Meaningful Human Control over Smart Home Systems: A Value Sensitive Design Approach

Steven Umbrello†
steven.umbrello@unito.it

ABSTRACT

The last decade has witnessed the mass distribution and adoption of smart home systems and devices powered by artificial intelligence systems ranging from household appliances like fridges and toasters to more background systems such as air and water quality controllers. The pervasiveness of these sociotechnical systems makes analyzing their ethical implications necessary during the design phases of these devices to ensure not only sociotechnical resilience, but to design them for human values in mind and thus preserve meaningful human control over them. This paper engages in a conceptual investigation of how meaningful human control over smart home devices can be attained through design. The value sensitive design (VSD) approach is proposed as a way of attaining this level of control. In the proposed framework, values are identified and defined, stakeholder groups are investigated and brought into the design process and the technical constraints of the technologies in question are considered. The paper concludes with some initial examples that illustrate a more adoptable way forward for both ethicists and engineers of smart home devices.

1. Introduction

In the weeks following his purchase of the Ring internet-connected security camera (an Amazon subsidiary) in July of 2019 Alabama man John Baker Orange became the victim of a strange cybersecurity breach, one that unfortunately is becoming ever more common (Noor, 2019). Following his lawsuit against the company, it was revealed that the incident began when a strange voice was heard coming through the microphone of the doorbell camera commenting on Mr. Orange’s children who were playing basketball in front of the house at the time (Paul, 2019). The Lawsuit claimed that “unfortunately, Ring did not fulfill its core promise of providing privacy and security for its customers as its camera systems are fatally flawed,” noting that the company did not implement two-factor authentication nor put requirements in place to

† Institute for Ethics and Emerging Technologies; University of Turin (Consorzio FINO), Italy.
encourage users to create complicated, and thus harder to crack passwords. The
cases similar to Mr. Oranges are becoming more commonplace as these types of
home technologies become normalized and distributed in our societies.

The marketplace during the course of the last decade has witnessed a mass
influx of new ‘smart’ and connected technologies commonly referred to as
‘smart home technologies’. The technologies that constitute these devices form
part of the larger sociotechnical landscape of the Internet of Things (IoT) in
which computing devices are interconnected via the internet and embedded in
everyday devices such as thermostats, refrigerators, light bulbs, vacuums, and
power outlets, just to name a few (ZionMarketResearch, 2018). The
interconnection and ability to gather and transmit data between devices and
online provides both boons and perils, implicating fundamental human values
such as data privacy, security, property ownership, autonomy, trust, and
accountability among others (Friedman, Kahn, Borning, & Huldtgren, 2013;
Landauer, Prabhu, & Helander, 1998). The mass adoption of these
technologies signifies that, for the most part, they are becoming more
ubiquitous and commonplace in our everyday lives. The consequence being that
as we become more dependent on their constant use they subsequently become
more entrenched in our social and technical infrastructures. This means that
technologies are not discrete entities that are independent of their social
contexts, but rather, the social contexts in which they develop constrain and
support certain design decisions and, as a consequence, support or constrain
our dependency on them (think of smartphones for example and our
dependency on them and their functions).

Similarly, there are numerous examples of how these commonplace
technologies can become recalcitrant, making us question their safety and
usability among other factors. From autonomous cars to algorithmic trading,
these systems are not only becoming normalized, but many of the fundamental
values that underly their design and use are emerging, making themselves more
explicit. To that end, how humans can design these technologies to ensure that
they are aligned with our values and remain under meaningful human control
(MHC) has become a question of particular relevance over the last decade.
Although originating within the debate over the ethics and legality of lethal
autonomous weapons, MHC has since evolved as a concept that can be applied
to autonomous systems in general. Filippo Santoni de Sio and Jeroen van den
Hoven, in their seminal paper on MHC, provide a philosophical account of how
autonomous systems can be designed in such a way as to make MHC possible
Their account of MHC, as well as Santoni de Sio et al.’s subsequent papers that expand this notion, are discussed in greater detail in the proceeding section, emerges from, and is in line with, a principled approach to accounting for human values in design called value sensitive design (VSD). This approach, which has become an increasingly popular framework over the last two decades is predicated on eliciting stakeholder values and casting such values as technical design requirements that designers and engineers can use to build systems (Friedman, Hendry, & Borning, 2017; Winkler & Spiekermann, 2018).

It is important then to take a closer look at how these technologies can be designed to account for the values of stakeholders. Stakeholder value elicitation becomes an important practice given that not only is the individual stakeholder impacted by the design and use of these technologies, but also the larger sociotechnical infrastructure in which they are being developed and embedded, a fundamental precept of the VSD approach. To this end, this paper’s aim is threefold: (1) to provide a working definition for what it means to have MHC over smart home technologies, (2) to propose the VSD methodology as a principled approach to attaining MHC and (3) to provide a preliminary set of values and design requirements that both design practitioners and engineering ethicists can use to further this research.

