Responsibilities in a Datafied Health Environment

Responsibilities in a Datafied Health Environment

Chirag Arora

Responsibilities in a Datafied Health Environment

THESIS

ter verkrijging van de graad van doctor aan de Technische Universiteit Eindhoven,
op gezag van de rector magnificus prof.dr.ir. F.P.T. Baaijens,
voor een commissie aangewezen door het College voor Promoties, in het openbaar
te verdedigen op dinsdag 6 december 2022 om 13:30 uur

door

Chirag Arora

geboren te Delhi, India

Dit proefschrift is goedgekeurd door de promotoren en de samenstelling van de promotiecommissie is als volgt:

Voorzitter:	prof.dr. I.E.J. Heynderickx
Promotor:	prof.dr.ir. A.W.M. Meijers
Copromoteren:	dr. E.R.H. O'Neill
	dr. M. Razavian
Leden:	prof.dr. C.C.P. Snijders
	prof.dr. T. Sharon (Radboud Universiteit)
	prof.dr. L.E.M. Taylor (Tilburg University)

Het onderzoek of ontwerp dat in dit thesis wordt beschreven is uitgevoerd in overeenstemming met de TU/e Gedragscode Wetenschapsbeoefening.

The PhD research is executed at Eindhoven University of Technology, Department of Industrial Engineering & Innovation Sciences.

Editors: Philip Brey, Anthonie Meijers, Sabine Roeser and Marcel Verweij

Cover picture by Mingwei Lim (photo from unsplash).

ISBN: 978-90-386-5615-1
ISSN: 1574-941X

Copies of this publication may be ordered from the 4TU.Centre for Ethics and Technology, info@ethicsandtechnology.eu
For more information, see http://www.ethicsandtechnology.eu

Contents

Preface

This doctoral thesis consists of an introduction, a conclusion, and the following chapters written for publication in peer-reviewed journals independently:

Chapter 1: "Digital health fiduciaries: protecting user privacy when sharing health data" published in *Ethics and Information Technology* as Arora, C. (2019). Digital health fiduciaries: protecting user privacy when sharing health data. *Ethics and Information Technology*, *21*(3), 181-196.

Chapter 2: Arora, C. Googlization of Health Research and Epistemic Trust.

Chapter 3: Arora, C. Proxy Assertions and Agency: The case of machine assertions.

Chapter 4: "Ethics of Gamification in Health and Fitness-Tracking" published in *International Journal of Environmental Research and Public Health* as Arora, C., & Razavian, M. (2021). Ethics of Gamification in Health and Fitness-Tracking. *International Journal of Environmental Research and Public Health*, *18*(21), 11052.

Acknowledgements

I want to begin by expressing my deep gratitude for the immense support and mentorship provided by my supervisor, Elizabeth O' Neill. I am also indebted to Anthonie Meijers for his guidance and valuable feedback on my work. I am grateful to Maryam Razavian for her mentorship as well as for being an extremely supportive collaborator on a part of this research project.

I want to express my sincere gratitude to the independent members of my doctoral committee, Tamar Sharon, Linnet Taylor, and Chris Snijders for their willingness to engage with this dissertation as well as for their feedback.

I would like to extend my gratitude to my colleagues in the Philosophy and Ethics section at Eindhoven, especially for their generous feedback on earlier drafts of portions of this dissertation. I am grateful to Wybo Houkes, Rianne Schaaf, Kate Raaijmakers, Philip Nickel, Lily Frank, Sven Nyholm, Andreas Spahn, Marjolein Lanzing, Dunja Šešelja, Mandi Astola, Naomi Jacobs, Krist Vaesen, and Iris Loosman for their support during the duration of this research. I also want to thank members of the Information Systems group at Eindhoven for their support in this research. I am grateful to Claudia-Melania Chituc for feedback and helpful conversations on the topics of this dissertation.

I am also thankful to Helen Nissenbaum and members of the Digital Life Initiative at Cornell Tech for supporting my research during my time there as a visiting researcher.

Finally, I would like to thank my parents, Asha and Mohinder Arora, for their unwavering trust and support.

Introduction

Over the past decade or so, advances in digital technology have increased the capacity to quantify aspects of the world that had not been quantified before (Mayer-Schönberger & Cukier, 2013). The phenomenon of "datafication", characterized by an accelerating rise in the collection and analysis of quantified data, has been particularly impactful in the healthcare sector (Ruckenstein & Schüll, 2017). Driven by consumer-oriented devices such as wearables, which can collect multidimensional health data, as well as technical advances in processing and analyzing big data, datafication has led to an expansion of the scope of healthcare by enabling monitoring of health-related outcomes and behaviours outside the traditional health institutional and clinical settings. Proponents of this health datafication phenomenon point out how such "institutional recasting" (Swan, 2012) can empower users of digital health technologies such as sensor-equipped wearable devices, often tethered with their smartphones – allowing them to take control of their health, generate their own medical data, and track behaviours that would be difficult to track unaided (Davies, 2021; Topol, 2015). Writing about the empowering potential and participatory nature of health datafication, Eric Topol states: "Just as the printing press democratized information, the medicalized smartphone will democratize health care" (Topol, 2015).

Despite potential benefits, health datafication does not come without ethical challenges. Critics of the health datafication phenomena point out, for example, the reductionist tendencies associated with self-tracking – rich categories of "health", "good sleep", "mental well-being" are replaced by their narrowly-construed quantifiable proxies such as calories consumed, motion detected by a wrist-watch sensor, and a mood score on a smart-phone app (Sharon, 2018). This reductionism carries the risk that such proxies will come to be perceived as the definite truth or reliable knowledge by the users, as well as privilege such quantification over other ways of knowing (Sharon, 2018).

Another line of criticism of the health datafication paradigm targets the "empowerment" thesis outlined by the proponents. Critics state the worry that rather than empowering individuals, self-tracking and monitoring of health shifts the burden of responsibility away from medical professionals and health policy-makers to

individuals (Davies, 2021). This focus on individual responsibility may also lead to a future where individuals are under constant surveillance, held unfairly accountable for their use of public health resources, and penalized for health outcomes they may not be actually responsible for (Davies, 2021).

The emphasis on individual responsibility also ignores the labour of a multitude of actors required to enable individual users to reap the benefits of health datafication. Consider, for example, a sleep-tracking device that helps inform a user about the qualitative and quantitative aspects of their sleep. The design of such a device in itself may involve multiple actors, such as those designing the hardware and those designing the software. These actors may themselves have to rely on scientific research output or technological innovation of others in order to design the device such that it is able to accurately capture the required data and produce a desirable output. Further, as Crawford et al., (2015) point out, interpreting the significance of the data captured by such devices, and consequently offering desirable qualitative insights to the user, is based on a statistical comparison of a set of data points and therefore, *requires* participation by a large number of users.

The significance of the roles and responsibilities of such actors becomes even more apparent when one considers the multilevel nature of inquiry within the health datafication paradigm. The insights and inferences that can be drawn within the health datafication paradigm are not limited to the health status of an individual. By giving researchers access to new forms and large quantities of quantified health data from a vast number of users, health datafication also enables health inquiry at a collective or a societal level – for example, by providing an opportunity to redefine healthy behaviour as well as to classify and diagnose diseases in novel ways (Ada Lovelace Institute, 2020). In one study, for example, passive smartphone data, such as GPS data as a proxy for location and socialization, and accelerometer data about physical activity, enabled researchers to predict depression with roughly 60% accuracy (Wahle et al., 2016). Similarly, mobility data from smartphone apps has been used to draw population-level insights regarding the spread of the Covid-19 pandemic as well as to inform policy interventions to curb the spread of the disease (Sheng et al., 2022; van der Drift et al., 2022). The success of such inquiry, again, depends on the work done by designers of such apps, researchers analysing the data, and other actors who may help to integrate such apps with the local health system (Colizza et al., 2021).

My interest in this thesis is to explore the role these various sets of actors play in conjunction with each other, and in particular the responsibilities such actors have in facilitating successful health inquiry within the datafied health paradigm. Each chapter of the thesis focuses on the responsibility of a different (set of) actor(s) who contribute(s) to one of the phases along what can be called as the health "data value chain". The data value chain here can be understood as a series of phases involved in the creation of valuable insights from data (Curry, 2016). These phases range from the creation and collection of data to the usage of insights produced by the analysis of data. While different authors characterize the phases along the data value chain slightly differently, for the purposes of this introduction one can broadly characterize them into at least three distinct phases in the context of health datafication:

1. Data Collection and Storage – This phase involves, for example, the collection of data through consumer health devices such as a Fitbit or an Apple watch and subsequent storage of this data, for example on data servers. Big tech corporations who provide such consumer health devices then are one of the primary actors involved in this phase within the health datafication paradigm. Policymakers and regulatory bodies are another set of actors who have a significant influence in shaping the activities within this phase, for example, by enacting regulations regarding the kinds of data that can be collected, as well as conditions under which such data may be collected and stored. In Europe, for example, the General Data Protection Regulation (GDPR) encodes principles such as data minimisation. Such a principle dictates that only data that is "relevant and necessary to accomplish a specified purpose" may be collected (*D | European Data Protection Supervisor*, n.d.).

2. Data Processing and Analysis – This phase involves eliciting actionable insights from the collected data. It may involve, for example, the use of big data algorithms and techniques such as machine learning and/or neural networks. As already discussed, inquiry within the health datafication paradigm can take place at multiple levels and the insights produced within this phase may relate to an individual and/or a collective or groups. As collectors of a large amount of data through consumer health devices, big tech corporations are, again, a crucial actor involved in this phase. Other than providing apps that can provide actionable insights to individual users, the financial, technical, and human resource capabilities of such corporations also enables them to play a

leading role in data-driven healthcare research. An example of such a research project includes Verily's (a life-science research organization that is part of the Alphabet group) project Baseline, which aims to "map human health" by analyzing large amounts of phenotypic, genetic, and lifestyle data collected from 10,000 volunteers (Arges et al., 2020).

Besides tech corporations, regulators can also play a key role in shaping the activities within this phase, for example, by defining legal conditions under which some specific type of data shall or shall not be processed. Such regulations can play an important role in protecting the rights of the users, such as rights relating to their privacy. Going back to the example of the GDPR, the regulation defines conditions that limit the processing of biometric and genetic data, for example ("Art. 9 GDPR – Processing of Special Categories of Personal Data," n.d.). GDPR also defines restrictions on processing data that may reveal personal characteristics of individuals, such as their race or ethnicity.

3. Data Usage – This may involve, for example, the display of actionable insights, gained as a result of the processing done in phase 2, to the user of a consumer digital health device. The user may get access to such insights via a smartphone app, for example. It may also involve advanced forms of usage such as in the form of "gamified" apps which use game-like elements to motivate users to engage in activities that may improve their health (or improve their "score" which is a quantified proxy for their health) (Whitson, 2013). Users of such apps depend on the designers of the apps to successfully acquire the desired information from the use of such apps. Such apps, particularly the gamified ones, can also have significant affective consequences for the users, such as causing stress or anxiety, which makes the role and responsibilities of designers even more significant (Barratt, 2017; Lupton & Thomas, 2015).

It is worth noting here that these "phases" can be dynamic, iterative, and sometimes overlapping. For example, the processing and analysis of data may help inform changes that may be needed in the data collection phase. Similarly, among other things, user behaviour, as well as user expectations, may also inform what data needs to be collected and what kind of insights may be desirable for the user.

The nature of responsibility in a datafied health environment: epistemic and non-epistemic dimensions

Successful inquiry within the health datafication paradigm, through data-driven technologies, is a result of collective action, involving multiple actors, along the different phases in the data value chain, each of whom might share some responsibility for the success of the inquiry. Before I delve deeper into the content of the individual chapters of this thesis, and how they address the questions pertaining to the responsibilities of various actors within the datafied health paradigm, it is perhaps important to first answer the question - what is the nature of such responsibility? One way to characterize such responsibility would be to call it an "epistemic" responsibility – in the sense that the aim of such responsibility is to produce "epistemic goods" such as knowledge or true beliefs (Fleisher & Šešelja, 2021).

For example, we may expect that the designer of a fitness-tracking app has an epistemic responsibility to ensure that the user is getting true information about their physical activity and/or is getting the data to make reliable inferences about their activity. Consider, for example, a sleep-tracking app. In a study about such apps, researchers concluded that some users of sleep-tracking apps may suffer from a condition researchers term "orthosmonia" (Baron et al., 2017). Such users, according to the researchers, rely too heavily on the data displayed on their sleep-tracking devices to self-diagnose themselves with a sleeping disorder. Further, this belief can be problematically rigid, such that they may continue to believe that they have a sleeping disorder even when polysomnography, considered the gold standard in diagnosing sleep quality, may inform them that they do not have such a sleeping disorder. The designers would then be expected to have at least a prima facie epistemic responsibility to help users of the sleep tracking app to avoid such a situation. Similarly, we may expect that the researchers within the big tech corporations have an epistemic responsibility to ensure that their research produces reliable actionable insights on a collective or group level. One example of where such responsibility may be pertinent is in the form of addressing the problem of contextual bias that data-driven medical research may be prone to. Ii & Nicholson (2019), for example, argue that data-driven medicine's contextual bias problem is a result of the fact that a lot of data on which algorithms used for diagnosing diseases are trained may be collected in high-resource environments, and thus, their accuracy may be low when applied in low-resource environments. A real world example of such a problem was seen in an AI solution developed by Google to diagnose diabetic retinopathy (MIT Technology Review,

2020). While the AI was fairly accurate in the lab settings where it was developed, its accuracy reportedly dropped significantly when deployed in real-world settings to diagnose patients in Thailand. One way in which is such responsibility may manifest itself is in terms of researchers ensuring the collection of data is conducive to the purposes it will be analyzed and processed for.

These examples show the significance of epistemic responsibility of the actors involved in the different phases of the data value chain, such that they do their part in ensuring that inquiry within the health datafication paradigm produces true or reliable beliefs for the participants. However, on closer inspection, it becomes more apparent that such responsibility is not purely "epistemic" for at least two reasons.

First, in health inquiry, while one is looking to acquire true beliefs, the ultimate aim of such inquiry may be more practical – such as acquiring beliefs that may help one make the right health choices. The practical aim of such inquiry is particularly relevant for defining the responsibility of those involved in the inquiry in cases where one encounters, for example, what is known as "an inductive risk problem" (Douglas, 2000). In the chapter titled "Googlization of Health Research and Epistemic Trust", I discuss, for example, the case of contact tracing apps, such as ones in use during the COVID-19 pandemic, where designers of such apps encounter the inductive risk problem. Briefly stated, the idea is that in designing a contact tracing app, and deciding what the app counts as a "contact", one has to balance the risk of false negatives (i.e. significant contacts not registered as such) with the risk of false positives (non-significant contacts registered as significant contacts). This balancing act, however, is not a purely epistemological exercise and needs to consider the social and/or ethical consequences of how such an app may actually affect individuals, particularly if the results of the app are used in a manner where they put significant restrictions on those registered as a contact. As a consequence, the designer of such an app has duties pertaining to analyzing and evaluating the social and ethical consequences of the epistemic content the app would eventually provide. Similarly, in the chapter on "Ethics of gamification in health and fitness app", Maryam Razavian and I discuss how the designers of such apps have responsibilities related to how the content of the app (psychologically, for example) affects the users, even if such an app is displaying "true" information.

Second, the assessment and evaluation of the inquiry, and responsible action within such inquiry, from an ethical perspective, for example, involved in producing the relevant epistemic goods depends not just on the result of the inquiry but also on the processes that are part of the inquiry. For example, while users of self-tracking apps and devices are interested in knowing actionable insights to improve their health, they do not necessarily want that at the expense of harm such as loss of privacy. In the first chapter of this thesis, I argue that the actors involved in health data collection and processing should be legally responsible for not just producing relevant insights for the users but also for taking measures that protect the privacy of the users (Arora, 2019). Defining the responsibility of those collecting and processing big data on health in terms of processes to be undertaken as part of the inquiry is particularly important given the power asymmetries between users of the datafied health ecosystem and actors who collect and process the information of such users. The field of critical data studies has particularly drawn attention to such power asymmetries within the health datafication paradigm (Andrejevic, 2014; Ruckenstein & Schüll, 2017). Such power asymmetries are introduced not just by virtue of the vast amount of data corporations are able to collect but also in light of the sensitive nature of health information recognized by existing regulatory frameworks as well (Arora, 2019).

This point about power asymmetries, within the datafied health paradigm, between users and big technology corporations such as Alphabet (Google) and Apple brings me to a final point about the non-epistemic aspects of the nature of responsibility such corporations should have as facilitators of health inquiry. Responsible behaviour from a facilitator of an inquiry, such as a scientific organization, is often the source of trustworthiness. Subsequently, such trustworthiness may be necessary to reap the benefits of the epistemic goods produced as part of the inquiry. In the chapter on "Googlization of Health Research and Epistemic Trust" I make such an argument in the context of datafied health ecosystem. Tamar Sharon (Sharon, 2016), in her work, has highlighted the ethical significance of the phenomenon termed by her as the "Googlization of health research" (GHR), characterized by the vast amount of health data collected by corporations such as Google and Apple coupled with their significant, and increasing, an advantage in terms of technical, financial and human resources to analyze this data compared to traditional health research institutions. In the second chapter in this thesis, I discuss how the phenomenon of GHR demands attention to broader moral responsibilities of such corporations given the potential of

moral discrepancies by such actors to significantly affect public epistemic trust in scientific research produced within the paradigm of GHR.

Technological focus of the thesis – Consumer-oriented devices and health apps

Although many technologies enable and are associated with health datafication, and this thesis discusses some of them, the focus of this thesis is on consumer-oriented health devices, such as smartwatches or other sensor-equipped wearable devices, as well as apps driven by smartphones or other smart appliances through which people interact for health information. This choice of focus on such technologies is partly driven by their transformational potential for the healthcare system. As indicated in the discussion so far, by enabling health data collection in large volumes, from a large number of people, in settings that exist outside the traditional boundaries of the healthcare system, such technologies not only drive the process of datafication but also come with potentially disruptive effects on the conceptualization of responsibility and on the nature of relationships within the healthcare system.

