academy of management review* 17.2 (1992), pp. 299–326.

[362] Iris Marion Young. "Five faces of oppression". In: *Rethinking power* (2014), pp. 174–95.

[363] Karolina Zawieska, Brian R Duffy, and A Strońska. "Understanding anthropomorphisation in social robotics". In: *Pomiary Automatyka Robotyka* 16.11 (2012), pp. 78–82.

[364] Leslie A Zebrowitz and Joann M Montepare. "" Too young, too old": Stigmatizing adolescents and elders." In: (2000).

[365] Eric Zeng, Shrirang Mare, and Franziska Roesner. "End user security & privacy concerns with smart homes". In: *Symposium on Usable Privacy and Security (SOUPS)*. 2017.

[366] Wei Zhang et al. "HoMonit: Monitoring smart home apps from encrypted traffic". In: *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*. ACM. 2018, pp. 1074–1088. DOI: 10.1145/3243734.3243820.

[367] Serena Zheng et al. "User Perceptions of Smart Home IoT Privacy". In: *Proc. ACM Hum.-Comput. Interact.* Vol. 2. 2018, 200:1–200:20. DOI: 10.1145/3274469.

[368] Irving Kenneth Zola. *Missing pieces: A chronicle of living with a disability*. Temple University Press, 1982.

[369] Shoshana Zuboff. *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. Profile Books, 2019.