To the best of the author’s knowledge, this is the first paper to adopt the VSD methodology to investigate how MHC can be attained over smart home devices. Previous studies have looked at how VSD could have been used to design technologies that went awry. To illustrate, VSD has been applied to existing technologies such as energy systems (Mouter, de Geest, & Doorn, 2018; Oosterlaken, 2015), mobile phone usage (Woelfer, Iverson, Hendry, Friedman, & Gill, 2011), care robotics (van Wynsberghe, 2012), mHelath solutions (Mueller, Heger, & Nichaves, 2018), architectural projects (van den Hoven, 2013), and even web browsers (Friedman, Howe, & Felten, 2002). The approach has also been used to explore how to design speculative future technologies as well. These exploratory applications have been undertaken for technologies such as nanopharmaceuticals (Timmermans, Zhao, & van den Hoven, 2011), atomically precise manufacturing (Umbrello, 2019), artificial

1 In their seminal paper on MHC (discussed in greater detail in Section 2), Santoni de Sio and van den Hoven (2018, p. 3) explicitly state that their conception of MHC is inspired by and in line with the VSD approach. And thus that the description of MHC they propose, in order to be salient, must be cast in terms of tenable design requirements.
intelligent systems (Umbrello & De Bellis, 2018), and, less futuristic, autonomous vehicles (Calvert, Mecacci, Heikoop, & de Sio, 2018; Mecacci & de Sio, 2019).

For this reason, this paper is comparatively unique in its approach, rather than providing a solely speculative application of the VSD methodology to the individual technologies that converge to form the IoT paradigm, it explores the appropriateness of VSD as a principled approach to informing the work of engineers in how they can design these smart home technologies, or update current software, to permit MHC to be attained. Hence, this paper evaluates how the adoption of VSD can be used to guide the development of smart home technologies that are designed for human values that are necessary for satisfying the conditions that characterize MHC. It takes as its model the paper by Giulio Mecacci and Filippo Santoni de Sio titled meaningful human control as reasoning: the case of dual-mode vehicles and aims to build on it, furthering its applicable scope (Mecacci & de Sio, 2019)

The remainder of this paper is organized as follows. In Section 2, the concept of MHC is expounded discussing in-depth its tracking and tracing conditions as well as the related concepts of distal and proximal reasoning. In Section 3, the methodological approach for how to design smart home technologies for human values will be laid out by introducing the VSD approach to technology design using a case study as an illustration. Section 4 illustrates how the VSD approach is optimal to satisfy the conditions necessary to attain sufficient MHC. Section 5 discusses the contributions of this paper, its limitations, as well as suggests directions for future research. Section 6 concludes the paper.

2. Meaningful Human Control and How to Attain It

The concept of Meaningful Human Control (MHC) was first conceived of in the still hotly debated discourse on lethal autonomous weapon systems (Canellas & Haga, 2015; Crootof, 2016; Roff & Moyes, 2016). Within this debate, MHC is attained when human agents have direct operational command over a system, guiding its actions and thus producing what is supposed to be a direct causal responsibility (Crootof, 2016). Recently, in their seminal paper Meaningful human control over autonomous systems: a philosophical account (2018) Santoni de Sio and van den Hoven stray from the existent accounts of MHC, instead providing a philosophical account of MHC by defining it as a co-variance between the systems behavior and the relevant agent(s) decisional intentions
and reasons to act. This approach is directly in line with and emerging from responsible design practices, most saliently value sensitive design (VSD) (Filippo Santoni de Sio & van den Hoven, 2018). Consequentially, this means that systems can be designed in ways that permits agents to forfeit some of their direct operational control while still possessing global control of the system, meaning that more, not less levels of autonomy (in certain cases) may permit more salient control of a system. Santoni de Sio and van den Hoven (2018) provide the timely example of autonomous vehicles (self-driving cars), where users can have overall control of the autonomous system even though the system can conceivably put the user in unforeseen and potentially threatening conditions. Similarly, attaining MHC in this sense allows for more clear lines of accountability to be drawn then when humans remain ‘in-the-loop’ of a system given that tracking the relevant reasons behind an agents decisions are a necessary condition for MHC.

Their approach to tackling MHC is novel given that it is comprehensive in its scope, looking not only at discrete systems, but rather the entire sociotechnical infrastructure of which these systems form a part. This means that although the specific design and deployment of systems implicate important factors in understanding MHC, they cannot be understood in isolation from the infrastructures, organizations, and other agents that are inextricably connected to their design, deployment and use. The approach is similarly novel given that it frames MHC as being able to be designed for by engineers, that is, as technical design requirements, not only for the system itself, but the sociotechnical infrastructures as well. In order to achieve this, however, two conditions must be met, the tracking and tracing conditions. As we shall see below, satisfying these two conditions allows for a more expansive, and comprehensive notion of meaningful human control, beyond that of solely users, that permits agents such as designers and policymakers, as well as organizations and states a level of meaningful control and thus clearer lines of attributing responsibility.