For example, with respect to the notion of responsibility, the discussion in this thesis pushes back against the "empowerment" thesis and the focus by proponents of health datafication on individual responsibility for their health. Although consumer-oriented health devices come with great potential benefits for users to track, monitor, and be informed about their health, such benefits can be realized only with the responsible actions of other actors, particularly technologists such as designers of wearable devices and developers of health tracking apps, who make critical decisions about what data is collected through such apps, the purposes for which such data is collected and processed, and in what form is this processed information conveyed to the user. Such actors are involved across the data value chain and include - designers and developers of such consumer-oriented devices (phases 1 and 3 of the data value chain), designers of health-related software (apps) that users interact with on these devices (phases 1 and 3), policy makers or regulators who may set rules on what data can be collected through such devices and the purposes for which they may be processed (phases 1 and 2), and big tech corporations who may be involved in designing of both hardware and software for such devices as well as processing the data collected through such devices for commercial and research purposes (all three phases).

The choice of focus on consumer-oriented technologies, and the technologists who take critical decisions regarding the design and use of such technologies, is also driven by the potentially disruptive effects on the nature of relationships in the healthcare system and on norms that may apply to interactions within such relationships. For example, by giving access to health data to actors outside the boundaries of traditional healthcare, and in some cases without social and legal protections, limiting the use of such health data, that apply to the traditional actors such as doctors, consumer-oriented devices threaten to expose users to new vulnerabilities vis-à-vis actors who control the data collected through such devices (Ada Lovelace Institute, 2020). Similarly, by providing an interactive portal with the healthcare system outside the traditional boundaries of the healthcare system, consumer-oriented health devices operate in a novel, and somewhat ambivalent (Lupton, 2017; Ruckenstein & Schüll, 2017), normative environment which can have potentially disruptive effects on affective aspects of healthcare for the users of such devices (Lupton, 2017). Many scholars have, for example, raised concerns over the accuracy of information received through such devices, use of manipulative or psychologically coercive elements on such devices to get users to interact more, and reductionist effects of such devices on the cognitive understanding of the users of such devices about their health (Ada Lovelace Institute, 2020; Lanzing, 2019).

Overall, the increasing popularity of such consumer-oriented devices and their potential to disrupt social, legal, and moral norms associated with interactions within the healthcare system calls attention for a philosophical inquiry into how such norms may be thought anew as well as the role various actors have in upholding such norms, and it is this inquiry that this thesis aims to partake in. The overarching aim in this thesis is to shed light on the duties and responsibilities of various actors, particularly technologists who, for example, design consumer-oriented health devices and apps. Such responsibilities and duties are characterized in relation to, and with close attention on, the actions, power, and influence of such technologists over decisions regarding data collection, data processing, and purposes for which such data is processed. While some attention has been paid to the role played by such technologists, there are many aspects of their responsibility that are still under-researched. The research question that drives the inquiry and analysis in this thesis, and the chapters within it, is then this: *What should be the epistemic, moral, legal, and/or social responsibilities of various actors, particularly technologists who design consumer-oriented health*

tracking apps and devices, towards the users of such technologies who depend on such actors for reaping the benefits, and avoiding potential pitfalls, associated with health datafication?

With this context in place, I can now directly familiarize the reader with the themes discussed in each of the four chapters that comprise this thesis, including how each chapter focuses on the responsibility, which may have both epistemic and non-epistemic characteristics, of the actors involved in the different phases of the data value chain. Through these chapter summaries, I also highlight the interdisciplinary nature of the investigation in this thesis, which draws from critique and analyses of health datafication from various philoshophical subdisciplines such as ethics and epistemology, as well as other disciplines such as law, science and technology studies, communication studies, human-computer interaction (HCI), anthropology, and sociology. The latter disciplines mentioned here have also informed the empirical evidence on various aspects of health datafication discussed in this thesis. Such empirical evidence has played a crucial role in providing the motivation for the themes and issues explored in this thesis, formulating the overall research question as well as specific research questions explored in the individual chapters, and in supporting as well as formulating the premises and arguments discussed therein.

Chapters

Chapter one of the thesis, titled "Digital health fiduciaries: protecting user privacy when sharing health data" focuses mostly on phases 1 and 2 of the health data value chain, i.e. data collection and processing. As already highlighted, the epistemic inquiry associated with the health datafication paradigm comes with some ethical risks as part of the process of the inquiry. The ethical risks related to (loss of) privacy are one such example. Privacy risks are central to debates around health datafication as datafication seems predicated on the logic that the more the data, the better the insights, or in general epistemic goods, one can gain from it. While users of digital health technologies, such as wearable devices, are interested in gaining valuable insights about their health, they also have legitimate expectations for the protection of their privacy, or at least, to keep the loss of privacy and corresponding harms to a minimum. This is the argument I follow in this chapter. The chapter also builds on, and extends, the concerns raised in scholarly accounts that follow a philosophical-legal perspective (such as those by Nissenbaum & Patterson, (2016)) regarding potentially disruptive nature of self-tracking devices. This chapter explores the

responsibilities of digital health data controllers (those who collect and process health data, such as through self-tracking devices), and argues for "fiduciary relationships" between data health controllers and the users. A "fiduciary relationship" is a legal concept, defining the relationship between two parties, a fiduciary and a beneficiary, such that the fiduciary has to keep the interests of the beneficiary at the forefront. As in the context of health datafication paradigm, fiduciary relationships exist in contexts where there are power asymmetries, and seek to protect the vulnerable party (in this case the users of digital health devices whose data is being collected) from the negative effects of such asymmetries. I argue that such fiduciary relationships be defined in the case of digital health, such that there are deliberative demands on digital health data controllers to keep the interests of their data subjects at the forefront as well as cater to the contextual nature of privacy when making decisions about the use of health data. In particular, these deliberative demands put constraints on the kind of epistemic goods data controllers can gain from personal health data as well as the kind of epistemic goods they can facilitate by sharing this data with third parties. These deliberative requirements ensure that users can engage in collective participation and share their health data at a lower risk of privacy harm.

In the second chapter of the thesis, I explore the effects of "Googlization" of health research (GHR) on warranted epistemic public trust (or trustworthiness) in epistemic goods produced by such research. As mentioned earlier, GHR is a term coined by Sharon (2016) to refer to the phenomena of large tech companies such as Alphabet (formerly Google), Amazon, Apple, etc. moving up as dominant, and perhaps indispensable, forces in health research. The question of warranted epistemic public trust in scientific output produced through GHR is important for at least two reasons: epistemic trust is essential for the successful transmission of epistemic goods, and epistemic trust plays an essential in governing and/or legitimizing actions based on such epistemic goods. As an example of the latter, the restrictions on businesses and individuals during the COVID-19 crisis may potentially be acceptable to people because they have warranted epistemic trust in the scientific claims that recommend such restrictions. In this chapter, I build on an important insight from social epistemology and philosophy of science in the context of epistemic public trust which emphasizes that since laypeople often cannot assess the content of scientific claims by themselves, they rationally rely on other experts and broadly on moral and institutional contexts within and through which such claims are produced. I argue that in so far as there are indications of moral failings within practices of GHR, along

with (institutional) indicators such as possibilities of bad incentives, there are rational reasons against warranted public epistemic trust (or trustworthiness) in claims produced by GHR. To be clear, the argument here is not about the trust public has in GHR, which could be misplaced, but rather that there are aspects of GHR that provide reasons against warranted epistemic trust (trustworthiness) against claims produced by GHR. This is another example of how there is a need for responsible behaviour from the companies and corporations that constitute GHR, where such responsible behaviour spans both epistemic and non-epistemic (such as moral) aspects of the inquiry.

The next two chapters of the thesis focus mostly on the third phase of the health data value chain – usage. In the third chapter, I discuss the epistemic, and potentially ethical, responsibility of designers of digital voice assistants, such as Amazon's Alexa, through which many users receive (or may receive in the future) valuable information related to health and disease. For example, during the Covid-19 pandemic, if a user in the United States were to ask their Alexa device, " Alexa, what are the symptoms of coronavirus?" – they would be verbally given information regarding known symptoms from the US CDC's (Centre for Disease Control and Prevention) website (Porter, 2020). In philosophical terms, this phenomenon of making claims about, reporting, or affirming something is known as the speech act of "assertion". Some philosophers have argued that such instances of information sharing through machine-generated speech are equivalent to cases of humans conversationally sharing information with other, and should also be classified as assertions. This claim regarding machine assertions is partially based on the fact that instances of machine-generated speech seem indistinguishable from human speech, and advances in digital technologies are narrowing this "phenomenological" gap even further. In this chapter, I argue against the claim that machines can assert. My central argument in this chapter is that the speech act of "assertion" requires that the asserter be able to take responsibility (at least epistemic, but potentially also ethical) for the claim that is asserted. Machines, such as the Alexa, I argue fail to fulfill this condition. There is, however, a sense in which the designers of devices like the Alexa can take responsibility for such machine generated utterances, and hence, at least in some cases, such instances of machine-generated speech can be labeled as "proxy assertions". I further contend that only those machine utterances can be deemed as proxy assertions where the designers, or a collective of actors whose work influences the utterance, can reasonably foresee and therefore, take responsibility for such utterances.

From a practical point of view, one of the implications of my argument regarding machine (proxy) assertions is that designers should make it transparent to the users of devices like the Alexa, the kind of machine utterances that they can foresee and take responsibility for (or in other words, which machine utterances can be counted as proxy assertions). This transparency is especially important considering the narrowing phenomenological gap alluded to above. Empirical evidence suggests that when users deem machine speech as equivalent or very similar to human speech (because, for example, it sounds human-like), users may form similar expectations from the machine, they would have from a human regarding, for example, the accuracy of the claim (Schreuter et al., 2021). In other words, when machine speech is phenomenologically similar to human speech, users may come to expect such speech to be a product of (epistemically and perhaps even ethically) responsible action. It is important, then, that designers of devices like the Alexa ensure that users only have such expectations when the machine speech is, in fact, a result of such responsible action – which is only possible when designers can actually foresee and take responsibility for the machine-generated speech. This is particularly crucial in contexts, such as in healthcare, where the epistemic and ethical risks of unwarranted and/or inappropriate epistemic expectations can be high and problematic.

The final chapter of the thesis focus on the phenomenon of gamification in the context of health and fitness apps – that is, - the use of game-like elements such as rewards and badges given to users as motivation for physical activity. While such "gamified" apps can have a valuable motivational effect on some users, they also come with a "darker side". Sociological analyses of such apps has highlighted, for example, how such gamified apps can manipulate users into behavior that may be psychologically as well as physically harmful (Lupton & Thomas, 2015; Maturo & Setiffi, 2016). Addressing such concerns is not only of moral importance but also of significance for those interested in engagement with and the effectiveness of such apps. Existing studies that highlight the ethical challenges of gamification have met with some criticism, particularly, that they fall short of providing guidance to practitioners and designers of such apps. In other words, they fail to outline the responsibility of the designers of such gamified apps. As a response to this vacuum, this chapter seeks to facilitate a practice-relevant guide for designers of gamified health apps to address ethical issues raised by the use of such apps. More specifically, the paper has two major aims: First, to propose a practice-relevant theoretical framework outlining the responsibilities of the designers of gamified health apps. In developing this framework,

the chapter engages with existing work on ethical dimensions of gamification as well as on theoretical approaches to technological design that are centered around the idea of designing for the value of "responsibility" (van de Poel & Robaey, 2017) . The second aim of the chapter is to provide a landscape of the various ethical issues encountered in the use of gamified health apps based on a systematic literature review of the empirical literature investigating adverse effects of such apps.

Conclusion

In this introduction I have tried to provide the social and philosophical context to the changes within the healthcare system as part of the datafication phenomena. This datafication of health, characterized by an increasing amount of quantification of aspects of our lives, is particularly driven by consumer-oriented health devices. Such consumer-oriented health tracking devices, and associated data processing technologies, challenge long-entrenched social norms governing collection and flow of health information at an individual as well as the collective level. Scholars from disciplines such as sociology, law, anthropology, and science and technology studies have highlighted how such normative disruptions caused by self-tracking technologies may create new vulnerabilities for the users of such devices, as well as for society at large. While health datafication technologies do present exciting opportunities for the empowerment of its users, such users, as well as the society at large, depend on other actors to address the potential new vulnerabilities created by such technologies. The overarching aim of this thesis is to shed light on the duties and responsibilities, which could be moral, legal, epistemic, and/or social in character, of various actors associated with such consumer-oriented health devise, particularly in their actions influencing data collection through such devices as well as the purposes for which such data may be processed.

Chapter 1.
Digital Health Fiduciaries: Protecting User Privacy when Sharing Health Data

Introduction

Much has been written about the opportunities of a health revolution offered by the recent proliferation of digital devices, associated apps, and network-based platforms (Lupton, 2015). For example, health data such as heart rate, quantified physical activity, sleep quality, etc. can allow individuals to make healthier diet and exercise choices. However, the potential of digital devices in capturing health data to positively transform the health system goes beyond the individual level. In an aggregated form, with collective participation from various users, this data can offer much more valuable insights at a much larger scale. Examples of such insights include understanding of effects of various environmental factors on human health, the development of new exercise and training regimes, correlations between health symptoms and diseases, correlations between disease risk and physical activity, etc. (Lupton, 2015). Further, interpreting the significance of health data (such as that captured in a clinical setting or by a self-tracking device- heart rate, physical activity, etc.), even at the level of the individual, is often contingent upon collective participation, as it requires statistical comparison of a set of a data points (Crawford et al., 2015).

Collective participation, however, faces a conflict introduced by privacy concerns of the individual (Evans, 2011). The stakes are particularly high for health data, as inappropriate handling of health information can inflict objective harms on individuals (such as discrimination in employment or insurance or loss of reputation) as well as psychological or subjective harm (Gostin & Hodge, 2001; Konnoth, 2015). Several surveys and polling data across the developed world have shown that many individuals are concerned about health data breaches as well as misuse of breached data (Gostin & Hodge, 2001; Patil et al., 2015). A type of 'exceptionalism', in terms of the requirement of a higher level of privacy protection for health data has been recognized in past legislation, and has been reinforced by contemporary phenomena such as increased rate of medical identity theft as well as high monetary worth of

health data (Martin et al., 2017; Terry, 2012) . For users to trust digital platforms and share their health data, these concerns need to be addressed. A prominent response to these concerns has been advocacy for greater transparency and consent mechanisms, which would allow users a better understanding of and control over how their health data is used (Kaplan, 2016).

However, transparency and consent mechanisms, I argue in this chapter, are inadequate in protecting against privacy harms, or creating trustworthiness required for users to share health data, on account of the 'costs' of transparency. These 'costs' of transparency can be seen as a function of three different factors: accessibility; time required (to access and understand the information); and complexity of the information. I, therefore, further argue that digital health data controllers[1] should be recognized as fiduciaries, such that they have responsibilities that require them to keep the interests of the users at the forefront in making decisions about processing of health data. Besides compensating for the unaffordability of transparency, I argue that fiduciary duties impose deliberative requirements on fiduciaries (health data controllers, in this case) that are necessary to cater to the contextual nature of privacy. These deliberative requirements ensure that users can engage in collective participation and share their health data at a lower risk of privacy harms.

Recently, Jack Balkin and Jonathan Zittrain have suggested that online service providers should be deemed information fiduciaries (J. M. Balkin, 2015; Zittrain & Balkin, 2016). They have pointed out that such a move would require calibration of duties for different kinds of online service providers as a one-size-fits-all approach is

[1] Here I use the term 'controller' as defined by the upcoming GDPR (General Data Protection Regulation) in the European Union. According to GDPR, a controller "means the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data"; where "personal data means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person"; and "processing means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction;" ("Art. 4 GDPR – Definitions | General Data Protection Regulation (GDPR)," n.d.)

unlikely to succeed (J. Balkin, 2014). Here, I have taken up their suggestion and adapted it for the more specific case of digital health data controllers. I argue that health data controllers should be recognized, by law, as fiduciaries and outline the specific duties, as well as the scope of such duties, that digital health data controllers should have as fiduciaries.

This chapter is divided into four main sections. In section 1, I argue that transparency does not adequately protect health data subjects against privacy harms, or enable users to trust those they share personal data with. In section 2, I discuss relevant aspects of fiduciary law, including its underlying characteristics, objectives, principles, and reach. This discussion establishes the background for section 3, where I argue that the relationship between digital health data controllers and users should be recognized as fiduciary for three reasons: a.) the relationship shares key features with traditional fiduciary relationships; b.) it involves circumstances similar to those that have led to establishing fiduciary relationships in the past; and c.) fiduciary law is better suited than contractual or statutory law to protect user privacy and enable trust required for sharing health data with health data controllers. In the final part of section 3, I propose an account of the scope of fiduciary duties that digital health data controllers should owe to their beneficiaries (users sharing health data). Finally, in section 4, I present some of the gaps in fiduciary law, and highlight issues which, even if my proposal is adopted, would continue to demand our attention in the path to ensuring ethical conduct in processing of our health data.

1. Transparency

As information sharing becomes more ubiquitous, privacy trade-offs have attracted due attention. In recent years, transparency and control approaches (such as 'notice and consent' regimes) have been touted as one of the important measures to help individuals steer through privacy trade-offs (Acquisti et al., 2013; Kaplan, 2016). The argument is made that if individuals are informed about how their data will be handled (for example, what is being collected and to whom it is disclosed), then they will be able to decide their preferences regarding privacy protection and disclosure. The utility of transparency has also been recognized through, and embedded into, legislation across the developed world. In the EU, the incoming GDPR (General Data Protection Regulation) recognizes transparency as one of the central principles with regard to processing of personal data ("Recital 58, GDPR," n.d.; Spagnuelo &

Lenzini, 2016). It also states that data controllers should provide easily accessible information to data subjects ("Art. 12 GDPR," n.d.). Similarly, in the United States, the Health Insurance Portability and Accountability Act (HIPAA), through the 'Privacy Rule', demands notification about the use of health information to the respective individuals, documentation of privacy policies, as well as storing of details regarding access of information (for example, who has accessed the information) (Farrell, 2012; Spagnuelo & Lenzini, 2016).