2.1. Tracking and Tracing Conditions

Building off Fischer and Ravizza’s (2000) concept of reason-responsiveness in their theory of moral responsibility, Santoni de Sio and Van den Hoven propose two conditions necessary for attaining MHC, the tracking and tracing conditions (Filippo Santoni de Sio & van den Hoven, 2018). The tracking
Meaningful Human Control over Smart Home Systems

condition deals with how responsive a system is to the actions consequent of human reasons.\(^2\) It is more comprehensively defined as:

*First necessary condition of meaningful human control.* In order to be under meaningful human control, a decision-making system should demonstrably and verifiably be responsive to the human moral reasons relevant in the circumstances—no matter how many system levels, models, software, or devices of whatever nature separate a human being from the ultimate effects in the world, some of which may be lethal. That is, decision-making systems should track (relevant) human moral reasons. (Santoni de Sio & van den Hoven, 2018, p. 7)

Hence, in order for a (semi-)autonomous system to satisfy the tracking condition, its behaviour must map onto the reasons (intentions, plans, objectives, etc.) of the relevant human agent(s) for undertaking or abstaining from any action. The tracking condition, then, is conditional on determinant design requirements. It requires that a smart home climate control system (e.g., automated thermostat) should be designed in such a way that, taking into account all accessible relevant input, its behaviour should be able to follow human reasons to act (or not act) as much as it is technically capable. If a system’s behavior is able to coherently co-vary with that of agents’ (moral) reasoning, then it, under this condition, be said to be under MHC.

The tracing condition differs in that it examines if it is possible to distinguish the human agent(s) along the system’s design and deployment history (e.g., smart lock designers, manufacturers, users, etc.), who are capable of: (1) understanding the system’s potential and (2) can recognize their moral responsibility of a system’s deployment and use (i.e., liability of moral consequence). Santoni de Sio and van den Hoven more thoroughly define tracing as:

*Second necessary condition of meaningful human control:* in order for a system to be under meaningful human control, its actions/states should be traceable to a proper moral understanding on the part of one or more relevant human persons who design or interact with the system, meaning that there is at least one human agent in the design history or use context involved in designing, programming, operating and deploying the autonomous system who (a) understands or is in the

\(^2\) The use of the term ‘reasons’ here is understood as any element that can both prompt and demonstrate human behavior, such as objectives, programs and strategies.
position to understand the capabilities of the system and the possible effects in the world of its use; (b) understands or is in the position to understand that others may have legitimate moral reactions toward them because of how the system affects the world and the role they occupy. (Santoni de Sio & van den Hoven, 2018, p. 9)

MHC then is attained by agents who can satisfy both of these conditions, only then can they be said to have MHC over a system. Smart home devices, then, can \emph{prima facie} be under MHC by (an) agent(s) if they are designed as to support as much as possible the values of accessibility and explicable (explainability and transparency) as they manifest in the system’s behaviours. If a system is able to explain its internal decision making (explicity) and such systems are themselves transparent (also a factor of expansibility) then such systems can, at least in theory, be more easily brought under MHC given that (an) agent(s) understanding of the system’s use and deployment can be more easily attributed to the system’s design architecture (tracing back to the designer as having receiver-contextualized explanation).

With these two necessary conditions, MHC ultimately entails a definition of control that is more nuanced as well as more stringent than operational control, where direct full control is demanded. What makes it more stringent than direct control is that it precludes the attribution of human control to systems just because they have an agent ‘in-the-loop’ (such as smart lighting systems). A user of a smart personal home assistant (e.g., Amazon Echo, L.U.C.Y., Moorebot, Google Home, Zenbo), even if they use their voice to initiate commands or have their hands on their smart phone to control their devices via an app, does not mean that they are necessarily equipped to understand why their device does what it does. This is the \emph{black boxing} that many technologies are subject to. Meaning that the technical infrastructure of a system can make its inner workings opaque to the user. In such cases, MHC by the end user cannot be attained because the tracing condition would not be fulfilled on account of a systems opacity. However, this does not preclude MHC from the system. Other agents (e.g., designers, help desk workers, etc.) may very well understand what is going on in the ‘black box’ (although not always). If the system successfully tracks these agents’ reasons, and they are responsible for and capable of understanding the behavior that the system exhibits based on that tracking, and for the way it acts based on its tracking of more proximal reasons (discussed
below) by the end, responsibility could be attributed to these agents (i.e., these agents satisfied the conditions for MHC).

This understanding of MHC is similarly more comprehensive than that of direct operational control given that it permits (not a necessary condition) the inclusion of supervisory control, which sanctions the user to supervise an (semi-)autonomous system that is in operational control, yet still permit that user to intervene in its operation if necessary. Furthermore, as already mentioned, this form of direct supervisory control is not a necessary condition for possessing MHC. A (fully)autonomous home assistant can, in principle, be precise, comprehensive, and transparent in tracking the reasons behind a human agent’s decisions in lieu of the ability for human agents to intervene in operations, thus still preserving the condition for their MHC.3

2.2 Distal and Proximal Reasoning

Adopted from the philosophy of intent and action (Bratman, 1984; Mele & William, 1992), Santoni de Sio and van den Hoven’s conception of a(n) agent(s) reasons is further developed, helping to not only specify types of reasons in complex systems, but also to better understand the inner workings of the tracking condition (Calvert et al., 2018). Calvert et al. (2018) began by developing two types of reasons: distal and proximal. Proximal reasons are those intentions which adjoin an action in a temporally immediate way (concurrent) such as the intention to open/close an app, to dim the lights or to adjust the air temperature. Distal reasons are longer term intentions or objectives that are formulated in a less immediate way. A users’ distal reason for example to use a smart home assistant is to make managing a household more convenient and thus less time consuming. Whereas a company’s or programmers distal reasons may be for the system to adhere to certain societal norms or laws (i.e., not permitting a smart home assistant to share data subject information).

3 Shifting traditionally held notions of accountability as a function of the end user to other relevant moral agents within the design history and use of the system.
Table 1 example of distal and proximal reasons with regards to smart home assistants

The distal reasons are the overarching intentions that the relevant agent(s) will have for the desired operations of a system. The concept of direct operational control is naturally aligned and sensitive to proximal reasons in which a system functions as a consequence of the immediate, concurrent intentions of the human agent, in most cases these will be the end users who are in proximity to the use of the system. With a traditional home climate control system, if the user does not augment the currently set temperature of the system it is because they had no intention in that instant to do so (they could have forgotten or got occupied with some other task). Because traditional systems like these are – the best extent possible – under the influence of the proximal reasons of their human users then those users are causally responsible for their use and consequent impacts. It is for this reason that MHC extends its scope of reasons that it must be sensitive to in order to sufficiently satisfy the tracking condition, particularly in the case of autonomous systems. The smart home assistant, which we can hypothesize being connected to various other smart home devices (illumination, water/air quality monitoring, climate control, video security, cleaning, etc.) must be sensitive to both proximal reasons as well as distal ones. Satisfying solely proximal reasons (i.e., increase temperature) can sacrifice the
more general objective distal reasons (i.e., maximize efficiency, or legal energy saving laws⁴).

Adopting this systems thinking approach to conceptualizing the tracking condition in particular requires that all the elements that are part of any given system(s) must be maximally sensitive/responsive the relevant (moral) reasons of agents, being users or otherwise. This means that it is not solely the burden of agents to be maximally able to behave according to patterns of reasoning, but every point in a system’s infrastructure must be similarly sensitive. This responsiveness can be framed by designers by choosing the proper ‘level of abstraction’ (Floridi, 2017) in creating autonomous systems based on the context of use to ensure receiver-contextualized explanations and transparent purposes (Floridi, Cowls, King, & Taddeo, 2020). This means that a smart home camera system, for example, must not only be responsive to the users’ reasons, but also conform to legal and social norms such as privacy and surveillance laws. Mecacci and Santoni de Sio (2019) are explicit in that, although the tracking conditions reads that the system must be responsive to human reasons and not to other vectors in a system, they argue that social and legal norms reflect the intentions and reasons of supraindividual agents such as organization, companies and states (Mecacci & de Sio, 2019, p. 4).

The implications of their approach are not insignificant, as they appear to run contra to the intuition that greater autonomy entails less MHC. The systems that compose smart home devices, and the systems that their integrations then form require comprehensive and ubiquitous design that permits them to be maximally sensitives not only to the end users intentions and reasons for action, but also societal norms as well as legal and policy statutes. Although, as already stipulated more than once, requiring this means having a more stringent notion of what constitutes MHC, but, as a consequence, it permits increased levels of autonomy (i.e., in the case of a smart fridge, the need to physically order groceries) with increased control over the system through design decisions and regulatory infrastructures. This means that MHC can be achieved if systems are maximally responsive to the intentions of agents beyond the final users such as the designers, companies and states in general.

⁴The European Commission Harmonised standards 2012/C 394/05 require that devices with standby modes (such as smart home devices like TVs, appliances, toys, etc., to switch into a low power mode after a reasonable amount of time and that they must (since 2013) not consume more than 0.5 Watts in standby or in off mode.
Mecacci and Santoni de Sio (2019) then apply their conception of MHC to autonomous driving systems. They explore the ethics of such systems from the lens of Santoni de Sio and Van den Hoven’s (2018) conception of MHC. In doing so, they provide salient tools for operationalizing MHC in design practice, while providing an illustrative case study on ‘dual-mode’ vehicles such as the Tesla Model S. Given the salience and comprehensiveness of their construction of MHC, this paper adopts their approach by providing examples of how to designing smart home systems to ensure MHC is attained. In such a way, it builds on their work, aiming to inject it within the discourse on the ethics and design of smart home technologies. The following section explains the value sensitive design approach, which, as mentioned, aligns with the above authors’ conception of MHC.

3. Value Sensitive Design

Value sensitive design (VSD) is a principled approach to technology design that originated in the field of human–computer interaction (Friedman et al., 2013; van den Hoven & Manders-Huits, 2009). It is principled given that it springs from a core set of conceptual precepts. It begins from the premise that technology is not value neutral and that technology always implicates some value(s). Technology is thus interactive with humans, meaning that values arise or are ‘located’ in the interaction between humans and technology, in their design, uses and practices (Friedman & Hendry, 2019).

VSD as such aims to direct technologies to socially beneficial ends (i.e., mapping on to stakeholder values) by directing designers, engineers, legal and policy specialists and others working at the crossroads of technology design and use in ways that technology can be designed for human (and nonhuman animal/ecological) values (Friedman & Hendry, 2019; Umbrello, 2018a). It does so by providing various theories, methods and practices that can be adopted by engineers and that seamlessly integrate into the already existent design practices in place. Over the last two decades, work on VSD has developed and shaped the now working definition of human values as “what is important to people in their lives, with a focus on ethics and morality” (Friedman and Hendry, 2019 p.4).