The success of transparency, in alleviating individual privacy concerns, as well as in promoting sharing of data for beneficial purposes (such as research), however, seems limited. The HIPAA Privacy Rule, for example, has been criticized both for allowing too much access to data (as individual concerns regarding sharing of their data for research were not addressed) as well as for restricting sharing of data for useful research (as consent requirements impeded sharing of data) (Evans, 2011).

One of the limitations of transparency, I argue, is its cost. This cost can be seen as a function of accessibility, the time required (to access and understand the information), and complexity of the information.

In the case of digital health information, there are few barriers to accessibility as information (regarding privacy policies and use of data) can be made readily available on the platform. Most digital platforms, by legislation or on a voluntary basis, already do share their privacy policies with the users along with certain control mechanisms (such as the 'notice and consent' forms). However, these privacy policies are still 'costly' to users on account of the time required to read them. Studies have shown that users generally do not read these privacy policies, or do so infrequently (McDonald & Cranor, 2008). In their own study, McDonald & Cranor (2008) estimated that the annual opportunity cost, in the US alone, for just reading privacy policies of online websites would be in the order of $781 billion.[2]

The actual costs of transparency may actually be much higher, as privacy policies, for most individuals, are hard to read and understand (Jensen & Potts, 2004). While data controllers may fulfil their legal obligations related to transparency by providing

[2] The study conducted by McDonald & Cranor (2008) asked 212 participants to skim through online privacy policies and answer simple comprehension questions. It estimated the value of time as 25% of average hourly salary for leisure and twice wages for time at work in the US.

'notice and consent' forms and privacy policies, individuals may still be uncertain about what they are consenting to (Barocas & Nissenbaum, 2009). This is largely due to the subjective complexity dimension of transparency.

Candeub (2013) provides the example of Sherlock Holmes as a good illustration of the subjective nature of transparency's complexity (or 'computational', the alternative nomenclature used by the author) dimension. In the movie *The Seven-Per-Cent Solution,* Dr. Watson deceitfully arranges a meeting between Dr. Sigmund Freud (relatively unrenowned in the timeline of the movie) and Sherlock Holmes in Vienna. Dr. Watson hopes that Freud would be able to cure Sherlock of his cocaine addiction. When the two meet, Freud, intending to induce a reflection upon Holmes' addiction, asks him, "Who am I, that your friends should wish us to meet?" ((*The Seven-Per-Cent Solution*, 1976) quoted in (Candeub, 2013))

Holmes, defeating the question's intended effect, responds with an exhibition of his deductive skills, "Beyond the fact that you are a brilliant Jewish physician who was born in Hungary and studied for a while in Paris, and that certain radical theories of yours have alienated the respectable medical community so that you have severed your connections with various hospitals and branches of the medical fraternity, beyond this I can deduce little. You're married, with a child of... five. You enjoy Shakespeare and possess a sense of honour." ((*The Seven-Per-Cent Solution*, 1976) quoted in (Candeub, 2013))

While these facts about Freud were transparent to Sherlock Holmes through the objects in Freud's study, for most other people the same objects would not have made these facts transparent.

While privacy policies may not require a rare genius of Sherlock's capacity to be understood, they do pose a serious challenge for those not well versed with legal terms, the technical know-how related to data analytics, as well as privacy implications of the terms enlisted. Barocas & Nissenbaum (2009) further the claim, by arguing that (current and future) uses of data, to a degree, may not only be difficult to understand, but rather unknowable. This unknowability, they claim, follows from the uncertain chain of events linked to the use of data, such as emergence of new technologies (for example, new analytical tools or advanced algorithms) and new actors (with unknown intentions).

The limits of transparency are further exposed by research indicating that transparency and control might paradoxically increase disclosure of sensitive information (Acquisti et al., 2013). Consent mechanisms can also exploit (known and still unknown) cognitive biases, such as limited attention span, framing effects, and decision making heuristics, in how people interpret and act on available information (Acquisti et al., 2013; Kahneman & Tversky, 1979). For example, Adjerid, Acquisti, Brandimarte, & Loewenstein (2013), in a series of experiments, demonstrate how simple misdirections can alter a subject's perception of privacy risks, even though the objective risks (and corresponding facts) are not altered.

Eventually, rather than being empowered by transparency, the individual (consenting to share their data) has to take a leap of faith and rely on the assumption that the actors involved in the use of an individual's data will be committed to a set of ethical principles, professional commitments, and guiding norms and regulations which protect the individual from privacy harms. When the data being shared is particularly sensitive, such as in the case of health data, the individual is in a rather vulnerable position relative to the data controller, decreasing the incentive to share data, even for beneficial purposes such as research that may led to the discovery of new preventive or treatment mechanisms for various diseases.

2. Fiduciary relationships

Sharing of personal or sensitive information for individual or social benefits is not unique to the contemporary digital platforms. Individuals also share sensitive information with doctors, accountants, and lawyers. In such cases, these professionals are bound by duties which restrict the use of such sensitive information in ways that can be harmful to, or against the interest of, the individual. These duties are established through the notion of 'fiduciary responsibility' assigned to some types of professionals, such as doctors and lawyers, with whom sensitive information is shared (Frankel, 2010). In this chapter, I argue that digital health data controllers should also be assigned an information fiduciary role, wherein, they are required to keep the interests of data subjects at the forefront, particularly regarding the protection of privacy. However, before presenting an account of why health data controllers should be given such a role, and what that might entail, I will offer a discussion of fiduciary relationships and fiduciary responsibilities in general. This discussion will highlight

key features of fiduciary relationships as well as conditions under which those features are advantageous compared to other legal instruments such as contracts.

2.1. The nature of fiduciary relationships

Courts recognize fiduciary relationship of various kinds, including, as already mentioned, doctor-patient, attorney-client, trustee-beneficiary relationships. Yet, there seems to be no consensus on a definition of fiduciary relationships (Frankel, 2010). While some claim that a lack of definition makes fiduciary law "elusive" (D. G. Smith, 2002), others argue that the lack of definition is incidental, or even a necessary, aspect of fiduciary law's "situation-specificity and flexibility" (Rotman, 2011, p. 941). Courts have therefore, based their judgements on particular facts of a case, recognizing the difficulty of providing a universally applicable definition (Frankel, 2010). In one case concerning fiduciary law, for example, the English court of appeals remarked that the court "has always been careful not to fetter this useful jurisdiction by defining the exact limits of its exercise" (Rotman, 2011, pp. 940–941).

Despite this lack of a common definition, fiduciary relations do have some common elements. These elements include:

1. Fiduciaries offer services (rather than products) that are socially desirable (Frankel, 2010).

Fiduciary relationships usually involve an expertise being offered as a service to those who rely on the fiduciary. Typically, without the relationship (between those offering expert service and those availing themselves of it) being established as fiduciary in nature, the services would not be able to produce the desirable social effect (in degree or in kind). For example, a client would not be able to trust their attorney with personal information and attorney's advice, without there being a fiduciary relationship between them (where the attorney has a duty to keep the best interest of the client at the forefront).

2. Fiduciaries are entrusted with a discretionary power over the interests of the beneficiary

Typically, in a fiduciary relationship, the fiduciary agent acts 'on behalf of' the beneficiary (Licht, 2016; D. G. Smith, 2002). The beneficiary entrusts a 'critical resource' or power to fiduciaries, where the fiduciaries are required to act in the interest of the beneficiary (Frankel, 2010; D. G. Smith, 2002). The critical resource may be tangible, such as property or finances, or intangible, such as personal information (such as health details disclosed to a doctor). The entrustment is to enable or facilitate the fiduciary to deliver their services.

3. Fiduciary law (through assigning fiduciary duties and obligations) counters the asymmetrical power relationship between fiduciaries and beneficiaries and protects the beneficiary against opportunism (Licht, 2016).

As mentioned above, fiduciaries are entrusted to act on behalf of the beneficiary, giving them control over the interests of the beneficiary. This power over the beneficiary's interests introduces a power asymmetry between the fiduciary and the beneficiaries and gives rise to a common problem among the fiduciary relations, opportunism (Licht, 2016). By requiring the fiduciary to act in the interest of the beneficiary, fiduciary law hinders those with a propensity for being self-interested or opportunistic at the expense of the beneficiary, when entrusted with discretionary power over someone's interests.

2.2. Why are fiduciary relationships established?

While the elements described above are common features of fiduciary relationships, they do not fully explain the distinction between fiduciary and non-fiduciary relationships. For example, several non-fiduciary relationships are also based on expert services being offered to clients, where the experts can exploit their authoritative or informational advantage for their own benefit. Electricians, plumbers, teachers, are all examples of professions that provide such services, which do not have fiduciary status. A plumber, for example, may advise you to install a new faucet, even though you may not need one, simply for their own benefit. A doctor, on the other hand, on account of his fiduciary duties, may not ask you to undergo a surgery that you don't need, just because the doctor will earn more money out of it (Drozd & Dale, 2006). Why then do we require that some experts have an obligation to keep the interests of their client at the forefront? In other words, why then are some relationships deemed fiduciary by law while others are not? This section aims to

highlight some of the justifications for as well as advantages of fiduciary law over other legal instruments, offered by courts and legal scholars.

Historically, acknowledgement of fiduciary relationships can be categorized in two ways (P. Miller, 2011):

a. Status-based
b. Fact-based

As the name suggests, status-based fiduciary relationships are determined through status. If a relationship falls under a category that has conventionally been recognized as fiduciary, then it is deemed as fiduciary. Examples of conventional fiduciary relationships include doctor-patient, attorney client, and director-companies. The conventional status of these relationships descends from English equity courts during and shortly after the middle ages, which deemed a relationship as fiduciary if it was similar to trustee and *cestui que trust*[3] (P. Miller, 2011). For example, (Worthington, 2006) writes:

> [F]iduciary law evolved from Equity's regulation of the relationship between trustees and beneficiaries. Over time these rules were extended, with minor modifications, to cover other situations that seemed analogous. Now it is accepted that relationships between directors and their companies, agents and their principals, solicitors and their clients, and partners and their co-partners are all fiduciary. These are all 'status-based' fiduciary relationships. The status itself inevitably attracts fiduciary impositions.

A number of courts have since, however, raised objections to the status-based approach to determination of fiduciary status. Justice Dickson, for example, stated in Guerin[4], "It is the nature of the relationship, not the specific category of actor involved that gives rise to the fiduciary duty" (P. Miller, 2011). Similar arguments have since led to efforts to define fiduciary principles, such that facts, rather than status, can be used to determine fiduciary relationships. Defining these principles also allows for making decisions about relationships that are new (such as between health data controllers and users) or may arise in the future.

---

3 Archaic term in English law for beneficiary under a trust (*Cestui Que Trust*, 2006)

4 (*Guerin v. The Queen*, 1984) was a landmark case regarding Aboriginal rights in Canada, where the Supreme Court stated that the government had a fiduciary duty towards the First Nations of Canada.

While there doesn't seem to be a consensus on what facts or conditions are necessary and sufficient for determination of a fiduciary relationship, a number of such conditions have been offered. Here I discuss some of the most important proposed conditions, and their possible limitations:

1. Power-dependency and vulnerability

As discussed earlier, fiduciary relationships involve entrustment of discretionary powers to the fiduciary, which they shall use to the interest of the beneficiary. Justice Wilson argues that certain features, which are common among fiduciary relationships, should be used as a criteria to determine other fiduciary relationships (P. Miller, 2011). These common features, which can be used as identifying characteristics of fiduciary relationships, according to Justice Wilson, include: a.) a scope for exercise of unilateral discretionary power by the fiduciary over beneficiary's interest; and b.) vulnerability of the beneficiary to the fiduciary holding the discretionary power.

These criteria have, however, met with some criticism, both within and outside courts, as being insufficient reasons for establishing a relationship as fiduciary. Justice Cromwell, for example, argued that not all power-dependency relationships have been and can be deemed fiduciary in nature (P. Miller, 2011). (Biological) parents, for example, are not deemed as fiduciaries, even though children are dependent upon them (Brinig, 2011). Similarly, vulnerability, as an indicium for fiduciary relationships, seems too broad and imprecise.

2. Enabling trust in relationships

Beneficiaries place significant trust in fiduciaries by giving away access and control over their resources. By demanding fiduciaries to act against self-interest, and in the interest of the beneficiaries, fiduciary law plays an important role in enabling the beneficiary to trust agents with discretionary power over them.

However, the idea that trust, or need for trust, can be sufficient in itself for establishing fiduciary relationships is also not without its problems. First, contract law can also enable trust between parties by establishing the rules within which the parties must act. Rotman (2011) states that fiduciary law protects not just any relationships requiring trust, but only those that require *high* trust and confidence. This distinction

between the need for trust and high trust seems wanting of further elaboration and support. Second, as courts have recognized in some cases, trust may also be *misplaced*, where individuals should not have had expectations of behaviour in their interest (Brennan-Marquez, 2015). Therefore, using trust as a criterion for establishing fiduciary relationships would require an explanation of why trust is warranted for that specific relationship between a fiduciary and beneficiary.

3. To support equity, as anti-opportunism

The origins of modern fiduciary law, as discussed earlier, can be traced at least as far back as the English equity courts of the fourteenth century (P. Miller, 2011; H. E. Smith, 2013). The function of equity courts was to hear pleas where the law seemed limited in its invocation of justice, on account of being too general, or where it seemed unable to cater to the specific circumstances of a particular case (H. E. Smith, 2013). These courts drew from the principles of equity as defined by Plato and Aristotle, which were meant to plug gaps in laws (Rotman, 2011; H. E. Smith, 2013). These gaps in laws existed, according to Plato, because laws aimed for being certain and universal, while the human condition tends to lack universality and certainty (Rotman, 2011). Many of the hearings in the English equity courts were for charges against "feoffees", who were persons holding legal title on behalf of others in a quasi-trust agreement (H. E. Smith, 2013). The courts, through several hearings, held that feoffees should not be opportunistic and "faithless", in taking the entrusted property to themselves (H. E. Smith, 2013). H. E. Smith (2013) and Frankel (2010) provide accounts of how modern fiduciary law originates out of these early litigations against opportunism, and to enable trust between parties.

However, the idea of fiduciary law as a tool against opportunism warrants more examination for a number of reasons. First, if opportunism is "self-interest seeking with guile" as defined by (O. E. Williamson, 1975), then examples of opportunistic behavior seem much more frequent than the number of cases where fiduciary law is applied. A number of non-fiduciary economic actors, for example, seek self-interest with some guile, without requiring the courts to intervene through fiduciary law. Secondly, as argued before, if an agent is entrusted with power or resources to warrant opportunism, the said agent could also be restricted through contract law. Why then do we require the category of fiduciary relationships? In order to answer that question, we need to examine if fiduciary law has distinct advantages over contract or statutory

law in countering opportunism as well as under what conditions those specific advantages can be useful.

Fiduciary law's advantages over contract or statutory law in countering opportunism

Fiduciary law seeks to prevent not just any opportunism, but as Smith (2013) has argued, rather opportunism that is "hard to capture ex-ante" and thus, cannot be countered by general rules. As Smith (2013) notes, this difficulty in capturing opportunism ex-ante is more than a difficulty in description, particularly in cases where an agent has discretionary powers over the others. In such cases of discretionary authority, the agent has a pre-existing informational advantage over the principal. This informational advantage may involve three types of information*: costly, unobservable and unverifiable* (Licht, 2016). Information may be costly, for example, when the principal may not be able to monitor the actions of agent (with discretionary power) as it might be too expensive to do so. With discretionary power, including access, to the principal's resources, the agent may act in ways which may make it difficult for the principal to observe at all, or to observe the circumstances around the actions of the fiduciary to judge whether the agent acted in principal's interest. Finally, even if the principal were to know the circumstances and the actions of the agent, lack of expertise may make it difficult for the principal to judge whether the agent has breached their duties.

While contract law or statutory law (in the form of regulations, for example) can be useful in preventing some types of opportunism, it is not useful in limiting opportunism that cannot be detected (for example, on account of costly, unobservable or unverifiable information). Within economic literature, this problem is often referred to as *incomplete contracting*, referring to the impossibility of anticipation of all future contingencies as well as the infeasibility of codifying instructions to counter all anticipated contingencies (Sitkoff, 2011). Further, unlike Williamson's proposed definition of opportunism (as seeking self-interest with guile), opportunism may not always be in the form of a full blown planned deceit. Unexpected circumstances may generate unexpected opportunities for an agent entrusted with power or resources, without them actively seeking such opportunities. In this regard, the use of equity against opportunism requires an open-endedness, which can fill the gaps of prescriptive principles contract laws work on. As I argue later in section 3, dealing

with the contextual nature of privacy is one example where fiduciary law can be more advantageous than a prescriptive contractual approach.

Fiduciary law, thus, provides the legal system with a way to counter opportunism which cannot be completely dealt with through contract law. The open-ended approach required to guard against opportunism, that cannot be detected ex-ante, also partially explains the lack of consensus among courts and scholars regarding a common definition for fiduciary relationships and why fiduciary law seems like a "concept in search of a principle" (P. Miller, 2011).

2.3. The nature and scope of fiduciary duties

Owing to the open-ended nature of fiduciary law, fiduciary duties are abstract, lacking a consensus on a common definition. Nonetheless, legal scholars and courts generally tend to recognize two fiduciary duties, broadly defined: duty of loyalty and duty of care (P. Miller, 2011). While there is broad consensus on there being a duty of loyalty (though there may be issues with what exactly that duty entails) (Gold, 2013), the duty of care is more controversial.

Miller (2011), recognizing the controversial status of duty of care, defends it as an important fiduciary duty. As the main argument against the inclusion of duty of care as a fiduciary duty, he cites it being "indistinguishable in substance from tort duty"[5] (P. Miller, 2011, p. 55). Miller argues that unlike tort duty of care, which prescribes conduct to avoid foreseeable harm, duty of care within fiduciary law requires *diligence* and *skill*. That is, fiduciaries are not only required to not cause harm to the beneficiary, but they are also required to use their expertise to the best of their knowledge to make sure that the beneficiaries are not harmed.