VSD’s interactional stance on technology requires designers to investigate the stakeholders who may be implicated by design and deployment (both direct stakeholders such as the designers, users and indirect stakeholders such as other
Meaningful Human Control over Smart Home Systems

publics, the environment and nonhuman animals) (Friedman & Hendry, 2019). Other commitments include, but are not limited to, investigating both the explicit and implicit values held by the designers and the stated values of the project as well as how they align with those of other stakeholders (Friedman & Kahn Jr., 2002; Umbrello, 2018b). Similarly, a fundamental commitment of VSD is investigating the group as well as the cultural and societal contexts, and the dynamic values therein, in which the technologies are being developed and eventually deployed in. It investigates how those dynamic and changing values can be accounted for in design and how to guide technological development towards progress (which avoiding avoids the pitfall of aiming for perfection) (Friedman & Hendry, 2019; Umbrello, 2020b). To better sum up, Friedman and Hendry (2019) argue that VSD emphasizes the following four criteria for progressing human flourishing in design:

- **Proactive orientation toward influencing design.** Value sensitive design is oriented toward influencing the design of technology early in and through the design process.

- **Carrying critical analyses of human values into the design and engineering process.** Value sensitive design is committed to design and engineering methodologies that bring critical analysis of human values into the design process.

- **Enlarging the scope of human values.** Value sensitive design embraces a broad spectrum of human values that arise in the human context.

- **Broadening and deepening methodological approaches.** Value sensitive design’s emergent methods draw on anthropology, design, human-computer interaction, organizational studies, psychology, philosophy, sociology, software engineering, and others.

Friedman and Hendry (2019) p. 4

3.1 A Tripartite Approach

VSD approaches are typically described in the literature as consisting of three parts or “investigations”: conceptual investigations, empirical investigations, and technical investigations (Borning & Muller, 2012). These three investigations are iterative and in practice must be treated as a continual feedback cycle (Figure 1). In conceptual investigations, designers examine the
relevant literature (philosophical, sociological, etc.) to that technology with the objective of discerning possible a priori values and tensions that may be implicated in that particular project. Preliminary investigations are also made to determine who the stakeholders of such a design would be. Empirical investigations support these conceptual investigations by eliciting stakeholders and bringing them into the design process. This can be done through a variety of established sociological tools and methods including questionnaires, semi-structured interviews, and surveys (Friedman et al., 2017). The goal here is to establish the possible relationship that these stakeholders may have to the project and how accurately the conceptually identified values and definitions map on to the empirically elicited values (Friedman et al., 2017). Finally, technical investigations focus on the technology itself to ascertain what technical constraints exist and how those constraints can support or restrict the values elicited in the previous investigations. An example would be how can a smart personal assistant balance a user’s needs for *usability* (i.e., its ability to make accurate and desirable recommendations or to restock an emptying fridge) and their need for *data privacy* (collection and storage of big data to provide more accurate recommendations).

Figure 1. The recursive VSD tripartite framework.

In order to better conceptualize how we can apply MHC viz. VSD to smart home technologies, this paper will use the example that began this paper, that of Mr. Orange and the Ring doorbell camera. To reiterate what happened more generally, it was discovered that a hacker gained remote access to Mr. Orange’s Ring camera and, consequentially, was able to see what was in the camera’s field of view, in the initial case his children playing, as well as use the camera’s two-way microphone to speak with them. Similar incidents have occurred with Ring products in both Texas and Mississippi, the former in which a couple were demanded a ransom of $350,000 in bitcoin and the latter involved speaking with an eight year old girl (Paul, 2019). The lawsuit against Ring mentions, alongside the fact that they do not require users to set up an initial password, has been their failure to encourage two-factor authentication in the access of these devices. This can include using a secondary device to authenticate a login such as a text message to a phone, facial recognition login, etc. Although a Ring spokesperson told the Guardian that “We have no evidence of an unauthorized intrusion or compromise of Ring’s systems or network” (Noor, 2019), on December 18 2019, they sent an email out to their registered customers to inform them that the log-in credentials for 3,672 Ring camera owners were compromised, thus leaking information such as log-in emails, passwords, time zones, and the names people give to specific Ring cameras.

The theory of MHC that has already been discussed can give some insight into this case. To begin, MHC requires us to look at how the responsive the system was (if at all) to the relevant agent(s) reasons (tracking condition) as well as if we can trace (second condition) at least one agent in the design history or use of the system that (1) understand the systems’ capacities and thus (2) realizes their position as being morally accountable for those capacities post deployment. With regards to the tracing, it appears that the end user, in this case Mr. Orange, did not understand that the system he purchased and installed was capable of being accessed in such a way, likewise, he may not have even been aware of the existence of security protocols like two-factor identification. Likewise, he may not have been sufficiently informed by the company, Ring, in the proper functioning of the system as well as how to protect their system from unauthorized intrusion. Beyond this, as an agent, he himself may have been ignorant of his own expertise and capacities in the proper functioning of this system (or other automated systems in general). The lawsuit against Ring then
represents something beyond solely the tort offence, but rather it shows how Mr. Orange’s responsibility for the system’s recalcitrance could have been hampered by how the company communicated their product (failing to ask customers to create a password can give the impression that login security is nothing to worry about, i.e., an omission of the factor of security in their product). Tracing the design history and use of the system, then, muddies the waters in terms of full or partial responsibility (and to what degree) can be attributed to both Mr. Orange as well as Ring (leaving aside for the moment the data leak and just focusing on how the product was presented and used). How about the tracking condition? Mecacci and Santoni de Sio (2019) provide some guidance as to how the tracking condition cannot only be satisfied in cases like this, when dealing with (semi)autonomous systems, but also how to design for it.