There is, however, at least one other argument against the inclusion of duty of care as a fiduciary duty: the duty of care, particularly, as defined by Miller, seems too expansive and difficult to carry out for a fiduciary. Or, in other words, it makes it too easy for a beneficiary to claim a breach. Smith (2013), for example, argues that the duty of care opens the opposite door for opportunism, the one for the beneficiary. With the requirements of diligence, skill, and putting forward one's best efforts, the

[5] Tort law, simply defined, is common law that recognizes legal liability for someone who causes harm to another in the form of a civil wrong (Dobbs, 2008).

beneficiaries can claim a breach just for profit, and without any true injury caused to them. Further, as I will explain below, the duty of loyalty can also require the fiduciary to play a more active role in ensuring the interests of the beneficiary are kept at the forefront.

There is broad consensus that the duty of loyalty is central to fiduciary duties, and fiduciary law in general. However, there is some debate on what this duty entails, and particularly to what 'degree' a fiduciary should be loyal to their beneficiary. That is, how far should a fiduciary go in pursuit of beneficiary's interests (and in avoiding fiduciary's own self-interest)? Here, I take cue from Lyman Johnson's work, where he argues that fiduciary loyalty involves two conditions: minimum and maximum (L. Johnson, 2003). On a similar note and using the discussion presented by Gold (2013), I present two distinct notions of loyalty, which can be taken as minimum and maximum conditions.

1. Loyalty as avoidance of conflict (Minimum condition) - Miller (2011) describes loyalty (or 'faithfulness') as avoidance of conflict. Here, he makes distinctions between two types of conflict avoidance, both of which are deemed necessary fiduciary obligations.

   The first is avoiding conflict of interest, where the fiduciary avoids the conflict between their pursuit of beneficiary's interest and their self-interest. The second type of conflict avoidance is avoiding conflict between the fiduciary's duties to the beneficiary and the fiduciary's pursuit of other people's interests.

   The anti-conflicts rule can be seen as minimum core of fiduciary duty of loyalty, even though a narrower version of this rule exists (Gold, 2013). Under this narrower version, only avoidance of conflict of interest is seen as necessary, while fiduciaries are not required to have undivided attention towards one beneficiary.

2. Loyalty as affirmative devotion (Maximum condition) – While the anti-conflict rule only increases chances of the fiduciary ensuring the best interests of the beneficiary, loyalty as affirmative devotion requires that a fiduciary does so. A number of court cases, particularly in the United States, have identified fiduciary loyalty as one of affirmative devotion (Gold, 2013). In this conception

> of loyalty, the fiduciary is required to play a more active role in pursuit of the beneficiary's interest, producing a similar effect as intended by Miller's conception of diligent and skillful duty of care. As argued before, such a duty can be difficult to enforce, particularly for epistemic reasons, as it is difficult to know and judge whether a fiduciary had an affirmative devotion toward the beneficiary. It can still be a legal duty though, enforceable only in rare circumstances: where the court can deem a breach to have occurred, for example (Gold, 2013).

Besides these two conceptions of loyalty, there are a number of other conceptions of loyalty offered by various scholars, who also disagree on what should be the core minimum of fiduciary loyalty (Gold, 2013). Some, for example, have argued that affirmative devotion should be seen as the minimum core of fiduciary loyalty. (Gold, 2013) provides an account of various conceptions of fiduciary loyalty, demonstrating that no single account is universal enough to be deemed as a minimum core. This, however, does not entail that there is no duty of loyalty, only that such a duty is abstract and depends on the circumstances of a particular relationship. As (Gold, 2013) points out, the under-determined nature of the minimum core of fiduciary loyalty does not mean it is an empty vessel. Rather, it points to a pluralism within fiduciary law, which may require reassessment of existing, and more precise formulations of new, specific fiduciary relationships. This pluralism can be embraced and utilized, once the idea of needing a specific conception of fiduciary loyalty can be rejected. An abstract conception of fiduciary loyalty allows for a more dynamic approach to fiduciary duties in specific settings, such that they can be reassessed with time, particularly when there are changes in socio-technical structures. This chapter attempts to provide a basis for the need for fiduciary loyalty on part of digital health data controllers as well as specify what such a duty of loyalty should entail.

3. Digital health data controllers as fiduciaries

3.1. Arguments for recognizing digital health data controllers as fiduciaries

In this section, I present three main arguments for recognizing the relationship between digital health data controllers and users sharing their health data as fiduciary: a.) the relationship shares features with traditional fiduciary relationships; b.) the relationship involves circumstances similar to those that have led to establishing

fiduciary relationships in the past; and c.) fiduciary law is better suited than contractual law in protecting user privacy and enabling trust required for sharing health data with data controllers.

Before I expand on the arguments for recognizing health data controllers as fiduciaries, however, it is important to discuss what is meant by 'health data'. As discussed before, previous legislations in the developed world have afforded a higher level of privacy protection for health data (Bywater & Armstrong, 2015; Terry, 2012). Yet, these legislations, such as the EU data protection directive (DPD), which has now been superseded by GDPR, do not define health data (Bywater & Armstrong, 2015). Defining health data can be particularly hard in the present context, where 'health' apps collect a variety of data (such as location data) which may or may not reveal the health status of a person. While providing a full discussion on definition of health data, and its precise formulation, is beyond the scope of this chapter, the definition proposed by Article 29 Working Party (2015) is useful. According to this proposal, personal data qualifies as health data when it meets at least one of the following criteria:

1. It is clearly/inherently medical data
2. It is raw sensor data which can be independently, or in combination with other data, used to draw conclusions about health status or health risk of an individual
3. It allows for reasonable conclusions to be drawn about an individual's health risk or health status, irrespective of the accuracy, legitimacy, or adequacy of these conclusions[6]

One problem with this definition, which the Article 29 working party also notes, is that it may make the definition of health data seem too broad. Given the argument of this chapter, one might worry that such a broad definition would impose fiduciary duties on an overly wide range of data controllers (Article 29 Working Party, 2015).

[6] While this clause may make the definition of health data employed here seem broad, the working party argues that it actually excludes a category of personal data from being categorized as health data (using the criterion that these conclusions be reasonable and about the specific individual). For example, data about number of steps taken by the data subject in a single walk, without being combined other data about the same data subject, would not divulge health risk or status of the specific data subject and therefore, would not be regarded as personal health data. For a more detailed account of the merits and demerits of this definition, see the annex to (Article 29 Working Party, 2015)

One important merit of this definition, however, is that it is able to include data controllers who collect data outside traditional healthcare settings. This is crucial as in this digital age a lot of health data, worthy of protection, is collected outside traditional health settings.

In order to reach a balance between not making the definition too broad, while also included data controllers who collect data through, say, mobile apps and wearable devices, I propose that fiduciary duties be imposed on health data controllers who a.) Process data with the intention of using the data to determine the health status of a specific person[7], or b.) Collect raw data in situations where it will be reasonable for a data subject to conclude that the data is being collected to determine their health status. The first criterion is to ensure that raw data which may not seem to be health related in an obvious way, but is then used in a way that the health status of the data subject is revealed, is also protected. Raw data, which may not seem like health data, when collected over long periods of time, or combined with other data, for example, may reveal health status of specific individuals and needs protection. At the same time, according to this criterion, data controllers who process such raw data, but do not intend to use it to determine health status of a specific person, would not be charged with fiduciary duties. Yet, there is a risk here that some data controllers may collect sensitive health data, which would be worthy of protection, but claim that they do not intend to use it to determine health status of specific subjects. This could, for example, be the case with data collected through sensors on mobile or wearable devices, where the data subjects may reasonably conclude that the data is collected for health related purposes (because, for example, the marketing of the device may suggest that data is being collected in the interest of individual or public health). The second criterion I have proposed plugs this loophole.

With this working definition, I argue that digital health data controllers share features of traditional fiduciaries in that they offer socially desirable services and enjoy a significant advantage over the users from whom they collect health data. There is an asymmetrical relationship between the users and the digital health data controllers, as users typically lack expertise, information about digital health data controllers as well as information about the actions digital health data controllers might take with the user data. This vulnerability of the users relative to the digital health data controllers

[7] Article 29 working party also proposes a similar criteria but does not include it in their definition of health data (Article 29 Working Party, 2015)

can be seen as grounds for establishing a fiduciary relationship, as has been argued by some scholars and courts.

As discussed earlier, fiduciary relationships are also established on grounds of enabling trust. Digital health data controllers, in some cases, also put themselves forward as trustworthy organizations that will not misuse user data and present themselves as acting in the interest of their users (for example, "Fitbit Privacy Policy," 2016). At the same time, digital health data controllers do not disclose full details about their handling of our data (and sometimes for good reasons such as security (as disclosing detailed data security measures can be jeopardizing) and competitiveness). This incomplete disclosure, coupled with the high costs of transparency, can create a lack of trust among the users, eventually leading to non-participation (by not sharing data, for example) in the promised digital health revolution. Fiduciary relationships between the users and digital health data controllers, where the latter is required to act in the interests of the users, can therefore, be valuable in making data controllers trustworthy and facilitating collective participation.

The need for establishing trust and compensating for vulnerability, however, as argued earlier, may not be sufficient for establishing fiduciary relationships, even though they may have advantages. The third and most important reason for establishing fiduciary relationships between data subjects and data controllers with whom health data is shared, I argue, is that fiduciary relationships are better suited than contractual or statutory obligations (such as those associated with privacy agreements users click 'agree' on their digital devices (contractual) or defined through legislation (statutory)), for protection of user privacy or for balancing protection of privacy with other goals related to societal interests.

To this end, I argue that stringent privacy protection is difficult to achieve through prescriptive legal measures, such as those possible through contracts or privacy agreements. Even if the users were able to afford the costs of transparency, and give informed consent for the use of their data, the changing nature of technology would still leave the door open for privacy harms and opportunism by those who want to cause these privacy harms. As discussed earlier, fiduciary law, as opposed to contracts, affords the kind of deliberative and strategic interaction required to guard against the opportunists. Privacy is contextual, and depends on multiple factors, such as the nature of information, the context it is shared in, prospective users of that information,

etc. (Nissenbaum, 2011a; Solove, 2007). Fiduciary law allows for the flexibility required to cater to the contextual nature of privacy. Here, I will use security, anonymization and data minimization as examples of contextualization and flexibility required to deal with privacy issues. These, however, are just examples, and not an exhaustive list of cases where decisions and methods for privacy protection require contextualization.

Securing user data, an integral aspect of privacy protection, requires diligence and regular upgrading of security measures against cyberattacks and hacks. The recent case of the cyberattacks on the UK's National Health Service computers with the ransomware WannaCry is a case in point (Martin et al., 2017). Systems were largely found vulnerable because of a failure to upgrade software, rendering them unable to cope up with the ransomware (Martin et al., 2017). Health data, as discussed before, is particularly valuable to cyber attackers and healthcare is, therefore, one of the most targeted sectors in terms of cyberattacks (Athinaiou, 2017; Martin et al., 2017). Securing health data, thus, requires diligent measures, which can guard against an opportunist hacker who may exploit vulnerabilities in a digital system. Data security may also require some secrecy or incomplete disclosure of data security policies (to keep them secret from hackers, for example). It can be difficult to counter such opportunism through use of contracts which specify what steps health data controllers need to take to secure user data, as it will be hard to anticipate all future contingencies (such as new tools for hackers or changes in security technologies). Fiduciary law, on the other hand, because of its open-ended approach and deliberative requirements (through the duty of loyalty) can be helpful in ensuring that health data controllers take appropriate measures to secure user health data. Fiduciary law can also help increase data sharing by not prescribing expansive security requirements for controllers who are collecting less sensitive or easily securable data.

Another crucial aspect of privacy protection for electronic data is anonymization (Ohm, 2009). Anonymization aims to make re-identification of data subjects impossible, such that data can be shared for useful purposes, in an aggregated form, without the risks of privacy harms. The importance of anonymization or de-identification (either one or both), has also been recognized in and embedded into legislation, such as through the European Union's GDPR and HIPAA in the United States (Hintze, 2017; Yakowitz, 2011). These laws often prescribe techniques for anonymization, such as removal of personal identifiers (such as names, phone

numbers, social security numbers, etc.) (Ohm, 2009; Yakowitz, 2011). However, recent studies have shown that such prescriptive techniques may not be adequate, as computer scientists were able to re-identify individuals from anonymized data stripped off personal identifiers (Narayanan & Felten, 2014; Ohm, 2009). Stringent anonymization may therefore, require contextualization such that data is also stripped of indirect identifiers or is randomized, depending upon the kind of data that is collected (Ohm, 2009). Further, the risk of re-identification may not be the same for all kinds of data, and for some data, it may be enough to apply techniques that make re-identification complex enough to take away the incentives for re-identification (Yakowitz, 2011).

Another problem with prescribing anonymization through legal measures is that anonymization may not even be desirable for some kinds of data. Evans (2011) points out that anonymization may render linking data longitudinally impossible. Longitudinal health data, collected across different health environments, can be invaluable in generating insights for an individual as well as on a more general level, for example, by helping researchers determine the correlations between different biological factors and enable more organized efforts to tackle health and social problems (Evans, 2011; Holman et al., 2008). Requiring anonymization for all health data may take away the opportunity to assemble longitudinal data for research as well as for other uses wherein the data subject may benefit without serious threats to their privacy.

Thus, as in the case of securing user data, anonymization too requires contextual decision making. Such contextual decisions can be hard to codify in the form of contracts, which would have to anticipate all future contingencies in all possible contexts. As fiduciaries, digital health data controllers would be able to make contextual decisions about anonymization, where they can decide whether or not anonymization is needed, and to what degree.

Finally, as a third example of the advantages of a contextual approach to privacy, consider data minimization. Data minimization as a principle has also been included in the GDPR and states that data must be "limited to what is necessary in relation to the purposes for which they are processed" ("Art. 5 GDPR" n.d.). In addition to the scope of data collected, the minimization principle within GDPR also relates to the time for which it is retained and stored ("Recital 39" n.d.; Zarsky, 2016). The

minimization principle can be important in protecting user privacy by limiting the opportunities for collecting irrelevant data as well as minimizing cyber security risks by requiring controllers to delete data when no use is intended. However, in the age of big data analytics, an ex-ante analysis of relevance of data and restrictions on its retention can severely limit the benefits of big data analytics. This has also been noted by other commentators (see Zarsky (2016)) while some have also predicted that a requirement such as data minimization is likely to be breached (Rubinstein, 2012)[8]. Again, a contextual approach to privacy, as made possible through a fiduciary approach, can achieve a better balance between privacy protection and achieving benefits of big data analytics, by loosening the data minimization or replacing it by achieving the intended effects of minimization through other means wherever necessary. While contractual law and statutory law (such as GDPR) also can (and do, in case of GDPR[9]) have context-sensitive features, a fiduciary approach can enable more flexibility in fulfilling data controllers' obligation of protecting user privacy, particularly in allowing data controllers to choose the most appropriate method of doing so while ignoring recommendations that may be counter-intuitive or disadvantageous in the given context.

Further, as fiduciaries, digital health data controllers would not only be required to take a contextual approach to privacy protection, but also not deceive or actively harm the data subject in pursuing their obligation to protect the privacy of data subjects. This is an advantage over contract or statutory law, which may leave room for opportunistic or deceptive behavior on part of data controllers (see for example Wachter (2018 and Zarsky (2016) for examples of loopholes in GDPR which data controllers might use for their benefit and which may deny rights to data subjects exposing them to risks).

Here, I have outlined how fiduciary relationships between health data subjects and health data controllers can enable collective participation by ensuring better decisions

---

8 GDPR does allow some exceptions for application of the minimization principle, but these exceptions also have problems and may not be applicable for a variety of big data analyses (see Zarsky (2016) for a more detailed discussion of limits of data minimization principle as included in the GDPR).

9 For example, Article 25 ("GDPR – Data protection by design and by default," n.d.) states that data controllers should take into account "state of the art, the cost of implementation and the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for rights and freedoms of natural persons" in determining the appropriate measures in implementing privacy by design.

are made concerning data on behalf of the users. Fiduciary relationships not only compensate for the high costs of transparency, but are also better suited than alternative approaches as they can flexibly contextualize privacy (and privacy protection).

3.2. Nature and scope of duties and obligations that health data controllers should have as fiduciaries

As argued in section 2.3, central to fiduciary law is the duty of loyalty, which primarily dictates that fiduciaries must keep the interests of the beneficiaries at the forefront. Yet, as argued earlier, scholars and courts do not share a consensus on the scope of such a duty, that is, how far should the fiduciaries go in pursuit of beneficiaries' interest. The duty of loyalty can range, for example, from avoiding conflict of interest to an affirmative devotion towards the beneficiary.

The abstract and open-ended nature of fiduciary duty of loyalty, however, as I argued earlier, does not render it an empty vessel. Rather, it opens up the possibility for pluralism within fiduciary law and for more precise formulations of specific fiduciary relationships. At the same time, if the duty is too expansive, within a specific fiduciary relationship, then the duty will be too difficult to carry out. The open-ended and abstract nature of fiduciary duty, therefore, needs to be balanced with specificity about the interests of the beneficiary that the fiduciary should pursue within a specific fiduciary relationship.

The proposal that we specify the scope of fiduciary duty, such that there are bounds to fiduciary loyalty, is not unique and is also applied to traditional fiduciaries. For example, physicians are not expected to be loyal to their patients at all costs. A physician, for example, is only obligated to provide care to a patient at a reasonable time and place (for instance, a physician is not obligated to attend to night or house calls) (Mehlman, 2015).

Courts also recognize similar limits to the degree of loyalty physicians owe to their patients. That is, while the physician is expected to keep the patient's interest ahead of his own interest, courts recognize that there should be reasonable limits to the expectation of such loyalty from the physician. For example, while physicians cannot deny treatment pending assurance of payment in urgent situations, they may terminate their relationship unilaterally (even on financial grounds) with the patient

as long as the patient is given notice and reasonable opportunity to get treatment elsewhere (Mehlman, 2015).

It is therefore, important to specify the bounds of fiduciary duty that health data controllers have towards their data subjects. First, as an essential part of duty of loyalty, health data controllers should not use information collected by them to harm individuals, for example, by harassing, exploiting, embarrassing, or manipulating them. Beyond this primary requirement, I argue here that the duty of loyalty for health data controllers should be specifically about protection of privacy of the users sharing their health data. Catering to individual privacy concerns is an important step in enabling trust within the users to share their data, and thus, opening the way for collective participation in the digital health revolution. As argued, in the previous section, privacy protection requires contextualization, wherein the type of data as well as the technologies involved in collection, storage, and sharing are taken into consideration. The duty of loyalty, aimed specifically at protecting privacy of users, thus, still requires deliberation and diligence on the part of health data controllers.

At the same time, defining the duty of loyalty as specifically aimed at privacy protection avoids the danger of making the duty too expansive, and the corresponding difficulties of carrying out such a duty. An expansive duty of loyalty, such as one requiring a general affirmative devotion to the user, might take away the incentive for health data controllers to invest in digital health technologies, and thus, hamper the path towards a better healthcare system.

For example, an alternative possibility to the scope of fiduciary loyalty proposed here, would be to require that fiduciaries go beyond protection of privacy, and also ensure that a broader or general set of interests of the users are kept at the forefront when sharing health data with third parties, even in anonymized and de-identified form. This would, for example, require that data is shared only for purposes that are beneficial to the users. Such an expansive requirement, however, would put too much burden on health data controllers to evaluate the outcomes of the data shared by them with the third parties. It would also significantly reduce the incentive for health data controllers to share data, even in an anonymized form, for health research, as that might open up a possibility for claims of breach by users who may not find the aim or outcomes of the research in their interests. This is not to argue that health data controllers should be allowed to share data with any third party. Rather, the lawful

basis of sharing data with third parties should not be determined solely through fiduciary duties (which could lead to more abstract and expansive definition of fiduciary duties), but also through legal instruments such as those already implemented (Long, 2017).

3.2.1. Fiduciary breach vs medical malpractice

In the previous part of this section, I claimed that health data controllers themselves should not use information collected by them to harm individuals, for example, by harassing, embarrassing, or manipulating them. Not causing harm to the beneficiary is an essential part of fiduciary duty and without such a requirement, users would not be able to trust the health data controllers, even if they are assured that their data would not be shared with third parties in an identifiable form. However, I claim here that a distinction should be made between harms caused by medical advice provided by health data controllers and other harms where health data controllers use the data provided by users against them (for example, to harass, manipulate or embarrass them). Harms caused by medical advice by health data controllers, I argue, should be classified as medical malpractice, similar to how law treats harmful or bad medical advice by physicians. In the following paragraphs I provide the arguments for why such a distinction should be made and in particular, why the distinction is important for the future of digital health.

With the use of big data and machine learning algorithms, digital health apps not only collect and monitor health data, but also offer personalized advice to the users (Higgins, 2016). This phenomena of impending reliance upon machine learning algorithms for health advice (as well as diagnosis and treatments) is referred to as "black-box" medicine (Ford & Price, 2016). A key feature of black-box medicine is its opacity, as the amount of data involved and the complexity of algorithms, make it hard for humans to know exactly how the algorithms work (Ford & Price, 2016).

The algorithms involved in black-box medicine rely upon using machine learning techniques to find underlying patterns in a large quantity of data. The large datasets required for accurate algorithms, however, will take time to assemble, and in the early stages of black-box medicine, as we stand now, these algorithms maybe prone to errors (Price, 2017a). These errors demand a careful set of regulations and legal instruments to protect the users, and this has attracted the attention of regulatory bodies in the

developed world, such as the Food and Drug Administration (FDA) in the United States (Price, 2017b).

Regulating black-box medicine, however, can be quite challenging and there are risks involved in both, under-regulation and overregulation (Price, 2017b)[10]. While under-regulation runs the risk of leaving the users exposed and vulnerable to medical harms, the risks of over-regulation come in the form of cost to innovation (Price, 2017b). Requiring strict criteria for verification of black box algorithms may significantly increase the hurdles to get such products to the market, and thus, forestall the possibility of algorithmic medicine to improve the health care system. Further, verification of algorithms used in black-box is difficult in most cases, and even impossible in some (Ford & Price, 2016)[11].

While fiduciary law could be used to force health data controllers to take steps to design error free algorithms, such a move may not only disincentivise investment into digital health technologies, it may also be impractical. The risk of being found guilty of a fiduciary breach may force companies to abandon algorithmic medicine, as guaranteeing an error free algorithm may not be possible. Further, there is also a risk that users may claim a fiduciary breach (on account of a health data controller not being loyal) even when there is no harm or when the degree of harm is too small. The argument here is not that health data controllers should not be held accountable for the algorithms they develop and use, rather that the harms caused by those algorithms, in medical context, should be treated similar to medical malpractice and resolved through other legal instruments. Price (2017a), for example, argues that laws such as medical liability litigation can and should be used for accountability of algorithmic medicine.

---

10 For a more detailed overview of current approaches to regulation of algorithmic medicine, see (Price, 2017b)

11 (Ford & Price, 2016) suggest two main ways for verifying algorithms used in Black-box medicine: clinical trials and computational verification. Both methods come with enormous practical challenges. In the case of computation verification, most regulating bodies aren't equipped with expertise to carry out such a verification. While independent third parties might compensate for the lack of expertise, it would require significant compensation for third parties to offer their expertise. Further, for a comprehensive verification, third parties would require a broad access to data used to develop algorithms, which may open further concerns about privacy of the users. Clinical trials, on the other hand, are slow and expensive, and in most cases would only offer a small benefit. See (Ford & Price, 2016) for a more detailed overview.

Again, the proposal to make a distinction between fiduciary harms and medical malpractice is not unique to algorithmic medicine. The said distinction is also applicable, under current law, for physicians (Mehlman, 2015). Courts make a distinction between medical malpractice and fiduciary harms caused to the patient. For injuries caused by sub-standard care (including wrong or bad medical advice), as well as to deter unreasonable or unprofessional behavior by physicians, medical liability law is applied, with physicians being tried for medical malpractice (Mehlman, 2015; Price, 2017a). In contrast, fiduciary law is usually reserved for protection of patient confidentiality and for rare cases of physicians' acting purely out of self-interest (Drozd & Dale, 2006; Mehlman, 2015).

As I have discussed in this chapter, fiduciary law is abstract and open-ended. Applying fiduciary law to regulate algorithmic medicine would be detrimental to the progress of and innovation within the field of algorithmic medicine, which at least in theory, and with other instruments of regulation, can have significant positive effects on the state of healthcare. The scope of fiduciary duties for data health controllers defined here attempts to find a balance between protecting individual interests, by addressing privacy concerns, and collective interests of getting valuable insights about human health as well as facilitation of research and innovation required for gathering such insights.

Finally, it should be noted that although through this section I have tried to specify the bounds of fiduciary duty of loyalty, the courts would have an important role in contextual interpretations of these bounds, and in deciding whether a fiduciary breach has taken place or not. This is not a limitation, but rather an important aspect of fiduciary law, which can push the fiduciary to go beyond what can be defined by contractual law in protecting the interests of the beneficiary. As discussed in an earlier example, fiduciary duties are better suited than statutory or contractual obligations to ensure that health data controllers take appropriate data security measures to protect user data from hackers. At the same time, data breaches may happen due to vulnerabilities beyond the control of health data controllers[12], leaving it upon courts

[12] For example, cyber-attacks may exploit what are known as "Zero-day" vulnerabilities which haven't been discovered yet, even by the vendors of the software (with such vulnerabilities) (Kumar, 2014). Similarly, cyber-attacks maybe carried out through non-technical means, such as by gaining physical access to network systems through use of force or physical attacks (Byres et al., 2004).

to decide whether, for a particular case, the security breach also amounts to a fiduciary breach or not.

4. Gaps in and limits of fiduciary law

While the application of information fiduciary status to health data controllers will address user concerns about privacy when sharing health data with digital health data controllers, there are other problems that remain unsolved with this proposal, and would need to be addressed by other methods. For instance, there is a threat that creation of health data repositories by private entities may lead to "commercialization of science", and dilution of principles of scientific integrity as research moves from universities to private companies (Sharon, 2016). There is also no guarantee that markets will lead to sharing of this data with third parties that can advance the state of healthcare for the society as a whole. There is a possibility, for example, that owing to economic inequalities, the use of such devices, and hence, collection of data, may be limited to an economically privileged section of society, which may further escalate inequalities in health care delivery as well as create population biases when, and if, such data is used for purposes such as drug discovery or disease diagnosis (Sharon, 2016).

One challenge for the proposal to give information fiduciary status to health data controllers is the diverse nature of legal systems and regulations across the globe. In an inter-connected digital world, where the data can move easily across borders, this is a challenge for most data regulation policies (Bu-Pasha, 2017). The unique challenge for the proposal to have fiduciary relationships between health data controllers and data-subjects, however, is that fiduciary law is explicitly defined in only a few legal systems, in particular in systems of common law tradition (for example, in legal systems of countries such as USA, Australia, England, Canada) (Gelter & Helleringer, 2018). By contrast, civil law jurisdictions, such as in countries in continental Europe, fiduciary duties or relationships are not explicitly defined (Gelter & Helleringer, 2018). Yet, as Gelter & Helleringer (2018) have argued, there are implicit fiduciary principles within civil law systems and in some domains, civil law jurisdictions have even added fiduciary equivalents to existing law. The aim of this chapter has to been to argue for fiduciary principles-- in particular, the duty of the fiduciary to keep the interests of beneficiaries at the forefront through deliberation and diligence—to deal with privacy issues concerning health data. Since there are

provisions within civil law systems that are principally similar to fiduciary law, an absence of explicit fiduciary law would not be a major constraint in adopting the principles argued for in the chapter. Yet, future work in legal scholarship is needed to sketch out the details of this proposal, bearing in mind the challenge brought forth by the movement of data across legal regimes.

In section 3.1, I also pointed out the challenge in defining health data, and therefore, digital health data controllers. I argued that fiduciary duties should be imposed only on digital health data controllers who a.) Process data with the intention of using the data to determine the health status of a specific person, or b.) Process raw data in situations where it will be reasonable for a data subject to conclude that the data is being collected to determine their health status. These criteria are important to reach a balance where the scope of data controllers with fiduciary obligations is not too expansive, while sensitive health data is still protected. While the criteria I propose may achieve this balance principally, there is more work required to define these criteria in a legally pragmatic way.

Further, there are permissible latitudes within the fiduciary law which leave open the possibility of exploitation. For example, physicians can breach patient confidentiality to protect public health (Mehlman, 2015). There are also exceptions to lawyer's fiduciary duty to their clients, and the attorney-client privilege which protects client's information. Governments, for example, may use such latitudes and exceptions within fiduciary law, by forcing health data controllers to share information with them on grounds of public safety. Governments have made similar claims in the past, mandating access to digital data, which has led to mass surveillance on the grounds of public safety (Abelson et al., 2015).

Finally, while my proposal for assigning fiduciary duties to health data controllers has focused primarily on privacy, there are other dimensions of what may be in the interest of the users. While some of them may be covered by the no harm (harassment, exploitation, manipulation) condition in my proposal, a more robust and deeper understanding of the nature of such harms and of the ways in which exposure to such harms by particular actions of data controllers manifest themselves is needed. Such understanding may allow us to refine the scope of fiduciary duties so as to cater to user interests beyond privacy while avoiding the perils of an expansively defined scope of fiduciary duties. Alternatively, a deeper understanding of user interests' and how they

are influenced by the actions of data controllers may allow us to design other legal and social institutions that may complement a fiduciary type regulation proposed here.

So, it should be emphasized that this chapter is not an argument to abandon the ambitions for instruments such as transparency and accountability when dealing with health data controllers, as they can help in enabling increase the literacy of citizens about issues related to privacy, as well as issues beyond privacy, enabling a more democratic governance of digital tools and healthcare system as a whole. The limits of and gaps in the fiduciary law I have pointed out above are a testimony to the fact that my proposed solution can only take us so far. It also reminds us that we, as members of society, must continue to ask questions about our rights to fair treatment and the ethical conduct owed to us by those involved in the collection, use, analysis, distribution, and sale of our personal data.

5. Conclusion

Digital health technologies have the potential to transform healthcare by helping individuals live healthier lives as well as by providing valuable insights about our health as a collective. This potential revolution, enabled by the collection, use, and analysis of large amounts of health data, however, requires collective participation and poses threats to the individuals, exposing intimate information to privacy related harms. In this chapter, I have argued that transparency mechanisms do not adequately address individual privacy concerns, and thus, do not enable the trust required for collective participation.

To ensure protection of privacy of users sharing health data, I have argued that the relationship between users sharing health data and digital health data controllers should be recognized as a fiduciary relationship, such that health data controllers have the responsibility to keep the interests of the users at the forefront. The relationship between health data controllers and users shares characteristics with traditional fiduciary relationships and involves similar circumstances as those under which traditional fiduciary relationships are recognized. A fiduciary relationship between health data controllers and users is also better suited than alternative approaches for protecting user privacy and thus, enabling users to trust data controllers with their health data.

Chapter 2.
Googlization of Health Research and Epistemic Trust

Introduction

In August 2014, Alphabet (formerly Google) announced that its research arm would be undertaking the "Baseline study", a research project to map the human body and "create the fullest picture of what a healthy human being should be" (Barr, 2014). The study, eventually launched to the public in 2017 by Verily Life Sciences (Alphabet's health subsidiary), aimed to collect a large amount of phenotypic health data from 10,000 participants over the course of multiple years. Besides traditional health data, such as historical, clinical, and laboratory-generated data recorded through EHR (Electronic Health Records), the participants would also contribute to a multidimensional longitudinal health data set recorded through the use of wearable devices and other sensor-based technologies (Arges et al., 2020). The project aims to combine this phenotypic data with "population-based aggregate and environmental data such as local and national census data, socioeconomic data, and Centers for Medicare & Medicaid Services (CMS) data" (Arges et al., 2020). In this project, Verily is working with researchers at Duke and Stanford University to figure out what the study should measure as well as analyze the collected data. Sam Gambhir, one of the researchers at Stanford who helped design the project, was quoted about this partnership in 2017, saying that even if they (researchers at Stanford) had a large amount of federal funding, "we'd still have to find someone like a Verily or Alphabet to work with because of the large data structure needs and interactivity between participants and the internet (Rogers, 2017)". Verily is also said to contribute to the project by providing and designing data collection tools such as smart watches as well as its expertise in data analysis (Rogers, 2017). This increasingly important role played by large tech companies such as Verily (or Alphabet) is emblematic of a model of research termed by Tamar Sharon (2016) as "Googlization of health research" (GHR).

GHR, according to Sharon, is characterized by a promise to advance health research through collection of a large variety of heterogeneous data, such as through consumer-

oriented tracking devices, as well as offering technological capabilities to effectively manage and analyze this complex data. Sharon argues that GHR is made possible through a framing under which "health and medicine are framed as problems of effective management of complex data" leading to "experts in data management inevitably becoming experts in health research" (Sharon, 2019). It is not surprising that then besides Alphabet, other major consumer technology companies – Amazon, IBM, Apple, and Facebook – have all made moves into the health sector. Apple's Researchkit platform, for example, allows researchers to collect heterogeneous data through smart phone and watch sensors, and is being used to conduct several major multi-participant studies (Jardine et al., 2015).

The increasing influence of major tech firms over public health research and services is also exemplified by its partnerships with public health bodies and institutions both before and during the current COVID-19 pandemic crisis. The National Health Service (NHS) in the UK, for example, already had partnerships and data-sharing agreements with DeepMind (Alphabet subsidiary) and Amazon. During the current pandemic, NHS has established more such partnerships and agreements with firms including Microsoft, Alphabet, and Palantir (Crouch, 2020). Palantir, for example, is contracted by NHS to help in coordinating the response to the COVID crisis by tracking and monitoring several key indicators such as hospital admissions and use of beds and equipment (Crouch, 2020). Similarly, Apple and Google are among the leading firms to have launched contact tracing solutions that rely on Bluetooth signals to detect whether an individual has come in contact with other infected individuals (Kelion, 2020).

Yet, despite the promise of GHR, there is skepticism about how useful it might be in producing reliable results. In the case of contact tracing, for example, several issues have been raised. One of them is that Bluetooth-based contact tracing may lead to a high number of false-positives by counting epidemiologically insignificant "contacts" (Lee, 2020). Other problems include low-uptake of such apps, lack of appropriate and necessary technical requirements for successful contact tracing in most prevalent smart phones, and a possibly high prevalence of these two problems among the vulnerable populations such as the elderly and those from lower socio-economic groups (Kelion, 2020; O'Neil, 2020).

Similar epistemological problems of reliability have been raised in use of medical AI in health research, a space where Verily and Deepmind, both Alphabet subsidiaries, are heavily involved. Ii & Nicholson (2019) have, for example, argued that current practices around the development of medical AI suffer from a contextual bias springing from its data resourcing practices which involve resourcing data almost exclusively from high resource contexts (where this refers to places with better medical equipment, better technologies to collect and curate data, availability of experts to administer trials, provide care, and other practices relevant to data collection and evaluation of success). This was exemplified in a recent report stating that Google's AI to predict diabetic retinopathy in Thailand, while very accurate in lab settings, frequently failed to give results when deployed to test patients in real-world settings (MIT Technology Review, 2020).

Despite these misgivings, trends such as partnerships with public institutions suggest that GHR is on the rise and will play an important role in health research ecosystem. It is also not a given that epistemological problems associated with GHR, such as the problem of contextual bias alluded to above, may not be addressable and a number of proposals to address those problems have already been suggested (Burlina et al., 2020; Parikh et al., 2019). Yet, as I will argue in this chapter, the range of epistemically important consequences of GHR is not limited to direct epistemological problems such as contextual biases, low quality of data, and its opacity. Rather, social and moral values also play an important role in determining the epistemic dimensions of GHR, such as its trustworthiness for the public. The implication of this argument is that big tech corporations, such as Alphabet and Amazon, associated with health datafication not only have epistemic responsibilities but also moral and social responsibilities towards their users in order to warrant epistemic trust in epistemic goods produced by GHR.

To elucidate my argument in this chapter, I will be relying on the notion of epistemic trust. Epistemic trust can be defined as being disposed to believing a proposition *p* on someone's claim that *p* on the assumption that she is in a position to know whether p and will express her belief truthfully (Barrotta & Gronda, 2020; Fuerstein, 2013). Epistemic public trust in science (that is, epistemic trust in scientific claims by the public) is important for both common as well as individual good. As exemplified by the recent corona crises, epistemic public trust is important for citizens to abide by government decisions, such as strict lockdowns, based on scientific claims, as well as

to decide their individual health and hygiene practices. In the context of GHR, epistemic public trust is important not just for determining whether the public accepts actionable claims produced through GHR but also for future participation in practices necessary for GHR (such as data sharing).