Although the example of Ring proves to be a salient example, the purpose of this paper is to focus on these types of smart home systems more generally. For this reason, in exploring the tracking condition I will use the example of smart personal assistants more generally, to better illustrate the efficacy of this approach, as well as to provide a more comprehensive account of smart home systems in general. To achieve this however, it bears noting from the onset as to what extent, if at all, such personal assistants are responsive to the relevant (moral) reasons of the relevant agents. I use the following illustration, albeit a fictional one, to better frame the following discussion:

Mary, the homeowner of a small urban apartment, has augmented her home with the latest smart home technologies. Her front door is equipped with a smart lock that deactivates when her phone is nearby (i.e., geofencing). Her lights automatically adjust to the time of the day to reduce blue-light eye strain in the evening (e.g., nightshift). The air conditioner and heating unit monitor air quality, noise level, humidity and temperature and automatically adjust them to Mary’s pre-set comfort levels. The refrigerator catalogues all items stored, determines what is most used and send notifications to her mobile devices to let her know when the produce is running low and placing online orders for regularly purchased necessities automatically. Similarly, the water for her tea is automatically put to boil by her smart kettle every morning, and her autonomous vacuum robot cleans her floor when she is not home so as to not ever be disturbed.
Visualizing this environment more clearly, we can now imagine Mary in the following scenario:

Mary decided to buy all of these new devices so as to abdicate much of the mundane routines that she would have to do manually. Upon purchasing each of the items, she realizes that she still needs to download the app of each device, make accounts and profiles, and program each app individually. It didn’t appear much more convenient than if she had to do those things manually. She ultimately decided to buy a digital personal assistant that connects all of these devices and manages them. The device learns over time, collecting data, understanding Mary’s habits and preferences, and makes suggestions to optimize her activities and routines. She can control all of them with a simple voice command, or, if she has the patience, use the app on her smartphone for more detailed settings.

By that definition, the systems do not have full autonomy, but are semi-autonomous to the extent to which they do require input from Mary in order to ensure accurate mapping of her habits and preferences to the consequent systems’ actions. Likewise, Mary does still need to make occasional, albeit perhaps less frequent, interactions with the system, inserting and updating things like credit card information for regular, automated purchases as well as modify thermostat preferences based on changing temperature and comfort levels of herself and guests. If we stay at the level of the user, many, if not all of the systems actions (even potentially recalcitrant ones) can be framed in terms of either distal or proximal reason concurrently. For example, shocked by her credit card bill, Mary realizes that her assistant purchased three times as much broccoli as she would normally purchase at any one time. This can be explained by Mary’s more frequent purchase of broccoli and, given its lifespan, the system saved her money by purchasing enough for optimal shelf life-consumption time. It can also be explained by her sudden change in diet that she logged on her fitness app on her phone, registering to a certain diet plan. The rationalizations for why this behavior has happened can be said to be responsive to a variety of both Mary’s proximal reasons (i.e., more frequent broccoli consumption) as well as distal ones (i.e., a long-term healthy diet). Either way, we seem to be able to satisfy the tracking condition of the system’s behavior via either scale of reasons, particularly the distal one above, as it would satisfy her more general plan of increasing efficiency. If her intention is to consume more broccoli, either from normalized habit, or because of a more general plan to include more in her diet, the system seems to be sufficiently responsive in mapping her intentions on its
consequent actions. If we go back to the 2019 debacle with the Ring device, if Mr. Orange did employ two-factor identification, then the system would have responded in kind, and one could say that the device would have not become recalcitrant in the way that it did. In this, albeit constrained sense, these systems can be interpreted as being under MHC.

However, this type of control does not appear to be meaningful in the sense described above, nor in the sense that I intend here. The system’s responsiveness to the reasons of the user, in this case Mary, are not the only agents in whose actions should the system be responsive too. Although important and relevant, they are not exclusively nor exhaustively so. Regarding the Ring camera again, it was apparently not designed in such a way as to ensure users protect the technical security of their devices. Because of this, the system was designed in a manner that brings into conflict two of Mr. Orange’s distal reasons: his intention to automate and increase his household security—further evidenced by his trust in not seeking out two-factor identification—as well as the safety of his family (that which a security system is meant to increase also). The Ring was obviously not designed with the latter reason in a sufficiently explicit way, most obviously because it did not inform the user that two-factor identification was an option, and that, if enabled, would increase the systems resilience to intrusion. This would also signal to the user that the systems is itself prone to such types of intrusions, given its ability to be remotely accessed via a network. Secondly, the system was not designed in such a way as to be capable of determining the user’s level of technical knowledge of how the system is accessed, and was not able to provide the user with any notification of unusual activity.