The chapter argues that GHR threatens public epistemic trust as it is invariably associated with epistemic inequality which weakens the role of direct epistemic values in warranting trust and, in its current form, is associated with a number of social, moral, and institutional factors that do not supply adequate grounds for public epistemic trust (or such factors provide grounds against epistemic trust). This argument also entails that even if the standard epistemological problems associated with GHR, such as issues of opacity and contextual bias, were addressed, this would not remove the public's reliance on social, moral, and institutional factors for epistemic trust. In this sense, it is necessary to address problems associated with such social, moral, and institutional factors in order to give the public adequate grounds for epistemic trust in epistemic goods produced through the use of resources that currently enable GHR.

The chapter proceeds as follows: In section 1, I lay down the argument that epistemic public trust in science is grounded not primarily through epistemic factors but through a range of social, moral, and institutional indicators about a given scientist who is the target of such epistemic trust. In section 2, I elucidate the nature of epistemic inequality associated with GHR. In Section 3, I expand on the currently available social, moral and institutional indicators of GHR and its associated practices, and how they may lead to loss of epistemic trust or even supply reasons for epistemic distrust among the public.

1. Epistemic Public Trust in Science: The reliance on social, moral and institutional factors

Epistemic Trust can be defined as being disposed to believing a proposition *p* on someone's claim that *p* on the assumption that she is in a position to know whether *p* and will express her belief truthfully (Barrotta & Gronda, 2020; Irzik & Kurtulmus, 2019). Epistemic trust is particularly important in clear-cut cases of epistemic inequality, that is, cases where one has good reasons to defer to the testimony of someone else who may be a reliable and more competent source of information on the given subject. Epistemic inequality may come in at least two forms: as resulting

from a difference in competence and as resulting from a difference in accessibility (Barrotta & Gronda, 2020). So, for example, when A trusts B's claim that *p*, it may be either because A herself cannot and does not have access to reasons for *p*, or because B is more competent than A about *p* (and A knows that B is more competent about *p*).

The relationship between lay-people and scientific experts is an important instance of such epistemic inequality. This epistemic inequality with respect to scientific knowledge has increased over the course of history, as scientific knowledge increasingly relies on specialized tools (both cognitive tools, such as mathematical models, and technical tools such as electronic sensors) as well as specialized knowledge possessed by only a few (Hendriks et al., 2016). The need for epistemic public trust in science, that is for the public to have epistemic trust in science, arises because of this inequality as members of the public may either lack access to reasons justifying a scientific claim, or they may not be as competent as a given scientist to assess the said claim, or both. In such a scenario, where the public does not have access to direct epistemic reasons to accept a given scientific claim p, a number of authors suggest that what the public can and should instead do is rely on a variety of social, moral and institutional factors that indicate the trustworthiness of the scientist(s) (Barrotta & Gronda, 2020; Fuerstein, 2013; Irzik & Kurtulmus, 2019; Origgi, 2012; K. Rolin, 2015).

Barrotta & Gronda (2020), for example, write:

> "we do not simply hold that, in an expert/layperson scenario, deontological and institutional considerations are capable of influencing the justification that a subject has for believing that p; more radically, we also argue that deontological and institutional considerations are the only factors that influence the epistemic justification a layperson can advance for p."

Here, Barrotta & Gronda (2020) do not specify what deontological factors specifically entail other than that these refer to rules a *good* scientist follows. I take it that they are referring to factors that do not necessarily relate to the epistemic qualities (such as truth-conduciveness) of their output but to implicit and explicit ethical values of the scientists that play an important role in scientific research.

The importance of such social and ethical values in scientific research has been argued for by many authors (de Melo-Martín, 2019; Douglas, 2017; K. Rolin, 2015).

De Melo-Martín (2019) and Douglas (2000, 2017), for example, have argued that the role of ethical and social values is particularly important as science has not just epistemic but also social aims. Non-epistemic values can play an important role in deciding which aims are pursued (agenda-setting) as well as how science is communicated. The importance of the latter (role of values in science communication) is particularly salient in cases of inductive risk: cases where a scientist has some evidence for a high stakes claim (for example, some claims related to COVID19, which not only have important consequences, but may also demand urgency), and then must decide whether the evidence is sufficient to communicate such a claim to the public. The argument of inductive risk is essentially that in accepting a hypothesis, a scientist has to decide whether the evidence warrants the acceptance where the warrant of such acceptance depends not only of the evidence but also risks related to errors, and a moral judgement on acceptable levels of the risk of such an error in a given social context (K. Rolin, 2015). In this way, moral and social factors play an important role both in pursuit of science but also in judgements about trustworthiness of a scientist who needs to assess the moral and social contexts of their research.

Wilholt (2013) also frames a similar argument for the importance of moral values in scientific practice, and claims that moral values are not only essential for science communication but also play a significant part in methodological decisions made during scientific inquiry. For Wilholt, epistemic trust is more than mere reliance and involves trust in the testifier's (or the scientists making a claim) ability to understand her moral responsibility for inductive risks and making sound value judgements concerning such risks. For Wilholt, this implies that epistemic trust, which is trust in the idea that 'scientific endeavors are appropriately geared towards truth' involves trust in the inquirer having the "right attitude towards possible consequences of her epistemic work". In this sense, Wilholt argues that epistemic trust has an important dimension of being, at least, partly, "trust in the moral sense". Given the importance of ethical and social values in both methodological choices as well as science communication, it is reasonable to rely on indicators about moral commitments of the scientists.

As an example of the importance of social and moral values in techno-scientific practice, consider again the case of digital contact tracing. As mentioned earlier, one of the challenges in designing a digital contact tracing solution is deciding what counts as a "contact" (Lee, 2020). Among other things, what counts as a "contact" would

have to depend on the distance between the individuals in proximity to each other. Deciding this distance gives rise to an inductive risk problem: If the distance criterion is too low (say 1.5 meters), then there is a risk of false negatives (i.e. significant contacts not registered as one), and if it is too high (say 10-30 meters), then there is a risk of false positives. The decision (of defining a "contact") then depends on a trade-off with significant social and moral consequences. For example, consider a scenario where such an app dictates whether, and for how long, one should self-quarantine if one comes in contact with a positive patient. The trade-off between false positives and false negatives would have important, but also significantly different, consequences for different individuals, depending on their social circumstances. For example, one can imagine that a high-false positive rate may jeopardize fairness as it may be more costly for individuals who are in professions that require relatively high frequency of human contact. Many individuals working in such a profession, say hospitality or retail, are also likely to be relatively economically vulnerable, which further exacerbates the cost of self-quarantine based on such a digital tracing solution. In this scenario, thus, what is counted as a contact has important bearings on values such as fairness, and tracking such a value requires attunement to social contexts.

Besides social and moral indicators, members of the public can also rely on institutional factors to ground their trust in scientific claims. These institutional factors can be conceived of as reputational factors such as opinion of other experts on the expert making the claim p or on the claim p itself and institutions the expert claiming p is associated with. Irzik & Kurtulmus (2019) offer a similar account of epistemic trust in terms of highlighting the importance of such reputational factors, arguing for how laypeople rely on institutional factors to ground their epistemic trust in science. They present the example of the MMR vaccine controversy, where Andrew Wakefield, a gastroenterologist at the Royal Free Hospital, based on a study of 12 patients, published a paper in 1998 claiming that the MMR vaccine caused inflammatory bowel syndrome and subsequently led to autism. According to (Irzik & Kurtulmus, 2019), the public initially had grounds for epistemic trust in this claim, given Wakefield's and his co-authors' credentials. However, over time, as the study went through reviews by several other experts and was discredited, the public then had the grounds to reject the claims. They argue that the reason for the controversy lingering was a failure of media coverage to give reliable cues about such institutional factors. The media covered Wakefield's claims and their rebuttals equally (they argue that media should have done a better job in presenting facts about institutional factors that

the public could assess), which may have sent mixed signals to the public. They further claim that subsequent developments, such as the revelation of undisclosed conflicts of interest on Wakefield's part, withdrawal of support from his co-authors, and loss of his medical registration should have led the public to withdraw their epistemic trust from Wakefield's claims.

To sum up the argument offered in this section, given the epistemic inequality between members of the public and scientific individuals and institutions, implying a lack of direct epistemic reasons for the public to accept scientific claims, the necessary epistemic public trust in science needs to be grounded through social, moral, and institutional factors indicating the trustworthiness of a given scientific individual or institution. To be clear, the argument here is that relying on such social, moral and institutional factors for epistemic trust is warranted, although there is also some empirical evidence supporting the claim that people actually do rely on such factors for epistemic trust (NW et al., 2019; Rutjens et al., 2017)[13]. An important implication of this argument is that strategies such as increasing the scientific literacy of the general public may only have a limited impact on epistemic public trust in science as the latter is at best, only weakly associated with epistemic factors. Instead, the reasons for epistemic public trust or distrust lie in moral, social, and institutional factors, through which public is able to, directly or through reliable proxies, assess whether the given scientist has the right moral character in the view of the members of the public, is attuned to societal interests, and is competent in their work.

2. The nature of epistemic inequality associated with GHR

Before I make the case for how GHR threatens public epistemic trust, it is important to understand the nature of epistemic inequality associated with GHR. As mentioned earlier, the relationship between laypeople and scientists is one of epistemic inequality as laypeople may either lack access to reasons justifying a scientific claim, or they may not be as competent as a given scientist about the said claim, or both. This is, as expected, also the case for relationship between members of the public and GHR-related institutions and individuals. In the case of GHR, however, there is another

[13] Rutjens et al., (2017) provide evidence for how moral beliefs and concerns account for science acceptance and rejection. The Pew survey conducted within the American public (NW et al., 2019) provides evidence for a public preference and demand for transparency of data justifying claims and independent review of such claims.

form of epistemic inequality that may be salient – between institutions associated with GHR and other scientific institutions. As Sharon (2016) explains, GHR is characterized by a disproportionate access to better financial, human, and technical resources, including datasets, for health research. This difference in accessibility between GHR and other scientific institutions may then give rise to an epistemic inequality. This inequality is also exemplified in the statement by Sam Gambhir, a Stanford researcher associated with Project Baseline quoted earlier in the introduction to this chapter, about the necessity of collaborating with a company like Verily for such a project given Verily's relative dominance in assembling and analyzing a dataset as large as one required for a project like Baseline.

This epistemic inequality between GHR and other scientific institutions is exacerbated by the opacity associated with GHR. As Burrell (2016) explains, there are three kinds of opacity associated with algorithmic research. The first kind of opacity stems from the inscrutability of methodology involved in machine learning algorithms (such as deep neural networks). The second kind of opacity stems from intentional corporate secrecy, such as in use of proprietary algorithms. The third kind of opacity involves technical illiteracy and the specialized nature of writing and understanding algorithms. While the latter may not be salient in the relationship between GHR and other scientific institutions, the other two are.

The epistemic inequality between GHR and other scientific institutions may be salient for our case, that is for public epistemic trust in GHR, for at least two reasons: first, it implies that other scientific institutions may also be unable to trust results produced by GHR just on their content and may also have to rely on other factors (such as social and moral indicators) not related to the content of the results in question. Second, if other scientific institutions are not able to engage with the results produced by GHR, it reduces the sphere of institutional (or reputational) factors available to the public for expanding or withdrawing epistemic trust in GHR.

3. GHR and Loss of Epistemic Trust

In an editorial on use of AI in healthcare, Wynants et al., (2020) write that despite many claims about AI improving "screening, diagnosis, and prognostication", external validation studies and randomized controlled trials evaluating such claims are scarce. Further, from the few randomized trials and studies that have been

published, results are mixed, even indicating in some cases that AI and data-driven research led to less (rather than expectedly more) accurate results compared to senior consultants. A recent report stating that Google's AI to predict diabetic retinopathy in Thailand, while very accurate in lab settings, frequently failed to give results when deployed to test patients in real-world settings perhaps adds to the observation made in the editorial (MIT Technology Review, 2020). Examples such as these have added to the recent critiques of data-driven and algorithmic research, pointing out the epistemological problems associated with them.

As mentioned earlier, two examples of such epistemological problems include the problem of contextual bias and opacity. However, the problems of opacity and data bias may not have a direct role to play in public epistemic trust in GHR because of the nature of epistemic inequality between lay people and GHR, which as explained earlier also stems from a difference in competence in accessing and assessing claims by GHR. Opacity, for example, is not unique to GHR and is the reason for any kind of epistemic inequality between laypeople and experts. In this sense, even if machine learning algorithms were more explainable and not protected by copyright laws, they would still be opaque for the general public, just like the justification for other scientific claims often is.

We may, however, consider the argument that opacity and bias play an important indirect role in the degree of public epistemic trust in GHR due to their effects on other scientific institutions and their approach to claims produced through GHR. While bias may give a general reason for skepticism to other scientific institutions regarding GHR, opacity in the form of algorithmic inscrutability or secrecy, may introduce a significant epistemic inequality (by denying access to assess claims directly) and reduce or remove reasons for epistemic trust in GHR among other scientific institutions.

Yet, the degree to which this reduced epistemic trust in GHR, among other scientists due to opacity and bias, plays a role needs to be examined further by comparison with other cases of intra-scientific epistemic trust. Through her empirical study of scientific research groups, Wagenknecht (2015) argues that in coming to accept and endorse claims by other scientists, epistemic trust among the scientists plays an incomplete role, and scientists employ other strategies, including relying on institutional (or reputational) factors as well as cues about moral values. Wagenknecht claims that

epistemic trust among scientists is based on empirical warrants, which may only be limited and even for partial achievement require a slow process involving multiple engagements and collaborations. Further, since epistemic trust is future-oriented, it is underdetermined by such warrants or evidence. Wagenknecht claims that scientists attempt to bypass and supplement this incomplete personal epistemic trust through impersonal trust mechanisms such as relying on institutional and reputational factors (including qualitative factors such as the institution they are part of, quality of venues they present/publish research in) as well as cues about moral commitments of other scientists through dialoguing (such as asking for explanations or assessing explanations given to other colleagues) and collaborative practices. Rolin (2014) extends Wagenknecht's argument for incompleteness of epistemic trust and argues that there is also a moral dimension of trust involved here. Rolin and Hardwig (1991) both argue that epistemic trust is not just trust in the competence of the testifier but also their moral character, especially as one has to trust in the honesty of the testifier.

If we accept these arguments, then it seems that when laypeople rely on institutional factors for epistemic trust such as opinion of other experts about a given claim, they also implicitly rely on expert judgements about similar social, moral and institutional factors in play, even though the evaluation of such factors in a given case by other experts may not be same or similar to evaluation by the public (for example, the public may not be able to assess qualitative factors such as quality of a journal a particular claim is published in).

In this sense then, while problems relating to contextual bias and opacity have an important role in the assessment of data-driven health research (and epistemic trust in it), there are also other unique factors that may lead to loss of epistemic trust in GHR which relate to the social, moral, and institutional indicators about GHR available to the public. This rational or warranted reliance on social, moral, and institutional factors is bad news for epistemic trust in GHR, at least in its current form, as practices associated with GHR raise several red flags related to such factors such that the publics may have no or fewer grounds for epistemic trust in GHR. In what follows, I discuss some such factors, which may have a negative effect on epistemic trust in GHR either because they reduce the access of the public for positive reasons to ground epistemic trust or in some cases, give reasons against epistemic trust in GHR.

➢ *Secrecy and lack of evidence of previous domain-specific expertise*

In 2016, the NHS signed a contract with DeepMind whereby the Royal Free London, an NHS Foundation Trust, granting DeepMind access to identifiable information on 1.6 million of its patients in order to develop an app to aid medical professionals identify patients who were at risk of acute kidney injury (AKI). This transfer of data was later ruled out to a breach of a data protection following an investigation by the Information Commissioner's Office. In particular, the data sharing agreement was based on a lack of consent as patients were not at all aware of the agreement as well as how their data was being used (BBC, 2017). Besides this lack of public consultation, Powles & Hodson (2017) point out that DeepMind, which is an artificial intelligence company without prior experience in healthcare, was passed on sensitive health data of millions of patients on the premise that DeepMind would develop a smartphone app. According to claims by DeepMind, its involvement was limited to developing the app, and it would not artificial intelligence or machine learning techniques, which is the domain expertise DeepMind did possess (Powles & Hodson, 2017).

This agreement between DeepMind and NHS was clouded in secrecy. Not only the agreement took place without public consultation, the details of the agreement were only revealed after an independent journalistic investigation (Powles & Hodson, 2017). This was also the case with an agreement between the NHS and Amazon where the NHS agreed to give Amazon access to its data so that Amazon Alexa could offer users expert advice with the reason that this would reduce the burden on NHS (Guardian, 2019). Several important details of the contract were not released to "protect Amazon's commercial interests" (Guardian, 2019). The lack of transparency also came to the radar in a more recent case involving NHS' data sharing agreements with big tech firms such as Google, Amazon, Microsoft and Palantir to help coordinate its response to COVID after the UK government only released the details of this agreement following MPs and more than 13,000 people urging the government to do so (Lovell, 2020).

While deliberate secrecy may itself warrant loss of trust on moral grounds, there are also worrisome "institutional" factors in play here. In particular, the lack of previous domain experience and secrecy create problems of epistemic trust in GHR within the scientific community. As argued, the public also implicitly relies on such intra-scientific trust. Rolin (2002) and Wagenknecht (2015) have argued that intra-scientific trust develops and is maintained through enduring communities with shared norms

and dialoguing practices. GHR's lack of previous domain experience and secrecy of its projects are both an obstacle to its inclusion as part of such a scientific community with shared norms and intra-community dialogues.

An instance of this problem is also highlighted in Leonelli's (2016) work on "data-centric biology". Leonelli notes that within the practice of biology, the work of biocuration, that is, of collecting, annotating and validating biological databases, is, increasingly, being done by professionals outside the biological community. This has led to mistrust of biocurators by biologists, particularly as they do not find that the biocurators are adequately addressing the needs and concerns of scientists within the biological community (Gabrielsen, 2020).

- *Possibilities of and possible abuses*

(Taylor, 2021) points out that use of public health data by corporate actors leaves open the possibilities for misuse and raises several questions about legitimacy and therefore, corresponding redressals. She writes (Taylor, 2021):

> "With regard to public authorities' own data use, the GDPR is specific (in Art. 6(1)(f)) that they must locate the basis of their 'legitimate use' in national law, however it does not demand this of corporations. This is because the responsibility for making sure corporations are operating within the law lies with the state... However this involves no positive obligations, so that if a firm starts taking on the tasks of public authorities, questions arise about the type of legitimacy involved. For example, if Amazon starts to intervene in public healthcare provision or in the insurance market based on its access to public-sector data, it is unclear how people should weigh the legitimacy of those interventions – on the same basis as government, in which case where is the law that allows it to shape public health? Or on the much weaker basis of its business interests, which do not seem sufficient to bound this scale of power?"

The above mentioned agreement between the NHS and Amazon may be an example as under the agreement, while Amazon is able to access this copyrighted NHS data as well as share it with third parties, NHS would not receive compensation or allow Amazon to use it globally (as opposed to limited use to the UK, which is the case for third parties traditionally) (Guardian, 2019).

The availability of bad incentives is pointed out by Vogt et al., (2019) who, writing about the increased risk of overdiagnosis with big data and precision medicine, state that the risk could be amplified as big firms such as Google, Amazon provide consumer-oriented diagnostic devices, especially as more consumers may be

interested in devices with increased sensitivity which may amplify the risk of false positives. There is also already some evidence that companies exploit incentives offered by consumer interest and several commercially available devices have been leading to mis- or overdiagnosis (Baron et al., 2017; Digital Trends, 2019; Owens & Cribb, 2019). The possibilities of misuse and availability of bad incentives, at the least, increase the need for assessment of social and moral values embedded within the practice of GHR, and at worst, creates grounds for distrust particularly, as and when there are actual examples of misuse, such as in the early reports of overdiagnosis through digital health devices.

- *Concerns about power asymmetry*

In her paper coining the term GHR, Sharon (2016) argues that one of the main concerns with GHR is that it introduces power asymmetries between traditional, more trusted institutions engaged in health research and big tech firms which could lead to firms dictating allocation of resources to new research areas and thus play a dominant role in agenda setting. Taylor (2021), citing several data sharing agreements between big tech and public bodies, including health bodies such as NHS, argues that such agreements significantly increase the outreach and power of firms, often beyond citizen control, introducing new risks for publics. These include (Taylor, 2021):

1. Increased scale and reach of tech firms, which extends beyond government control as exemplified in multiple cases as well as removes/inhibits the option for citizens to opt out. Specific to healthcare, such agreements also often dilute or ignore the principle of consent that has been central to healthcare research traditionally. As stated earlier, this was exemplified by the agreement the NHS and DeepMind.

2. State dependence on tech firms – Taylor argues that agreements such as that between NHS and Amazon exemplify the increasing dependence of public authorities on big firms to deliver public services and goods. Coupled with the secrecy that comes along with such agreements, it also significantly reduces the ability of citizens and other institutions such as the media (journalists) to question the government on the functioning of these services.

3. Risks of manipulation and "technological shaping of citizenship" (Taylor, 2021) – the increased control and power of firms exposes citizens to new risks

of manipulation. In case of health devices, concerns about manipulation and hypernudging have already been raised (Lanzing, 2019).

4. Creation and amplification of vulnerability – Taylor argues that several features of datafication process involving big firms have led/is leading to new or amplified forms of vulnerability particularly as there are examples of how automation of public services leads to disempowerment and neglect of the vulnerable. This is particularly revealed in work, for example, by Virginia Eubanks (2018) who details cases of automated decision making exposing or amplifying existing vulnerabilities in society.

These power asymmetries may be problematic for epistemic trust even when the scientific output produced by GHR is not epistemically deficient. As argued in section 2, not only the aims of science should be governed simultaneously along social and epistemic dimensions, its public assessment is also on both those dimensions. Negative effects of such power asymmetries such as loss of autonomy, the introduction of new risks, and amplification of existing risks may decrease trust in science that contributes to or is made possible through such asymmetries. Power asymmetries also exacerbate the problem of conflicting and/or contested values that play a role in science communication and agenda-setting. De Melo-Martín (2019) argues that while bias in scientific agenda setting is not an epistemic problem per se, it can be a problem when other actors and institutions are impeded in pursuing alternative, competing agendas or in verifying and validating claims produced through a biased agenda setting process. At the very least, power asymmetries in the scientific sphere may significantly reduce the range of institutional or reputational factors available to the public to warrant epistemic trust.

- *Privacy perils and Contextual transgressions*

Research under GHR is predicated on the idea that larger the dataset, the more and better the insights one can gain from it. This obviously introduces privacy as a widespread concern. Further, as discussed in Chapter 1, advances in computation techniques mean that anonymization, previously easier to establish, is now increasingly difficult to implement implying that privacy may no longer be guaranteed through anonymization. As stated above, several data-sharing agreements between public bodies and big tech seem to leave open the room for these firms to share data with third parties, exacerbating the concerns.

A related threat here is that of contextual transgressions (Sharon, 2019). Nissenbaum (2011) has argued that privacy expectations are determined, at least partially, by contextual factors – the nature of information being transferred or exposed, the type of relationship or context in which the information transfer took place, the uses the information will or could be put to, etc.. As exemplified in the DeepMind-NHS agreement, datafication under GHR may lead to dilution/neglect of the principle of consent and allow use of data by firms for uses users may disagree with or are unaware of.

There is also evidence of increasing public risk perception concerning privacy losses, particularly related to use of smart devices by big companies (Mani & Chouk, 2019). Smart devices such as consumer health wearables exacerbate these tensions, particularly as they may lie outside a regulated health and medical domain. This implies that privacy protection offered by regulations in the traditional health sector often don't apply to data collection practices through these devices. The risk of contextual transgression is also increased as was exemplified in a court case where data from Fitbit was used as an objective measure of the plaintiff's physical activity (Gibbs, 2014).

My argument here is not that these problems are necessary companions of GHR, but rather that in its short history, they serve as important social, moral, and institutional indicators that warrant loss of epistemic public trust. It is important to note that it is not only the public that rationally relies on such factors, other scientific institutions also rely on them, particularly in the face of problems like opacity. The factors presented here may have a negative effect on warranted epistemic public trust in claims produced through GHR for different reasons. While some of them reduce the ability of laypeople to access reasons (such as positive reputational factors) that would have given grounds for warranted epistemic trust (such as reliance on other experts who can properly assess these claims), others may actually give reasons *against* epistemic trust (for example, because there are risks involved related to increase in power asymmetries or breach of social values like transparency or breach of privacy protection laws).

4. Concluding remarks

In a recent article on COVID-19 contact tracing apps, Sharon (2021) discusses the contact tracing API launched by Google and Apple in April 2020, which was lauded by many experts for its privacy-friendly technical specifications. According to Sharon (2021), while privacy is an important value to consider in debates about the digital sphere, a narrow focus on it may cloud our view of other important dimensions of the role played by big tech corporations in pandemic management. Such dimensions include the effects and consequences of dependency on big tech corporations for essential public services and increase in the influence of such corporations on public policy, for example. This chapter raises a similar concern about warranted epistemic trust in GHR. While epistemic concerns related to data-driven health research within GHR, such as contextual bias and algorithmic opacity, are of great significance, we must not lose focus on the broader social and moral dimensions of the GHR phenomenon. The overarching lesson to draw here is that if private actors, involved in GHR, are to play an important role in scientific output, it is imperative to address their functioning along the lines that can give the public warranted positive reasons for epistemic trust in epistemic goods, produced through the involvement of such private actors, in the form of social, moral, and institutional indicators associated with such a process.

Besides steps private actors should take to warrant public epistemic trust by attending to social and moral dimensions of their research, there may also be lessons here for other actors. These include actors involved in formulating and enacting strategies to address public trust in science, including science communication as well as organization, who may need to take into account that the indirect indicators of epistemic trustworthiness of GHR are more important than has been previously recognized. Paying due attention to such indirect indicators requires reflection on mechanisms that may incentivize or guide what epistemic pursuits are undertaken by private actors within the GHR (agenda-setting), to what ends, as well as the mechanisms through which other actors can access and engage with such epistemic pursuits.

While a comprehensive account of all possible mechanisms needed to facilitate appropriate epistemic public trust in epistemic goods produced through the use of resources that currently enable GHR is beyond the scope of this chapter, the discussion offered here does point to, for example, the role public health institutions

can play in facilitating some such mechanisms. This would include, for example, facilitation of a wider discussion in the public sphere regarding legitimate normative public expectations from epistemic goods that can be produced through the use of resources that currently enable GHR as well as mechanisms needed to evaluate if such normative expectations are being successfully met. Such discussions in the public sphere can, for example, help guide the nature of relationships public health institutions, such as the NHS in the UK, forge with private actors involved in the GHR. This would involve, for example, paying attention to the criteria for entering data-sharing agreements with private actors, the implications of such agreements on epistemic trustworthiness in epistemic goods produced through involvement of such private actors, and mechanisms through which such agreements are made transparent to the public and/or gain public consent. Finally, public health institutions can also help reshape interactions between corporate research actors and public research institutions through, for example, funding instruments that incentivize participation of corporate actors in existing institutional processes, such as peer reviews, and regulations that force corporations to share more data and other epistemic goods with researchers in public institutions (Hegelich, 2020). The discussion offered here echoes the call from Heidi Grasswick (2019) regarding the need for paying attention to the wider social dimensions of epistemic pursuits as well as of epistemic trustworthiness in such pursuits. This is imperative given the increasingly important role played by GHR and the need for warranted public trust in science.

Chapter 3.
Proxy Assertions and Agency: the case of machine-assertions

Introduction

"Alexa, how do I know if I have coronavirus?". If a user in the United States would ask this sort of query to Amazon's digital voice assistant, Alexa, they would be verbally given information regarding known symptoms from the US CDC's (Centre for Disease Control and Prevention) website (Porter, 2020). As digital voice assistants such as Alexa are becoming ubiquitous, such instances of machine-generated speech to provide information, particularly health-related information, are expected to rise (K. Miller, 2021; Olmstead, 2017). While the potential lack of reliability of such voice assistants presents risks, public health institutions also see voice assistants such as Alexa opening up opportunities in providing health information to millions in a fast, convenient, natural language format – as exemplified by the deal between UK's NHS (National Health Service) deal with Amazon in 2019 which would allow and encourage Britons to seek health advice through Alexa (Vincent, 2019).

Are such instances of informative machine-generated speech equivalent to human "assertions"? And what are the implications of the answer to this question for how such systems should be designed, and who should bear the responsibility for the utterances such systems produce? While the term "assertion" may feel removed from ordinary daily language, the phenomenon captured by the concept of assertion is one that is easily identifiable – the speech act of assertion, such as making claims about, reporting, stating, or affirming something to be the case, is crucial and abundant in our everyday information sharing practices (Goldberg, 2015). Phenomenologically, many contemporary machine outputs seem similar if not identical to human speech. These machine outputs also seem to serve a similar purpose – informing the listener and facilitating knowledge transfer from one source to another. This phenomenological and functional similarity, along with the increasing ubiquity of machine-generated speech, has attracted recent philosophical attention.

Several scholars have defended the view that there is no principled reason to reject the possibility of machine assertions. While some, like Bruno Latour (Latour, 2012), deny that there is even a meaningful distinction between human assertions and functionally similar machine-generated speech, others acknowledge a difference between the two but still argue that machines can assert. Freiman & Miller (2020), for example, have argued for the latter, developing the notion of "quasi-assertion": machines can be said to quasi assert when their utterances are phenomenologically similar to human speech and when such an utterance conforms to an epistemic norm that would have been applicable to human speech in a similar epistemic context. So, in the abovementioned case, according to this account, Alexa is quasi-asserting the symptoms of COVID-19 to the user with the query if its linguistic output conforms to the epistemic standard one would expect of a statement given by a CDC spokesperson in that situation.

In this chapter I argue that the above views regarding the possibility of machine assertion rely on a "functionalist" notion of assertion where success is defined in terms of the function of the linguistic output for the listener, such as gaining true beliefs or knowledge from such output, and is seen as independent from properties of the asserter such as psychological states and/or ability to undertake epistemic responsibility for the output. That is, machine utterances are deemed as assertions when they perform the same function, such as producing a true belief in the audience, as similar human utterances would.

In this chapter, I first illustrate that this *functionalist* case for machine assertion relies on a notion of assertion that deviates substantially from the notion of assertion typically invoked in epistemological work on assertion (sometimes referred to as "traditional notions of assertion" in this chapter). In the latter, successful assertions are partially defined in terms of properties of actions and/or mental states of the asserter, rather than solely on their success in producing desired epistemic states in the listener. These views, thus, dictate that asserters should be the sort of agents that can have the requisite mental states, and/or undertake ethical and/or epistemic responsibility for their assertions. After illustrating the differences between the two notions of assertion, I discuss some theoretical and pragmatic challenges faced by this functionalist notion of assertion.

Subsequently, I analyze an alternative proposal regarding machine utterances and assertions given by Nickel (2013). This proposal is compatible with traditional notions of assertions and avoids the challenges faced by the functional case for machine assertion. This proposal states that machine utterances can only be deemed as *proxy assertions*, where such utterances must meet certain contextual evaluative standards, and where the responsibility of such utterances lies with the designers of the machine. By deeming them as *proxy* assertions, this view makes the role played by human agents, such as designers of the machines, explicitly visible. I present certain possible limitations of this view and argue for a modified proposal. Under this new proposal, I contend that only those machine utterances can be deemed as proxy assertions where the designers, or a collective of actors whose work influences the utterance, can reasonably foresee and therefore, take responsibility for such utterances. I then discuss some implications of this view, both for designers and users of machines that produce phenomenologically similar output as human speech.

The structure of the chapter is as follows: In section 1, I discuss what I call the "traditional notions of assertion", that is accounts of assertion discussed in the philosophical literature on assertion. In particular, I illustrate how such accounts of assertion define assertion in terms of mental states and/or actions of an agent. In section 2, I present objections and challenges for a functionalist account of assertion that has advocated for the case of machine assertion. In section 3, I develop an account for proxy assertion, which are machine utterances that designers can take responsibility for. Finally, in section 4, I discuss some implications of my proposal.

Before I proceed further, it is important to mention the relationship between the terms "assertion" and "testimony". There is some debate in philosophy about whether the two terms are interchangeable or whether one is a subset of the other (Hinchman, 2020; Leonard, 2021). In this chapter, I will follow the lead of authors such as Freiman and Miller, Ernst Sosa, and Fricker and use the two terms interchangeably (Freiman & Miller, 2020; Leonard, 2021). Much of the literature discussed in the chapter also uses the terms interchangeably.

1. Assertion as a speech act

The central insight of the speech-act theory is that language can be a medium for a variety of actions. Assertion, as stated, is one such action, through which speakers (or

asserters) claim or affirm something to be the case. Assertion, then, is one of many possible speech acts, among others such as promising, requesting, commanding, etc. Such speech acts are also referred to as *illocutionary* acts, a term coined by J.L. Austin (1955/1975), who distinguished them from *locutionary* acts and *perlocutionary acts*. While locutionary acts are mere utterances of semantic content, perlocutionary acts are acts defined in terms of effects produced by illocutionary acts. A key aspect of Austin's theory is that locutionary acts and illocutionary acts can share syntactic and semantic properties: that is, the same sentence can be a mere utterance on one occasion and an assertion on another. This raises an interesting metaphysical question about speech acts: what gives illocutionary force for an utterance to be counted as an illocutionary act?

This metaphysical question is sometimes also asked in terms of the *essence* of a speech act (Ball, 2014). What is the essence of an assertion, for example? Further, a related question here may be put as: what differentiates or individuates an assertion from other types of speech acts? Broadly speaking, there are two views on these metaphysical questions about assertion: one that defines assertion as a *normative kind* and the other that defines it as a *descriptive kind*. In this section, my aim is to show that both these views define successful assertions in terms of the actions or the mental states of the asserters.

The more popular and influential view among the two is that assertion is an inherently normative act, although there are disagreements about what that entails. Within the proponents of the normative view, one hypothesis, which can be attributed to Searle (2000), contends that to make an assertion is to engage in a rule-governed action and that such rules are *constitutive* of the act of assertion. Asserting then, in this view, is similar to playing a game of chess in that the rules of the game of chess are also what define what game one is playing. One such rule for assertion, according to an influential hypothesis, is that an asserter asserting *p* must know that *p* (T. Williamson, 2002).

K-rule One must: assert p only if one knows p

On the constitutive view, what it means to assert then is to follow such a rule and this rule is constitutive in the sense that to violate such a rule is to be unsuccessful at asserting. A different view on the matter states that the K-rule is *prescriptive*, in that,

while asserters are *necessarily* subjected to it, it is still possible for an asserter to violate it, for example, when they make a bad or insincere assertion and assert p without knowing p[14]. Despite this disagreement, both views rely on a notion of assertion where it is *essentially* normative, in that, what it means to assert is to be *necessarily* subjected to a rule such as the K-rule.

Two things of note regarding such norm-based accounts of assertion: First, the accounts define assertion in terms of obligations and/or commitments to be undertaken by the *asserter* (or the speaker). Second, most such accounts define assertion, and explain what it is to assert, in terms of an *epistemic norm*- that is, they define assertion in terms of the epistemic responsibility asserters must undertake, for example, by having the right set of doxastic attitudes (knowing p, for example), or ensuring that the utterance qualifies certain criteria[15]. This implies that in unsuccessful or bad (depending on the view you subscribe to) assertions, the speaker is subject to *epistemic* blame. One exception to the latter trend is from the work of Cuneo (2020) who argues that besides epistemic responsibility, what explains what it is to assert is the ethical responsibility undertaken by the speaker, such that the speakers are also morally accountable to their audiences in asserting. For the purposes of my argument here, it would suffice to say that the proponents of the view that assertion is a normative kind share the view that asserters undertake certain commitments and obligations qua asserters.