Consequentially, a (moral) accountability gap is left open on account of the design decision made by the manufacturer (Mecacci & de Sio, 2019). This comes as a product of the asymmetric allocation of the relevant knowledge regarding the capabilities of the system as well as how the relevant agents in the tracking chain discern both their own and each others’ responsibility in its use (Mecacci & de Sio, 2019; F Santoni de Sio, 2016; Filippo Santoni de Sio & van den Hoven, 2018). Likewise, a disparity and inconsistency in the level and quality of control of the system is created given that such a system was designed to have Mr. Orange abdicate some of his direct control in order to fulfill his intentions (i.e., home security) while concurrently not being able to (consistently) act on these reasons autonomously (as the lawsuit against Ring aptly argues).
Of course, and as Mecacci and Santoni de Sio admit, distal reasons are arguably more difficult to design for in systems given their more general complexity. This does not entail that such devices are not responsive to these types of reasons, in fact, the Ring camera performs quite well at monitoring the home 24/7, notifying the user of movement and letting them know when the battery is low so that it can continue its surveillance. One should resist the conclusion then that such systems, in order to have more meaningful control, require greater direct operational control. In fact, as mentioned previously, systems such as the Ring should be designed in such a way as to be as maximally responsive, as technically feasible, to the proximal reasons of agents via greater degrees of automation. Augmenting automation through responsible design can actually aid in satisfying the tracking condition if such automation permits greater responsiveness to those stakeholders’ reasons.

Turning to the design question then, we shift design motivation away from consequential behaviors and towards agential reasons. A less rigorous behavioral requirement of direct operational control is superseded by a more rigorous, and thus more meaningful notion of control that permits clear lines of responsibility to be drawn despite potentially greater levels of systemic autonomy. One way designers can then begin to look at the design of smart home devices, including assistants, is to design them in such a way as their integration and cooperation amongst one another, is maximally sensitive to the widest range of agent(s) relevant reasons as possible. This does not only mean the proximal and distal reasons of the end user (that was shown above), but also to supraindividual agents such as companies (i.e., manufactures), cities or states in general. This can be done by engineers by looking at different gradations of proximity in terms of reasons and the corresponding systems behavior over time. Similarly, augmenting autonomy in systems such as the personal assistant, as long as such is sensitive to the relevant distant and proximal reasons of the relevant agent(s), can similarly increase MHC over the network as a whole. Helping to make better decisions that would otherwise not be made by human users is arguably more meaningful. In the case of Mary’s smart home assistant, one way to do this is to make sure that systems like that are tempered by receiver-contextualized reasons and transparent processes that continually satisfy the tracking and tracing conditions (Boscoe, 2019; Floridi et al., 2020). The fact

---

5 This would mean that depending on the agents (i.e., user, designer) the explicability of the system and the reason that it gives for any given action will be contextualized differently given the different levels of knowledge of the system’s capacities.
that the system purchased more broccoli than usual can satisfy, as has been described above, both her proximal or distal reasons; but this does not entail that it was under MHC. Of course, MHC could still be described by tracking the relevant actors throughout the design-use network (designers, manufacturer, state norms/laws, etc.). However, in this case, the system’s behavior could be brought under more meaningful control if it provided such types of explanations, even if it is only to more semantically trained specialists like designers and programmers (which can then relay any queries to users). Figure 2. is an example of how to illustrate VSD’s translation of values into technical design requirements and vice versa.

Figure 2. illustration of a bi-directional hierarchy of values, in this case how efficiency can be translated into norms and into technical design requirements. The process can also begin from design requirements and move towards affirming values.
Figure 2 is a simple illustration of the varied approach that can be taken by designers for visualizing how to translate abstract values into more tangible technical design requirements. Van de Poel (2013) proposed that the most salient way for this translation to take (and it can be taken in either direction) is to translate values or design requirements through norms. Norms can be understood as normative imperatives such as ‘maximize security’ or ‘minimize harm’. Norms can also take the form of legal statues or policy guidelines, providing a structural flow that can support or constrain the open avenues for determining the most salient design requirements, or vice versa. If we take Mary’s smart home assistant as the paradigm again, looking closer at Figure 2, we can see that the hierarchy begins with the value efficiency, which, in the case was her motivating factor in buying her assistant. Norms are the vector of dynamic change, differing based on sociocultural, economic, demographic and geographic contexts (Friedman & Hendry, 2019). ‘maximize autonomy’ could just as easily be substituted with a legal norm that mandates non-proprietary API’s to permit integration between systems and reducing causal complexity in the event of system recalcitrance. Naturally, depending on the norms that designers find themselves with, the technical design requirements that present themselves consequentially vary, supporting certain norms better than others.

A similar route could have been taken, and perhaps should be considered, for technologies like Mr. Orange’s Ring system. His desire for security could have been translated through various legal norms in his state and at a federal level also to ensure that other values such as data privacy are protected (which was shown to ultimately have been compromised). More tangible design requirements such as the mentioned two-factor identification could have not only been implemented, but encouraged by the system such as prompts during the install on the device as well as brief, but clear explanations of its purpose and that entailed risks associated with using standard login protocols. The form of such requirements are not exclusive to this, and it may be the case that there are more practical ways of achieving the same ends. There is a subtle point here however that is not unimportant, and is, in fact, critical to understanding the proximity of reasons and their importance in attaining MHC, that being, although some reasons (proximal/distal) may not be reflected by the systems

---

6 This is aligned with an underlying precept of VSD, the approach eschews perfection as its aim, but rather progress. Hence, the reader should resist the notion that the aim here is to find a perfect way or comprehensive way of account for all reasons and values. A thorough discussion of the ‘progress, not perfection’ paradigm can be found in Friedman & Hendry (2019).
behavior at any given moment, they may conflict with higher level reasons of other agents, such as the state in the form of laws. It is for this reason that MHC need not be located always in the end user, and it rarely is exclusively, given the distribution of the knowledge given the systems capabilities and limits. Mary, for example, may use a voice command to her assistant telling it to raise the music to maximum, but the system may only achieve 40% of its possible master volume given certain legal constraints such as not having music of a certain decibel level past 11pm in her geopolitical area. Her system was responsive to her proximal reason to only a certain extent, but it was so because it was maximally responsive to the distal reasons of other actors such as the state, and the designers who programed it according to such norms. The tracking condition in such a case can be satisfied nonetheless.

To briefly sum up then, MHC can only be attained through the satisfaction of the tracking and tracing conditions discussed above. The tracing condition is that which is easier to determine and literally trace the design histories and agent(s) in any given technological artefact. The tracking condition however is more nuanced given that it follows the agent(s) reasons governing the design, deployment and use of the technology. VSD is proposed as a means by which the more nuanced tracking condition can be conceived of in the design process, using the example of smart home assistants. Designers can then begin to think of designing technologies for meaningful human control, rather than retaining the concept at an abstract level, or more burdensome, how to modify any deployed technology to be under MHC.

5. Limitations and Suggestion for Further Research

This paper has thus far explored the concept of meaningful human control and provided a conceptual case of how to illustrate its meaning within the context of the evermore widespread deployment and use of smart home devices, more specifically, centralized smart home personal assistants. In doing so, we discussed how MHC can be attained conceptually in these systems and some of the values that are necessary if such systems are to be capable of MHC.

There are some conceptual limitations of this approach however. What is needed is real case studies of how VSD can be used to attain MHC in general,

---

7 An example of this would be the Administrative Regulation (Legge Quatro 447 of October 26 1995) in Italy that stipulates silence in urban areas after 11pm and before 7am.
and towards smart home devices more specifically. Ideally, how VSD can be used with the explicit goal of attaining the values central to attaining MHC in a specific smart home assistant would be the most salient example of the feasibility of this approach. Here the case has been made solely for the conceptual feasibility of the approach of using VSD to attaining MHC.

Likewise, there are points in Santoni de Sio et al.’s conception of MHC that could be taken up by scholars that may prove to be contentious, more specifically, their claim that that their [Mecacci and Santoni de Sio, 2019, p.5] framework provides an objective framework from which engineers can use to identify and rank-order different reasons and agents. Their arguments of such an approach rests on many existing theories of human control, intentions and practical reasoning as well on their relationships in technological innovation and use. Whether undermining or showing why any given one of those theories proves to be incompatible or simply outdated is yet to be seen, as well as to whether doing so undermines their conception of MHC. All in all, whether or not such an approach pans out in engineering practice is tentative, albeit conceptually convincing.

Likewise, an underlying notion assumed in the distribution of reasons, particularly distal reasons, is how to rank-order a systems responsiveness to different categories of reasons (that of users, designers, state laws, etc.). Although Santoni di Sio and Mecacci (2019) are explicit in that they do not endorse one rank-ordering over another, there is nonetheless the intuition that state legislated distal reasons can, and probably will, supersede the proximal, and perhaps subsequently, the distal reasons of the end user. Although nonetheless remaining under MHC in this paradigm, the preference of collective distal reasons seems to take precedence over that of the individual. This may be justified in many cases, and can be argued to be the always already paradigm for responsible technology development, however scholars should consider this underlying tension as a point of interest in further dissemination of MHC.

Similarly a fruitful area of future research is determining the needs for policy measures that govern the design and development of smart home devices given their implications of fundamental human values. How this can be done, if at all, through the implementation of VSD may be a prudent starting point given that it affirms the central values that are fundamental to democratic policy (i.e., justice, equality, autonomy, freedom, etc.) (Mintrom & Luetjens, 2017; Umbrello, 2020a).
6. Conclusion

The ubiquitous interconnectedness of smart home systems controlled by AI empowered smart assistants offers promises of convenience and efficiency. However, many of these systems function in ways that are not clear to their users, gathering vast quantities of data spread across networks, infrastructures and agents. These implicate many critical human values such as privacy, security, transparency and accessibility. Giving up some of the direct control of systems to automated devices implicates a host of ethical issues that are not easily resolved, however, where lies the problem might also lie the solution. This paper has discussed what it means to have meaningful human control over these household systems. Illustrations of smart home assistants is given in order to better understand the demands of attaining sufficient control that aligns with important human values. What is shown is that the various conditions that are necessary to attain MHC can be satisfied via value-aligned design requirements by adopting a value sensitive design approach, in doing so, greater levels of autonomy that are more responsive to the relevant reasons of the relevant agents can augment MHC rather than lessen it. Examples are given for how these design requirements can be conceptualized in practice by engineers to better align smart home assistants with human values and intentions/reasons and thus conceptually attain sufficient MHC.

REFERENCES