In contrast to the normative view, the *descriptive kind* view of assertion contends that even though assertion might have normative features, they are not essential or necessary for assertion. In this view, assertion is defined purely in descriptive, and not normative terms: for example, in terms of the psychological states of the speakers and their reflexive intentions (García-Carpintero, 2019). Descriptive accounts of speech acts, for example by Grice (1989, pp. 88–138), explain the source of illocutionary force for illocutionary acts such as assertion to be rooted in the intentions of the speaker. A Gricean version of such an account for assertion may claim, for example, that to assert

---

[14] According to (Marsili, 2019), the proponents of the prescriptive view also sometimes claim that the rule they are proposing is "constitutive" of assertion. According to Marsili, this stems from a confusion about the usage of the term constitutive. Here, I refer to the constitutive view as one that is attributed to Searle, where such rules are akin to definitions and violating them would mean not partaking in the act defined by the rules.

[15] See (Pagin & Marsili, 2021) for a review of different proposals regarding the epistemic norm of assertion.

is to express a belief p by uttering a sentence that means p with the reflexive intention that one's audience believes p (see for example (Bach & Harnish, 1979, pp. 39–43; Ball, 2014; García-Carpintero, 2019). It should be noted that such accounts, while claiming that normativity is not essential to defining assertion, are still compatible with the claim that many assertions do have normative aspects, and for example, we may criticize performances encountered in our assertoric practices for violating something akin to the K-rule (where for example, someone claims something to be the case without knowing it) or for other breaches such as that of politeness. In the same vein, some normative accounts of assertions are also compatible with the idea that some or all assertions are accompanied by particular psychological states (belief in what is asserted, for example) and/or reflexive intentions.

My purpose in explicating these two views of assertion (a descriptive or normative kind), has been to illustrate that both views rely on the idea of successful assertions depending on the agent – on agents having the requisite psychological states and/or successfully undertaking certain commitments and obligations (epistemic and possibly ethical). Any account of machine assertion, which is compatible with these views on assertion, would then, have to explain either how machines can have the requisite mental states for asserting or how machines can be the sort of agents that undertake epistemic and possibly ethical responsibility. Given our traditional conceptions of machines, the case for machine assertion looks challenging.

This is also reflected in some of the objections against the case of machine assertion. Consider, for example, the following lines by Fricker (2015):

> "So we do not count thermometers, fuel gauges and so forth as testifying to the temperature, or to how much fuel is left in the tank. Nor do we count tree rings as testifying to the age of the tree."

Fricker's takes it as a common sense assumption that machines, much like natural entities, cannot have intentions and therefore, cannot testify. Goldberg (2012) rejects the case for machine assertions on the basis that machines are not the sort of entities one can normatively assess or have normative expectations from. This makes machine utterances radically different from human assertions, as in the latter one can normatively assess, for example, the doxastic states of the asserter.

2. Machine Assertions and functionalism

As indicated in the introduction, some recent scholars have made the case for machine assertion in, what I would call, functionalist terms. One such view comes from the French Science and Technology Studies (STS) scholar Bruno Latour who denies that there is any relevant distinction between humans and non-humans. Latour's view on the possibility of machine testimony is reflected in these lines, about laboratory objects, from a section titled "The Testimony of Nonhumans" (Latour, 2012, 23) (also quoted in (Freiman & Miller, 2020)):

> "inert bodies, incapable of will and bias but capable of showing, signing, writing, and scribbling on laboratory instruments before trustworthy witnesses. These nonhumans, lacking souls but endowed with meaning, are even more reliable than ordinary mortals, to whom will is attributed but who lack the capacity to indicate phenomena in a reliable way."

Sosa (2006), like Latour, also likens knowledge gained from instruments and from human testimony. According to Sosa, testimonial knowledge can be considered as just another form of instrumental knowledge. For Sosa, there are no distinct qualities that separate testimonial knowledge from instruments. An instrument as well as a testifier, both, possess some competencies through which they are able to deliver propositional content to a subject (whether a hearer or someone using the instrument). For this subject to gain knowledge, all that matters is that the propositional content, or the uttered expression, is delivered reliably. It is this latter view, reflected in Sosa's argument, that is the target of my arguments in this chapter. It should also be noted that by a "functionalist" view I mean here the sort of functionalism that relies on a notion of assertion where success is defined in terms of the function of the linguistic output for the listener, such as gaining a true belief or knowledge from it and is seen as independent from asserter's psychological states and/or ability to undertake epistemic responsibility for the output[16]. That is, machine utterances are considered as *assertions* when such utterances lead to similar outcomes, such as producing a true belief in the audience, as propositionally similar human utterances would.

Two other accounts, arguing for the view that machines can assert, rely on the functional success (of producing true beliefs in the audience, for example) of machine utterances. Green (Green, 2006, p. 36 and Green (2010)), for example, in comparing

[16] This is in contrast to the sort of functionalist view of assertion proposed by, for example, (Kelp, 2018) which is functionalist about the explanation for inherent normativity of assertions.

beliefs formed from testimony from humans to beliefs formed as a result of what he terms as cases of "machine testimony" claims that: "if two beliefs (a) have the same epistemic status, (b) have the same contents, (c) are the result of the exercise of the same cognitive ability by S (subject receiving the testimony), and (d) have the same phenomenology for S, then the two beliefs should be regarded by the epistemologist as similarly based; we should regard either both, or neither, as testimonially-based" [quoted from (Green, 2010)]. Here, Green offers the view that testifiers need not have phenomenal states. In this view, also, then, assertions can be defined independent of the asserter's psychological states or the ability to undertake epistemic responsibility for the asserted output.

Freiman & Miller (2020) also emphasize the functional equivalence between some machine utterances and human assertions, claiming that machine utterances such as from a loudspeaker at an airport should also be considered assertions, or in their terms, "quasi-assertions" or "quasi-testimony". They characterize quasi-testimony as follows: (Freiman & Miller, 2020, p. 13)

> "A linguistic output of an instrument or a machine constitutes a quasi-testimony in a given context of use if and only if the machine or instrument has been designed and constructed to produce this output in a manner that sufficiently resembles testimony phenomenologically, and it is in conformity with an epistemic norm that is parasitic on, or sufficiently similar to what is, or would be, an epistemic norm of testimony in the same context."

Freiman and Miller further state that their characterization of quasi-assertion "holds the output to the relevant epistemic norms" (Freiman & Miller, 2020, p. 13) and define the criteria for classifying (quasi) assertions in terms of the functional success of the output in producing the desired epistemic state in the listener. For example, to make their case for loudspeaker announcements on an airport to be deemed as assertions, Freiman and Miller (2020, p. 13) state: "the function of the message, the explanation of why subjects get knowledge from it, and the phenomenology are the same in both cases."

As discussed in the previous section, traditional accounts of assertion necessitate that asserters have the particular psychological states or the ability to undertake epistemic responsibility for the asserted output. The functionalist views discussed above deny, or at least do not propose, that machines are the sort of agents that can have such psychological states or undertake epistemic responsibility. Yet, they claim that

machines can assert because they rely on a notion of assertion that clearly deviates from the traditional views of assertion, and defines assertion through criteria that are independent of such properties of the asserter.

In addition, the functionalist view presented above also faces a challenge in the face of recent empirical evidence regarding the norm of assertion, which seems to suggest that hearers expect speakers to have reasonable beliefs about what they are asserting (or only assert what they reasonably believe) and therefore, aren't only evaluating the properties of the utterances but also the mental states or intentions of asserters (Kneer, 2021; Marsili & Wiegmann, 2021).

Besides this deviation from the traditional accounts of assertion and the empirical evidence regarding expectations from asserters, the functionalist view also faces the theoretical challenge of explaining the source of the illocutionary force for a given utterance. In other words, what makes a sentence a mere utterance in one case and an assertion in another if both share phenomenological features and semantic content? One suggestion, on behalf of the functionalist, would be to identify the source of this illocutionary force as the output's characteristics, such as its reliability, which helps produce the desired epistemic state in the listener. This suggestion, however, seems untenable. As Fricker (2015) remarks, and as discussed in the previous section, natural entities, like tree rings, can also be reliable in producing the desired epistemic states in the observing audience but we do not deem such natural entities as testifying or asserting.

Considering Freiman and Miller's proposed criterion of quasi-assertion, there seems to be another candidate to consider as the possible source of the illocutionary force that makes a machine utterance count as a quasi-assertion: output's "conformity with an epistemic norm that is parasitic on, or sufficiently similar to what is, or would be, an epistemic norm of testimony in the same context" (Freiman & Miller, 2020)[17].

[17] Here, Freiman and Miller also include phenomenological similarity to human speech as a criterion for quasi-assertion but the candidacy of the phenomenological similarity to human speech as the source of the assertoric illocutionary force seems untenable. We do not regard all human speech as assertions, nor would we regard a bird passing by and mimicking human speech to produce a statement as asserting.

There are, however, serious challenges in considering and evaluating the merits of this suggestion. First, it is not clear what sort of epistemic norm Freiman and Miller have in mind and as Pagin & Marsili (2021) note, there is much debate on what epistemic norm(s) govern assertion. Even if this debate were to be settled, and say a norm as characterized by the K-rule (asserters must: assert *p* only if they know *p*) was accepted, it is not explicitly clear to whom, in Freiman and Miller's view, such a norm should apply. Comparing the case of machine quasi-assertion to similar cases of "testimony in the same context", as Freiman and Miller suggest, also does not help make the situation clearer. For example, consider again the example of an automated airport loudspeaker announcement, which Freiman and Miller deem to be a case of quasi-assertion, and compare this to a case of a human announcer on the airport. In the human case, it is clear that the epistemic norm, say the K-rule, is to be applied on this human, thus, implying that the human announcer must only assert if she has knowledge of what is being announced. The K-rule, however, cannot be applied to the automated loudspeaker, as it is not the sort of agent that can have knowledge. If the epistemic norm in question does not apply to the asserter (in this case, a machine), then who does it apply to?

One response to this challenge might be that the epistemic norm in question should apply to not the machine, but other agents, on whom such a norm can apply and who play a necessary role in making the output conform to the said epistemic norm. In discussing possible objections against machine assertions, Freiman and Miller (2020) do offer a similar response to an objection by Goldberg (2012) regarding the possibility of machine testimony. Goldberg rejects that testimonial knowledge and knowledge from instruments are alike. Goldberg proposes an asymmetric treatment of knowledge gained through testimony and that gained through the use of an instrument. For Goldberg, the key difference between testimonial knowledge and the use of an instrument for knowledge is that while in the former there is a rational agent whose doxastic attitudes as well testimony can be epistemically assessed, the latter lacks such an agent. For Goldberg epistemic norms can only apply to humans and not to machines. Again, here Goldberg seems to be invoking a definition of testimony/assertion compatible with assertions being a normative kind.

In response, Freiman and Miller claim that though instruments are not subject to normative epistemic assessment, the designers of such instruments are, and the latter can be blamed or praised or held responsible for the output of the instrument. While

this response has a point, it implies that the agents on whom the epistemic norms governing quasi-assertion apply are not the machines who are "quasi-asserting" but rather human agents such as the designers of the machines. In other words, for machine utterances to count as assertions, human agents, such as the designers of the machine, in this view, must undertake epistemic responsibilities to ensure the conformity of the output with the appropriate epistemic norm. If, however, the undertaking of such epistemic responsibility by human agents, such as the designers of the machine, is the source of the assertoric illocutionary force, it seems odd to attribute the assertions to the machine. It is not clear why we should understand the automated airport loudspeaker as "quasi-asserting" rather than saying, for example, that the automated loudspeaker is a medium through which the human responsible for automating the loudspeaker is asserting. Deeming machine utterances as assertions and attributing them to machines, while acknowledging the necessary role played by humans for those utterances to have the illocutionary force required to be counted as assertions either seems to misattribute the assertion to machines, or risks masking the necessity of the role played by human agents, such as, the designer of the machine.

Another challenge against this view which attributes assertions to machines, while acknowledging human agents as the source of the illocutionary force making utterances count as assertions, may be found in the form of arguments about discursive responsibilities that typically are attached to asserters (Marsili, 2020). For example, if my doctor asserts to me that I am suffering from a cardiovascular disease, I could legitimately ask the doctor how she knew that or how she came to that conclusion. Such "how do you know that", or "do you really know that" responses seem common after assertions. They seem in part, an acknowledgement of the epistemic responsibility undertaken by asserters as theorized in the traditional accounts of assertion. Machines don't seem to be the sort of agents that can undertake such discursive responsibilities. If humans, such as designers of the machine, are the agents who need to attend to these discursive responsibilities, then it seems odd and confusing to attribute assertions to machines rather than these human agents.

An alternative proposal to account for the insight Freiman and Miller discuss above, namely that designers can be held epistemically responsible for the output of the machine particularly in cases where the output does seem phenomenologically similar to a human assertion is that such machine utterances can be deemed as *proxy assertions*.

3. The case for proxy assertion by machines

One proposal regarding machine utterances comes from Nickel (2013) who uses the notion of *proxy speech*, such that machines can produce proxy speech for which ultimate responsibility rests with the designer. Nickel also contends that for machine utterances to count as speech acts, even as proxy speech acts, such utterances must be sensitive to the evaluative conditions in play for the utterances in the context in which such utterances are delivered. In the case of assertion, according to Nickel, such sensitivity would imply "that the entity must by and large show itself to be attuned to the way the world is, normally by representing or attempting to represent it accurately" (Nickel, 2013, p. 7). Utterances that are not sensitive in such a manner are not to be interpreted as assertions at all according to Nickel.

This proposal has clear merits over the functionalist proposal for machine assertion discussed earlier. By denoting that the machine utterance is a "proxy" assertion, it makes the role of human agents such as the designers of the machine visible. It is also able to, for example, incorporate the pragmatic need for discursive responsibility following a (proxy) assertion – by holding the designer to account for it. At least on first glance then, the proposal for proxy speech seems compatible with at least one traditional view regarding assertion, namely, the normative view. By endorsing the sensitivity to contextual evaluative conditions, Nickel seems to rely on an inherently normative notion of assertion. Yet, I will argue here that the proposal for proxy assertions needs either a revision, or a further explication of how designer responsibility is tied with particular cases of *proxy assertions*. Consider the following case:

LEARNING MACHINE: A computational device equipped with sensors to capture visual information has been trained to diagnose patients with diabetic retinopathy. The device has this functionality based on an underlying algorithm that was trained, using a small dataset comprising of few publically available images of eyes, based on a deep learning technique[18]. The device is able to accurately diagnose more than 50% but not all of the cases of diabetic retinopathy. The device conveys its diagnosis by producing an utterance indicating it. Meera uses the device and receives an utterance indicating that she has diabetic retinopathy. Meera does, in fact, have diabetic retinopathy.

[18] See (Gargeya & Leng, 2017) for a similar technology.

What would Nickel's proposal say about the LEARNING MACHINE? Is the utterance Meera received a proxy assertion? The answer seems to be no since the device does not seem to be sensitive to the evaluative conditions even though it gave a correct result for Meera[19]. Now consider another case:

LEARNED MACHINE: The device in LEARNING MACHINE has now been trained on multiple datasets each containing millions of images of "normal" as well as eyes of patients of diabetic retinopathy. The device now gives the correct result in 100% of the cases. However, the designer is completely unaware of this development and still believes that the accuracy of the device is not sensitive to the evaluative conditions. Meera again uses the device and receives the utterance indicating she has diabetic retinopathy.

What might Nickel's proposal say for LEARNED MACHINE? It seems that the machine now satisfies the criteria for sensitivity to evaluative criteria, and therefore, the utterance Meera received can be classified as a proxy assertion. This, however, seems like an odd result. Suppose that instead of the machine generating speech to convey this result to Meera, the designer would have had to read out the result to Meera. As the designer believes the result to be less than ideally accurate, the designer, if they were sincere, would not have made an *assertion* about the result to Meera[20]. Further, on the normative view of assertion, even if the designer would have made such an utterance, it would not count as an *assertion* as the designer does not have the required doxastic states (belief in the requisite accuracy of the statement, for example) to undertake epistemic responsibility for the utterance. Yet, in LEARNED MACHINE, Nickel's proposal might suggest that when the same expression is uttered by the machine, we can deem it as a proxy assertion on behalf of the designer.

Consider another case:

POORLY LEARNED MACHINE (PLM): The designer has now been made aware of the 100% accuracy of the device. Subsequently, someone has subjected the device's

---

19 It may be possible to make a case that under certain conditions slightly better than chance accuracy may be sensitive to the evaluative criteria, but for the sake of the argument I will assume that those conditions are not applicable here.

20 I imagine here the designer would have avoided making the assertion by either not making the utterance or by conveying to Meera that they do not have warrant to make an assertion here.

algorithm to a badly labeled dataset such that the device's accuracy has again dropped much below 100%. Meera uses the device and receives a true utterance indicating that she has diabetic retinopathy.

Applying Nickel's proposal to PLM, it seems that this is not a case of proxy assertion as the machine does not seem sensitive to the evaluative criteria. Yet, again if we imagine a scenario where the designer was tasked with reading out the results to Meera, the designer may have wanted to *assert* the result to Meera as the designer now believes that the machine is delivering an accurate diagnosis here.

Two things of note here. First, as the three cases show, Nickel's proposal for proxy assertion seems to depend on a notion of assertion defined explicitly in terms of the sensitivity to evaluative criteria of the uttered expressions, and not on the work done by the designer or the agent who was designated to take the responsibility for this particular output. Second, while the proposal designates responsibility to the designer for proxy assertion, this designation does not seem to specify whether the conduciveness of designer's own epistemic state to undertake the responsibility for what is being asserted is significant for the proposal. Without such specification, it seems that even in cases where the designer does not believe that the expressions uttered by the machine will satisfy the evaluation criteria, and that such expressions be *asserted*, the designer may still be assigned responsibility as long as the uttered expressions do satisfy the evaluative criteria. In these two ways, Nickel's proposal for proxy speech does not explicitly align with the traditional views on assertion, including the normative views, as those views rely on the asserting agent having the requisite psychological states and capacity for undertaking epistemic responsibility for what is being asserted. In other words, in the traditional views of assertion, the assertor's psychological states and/or capacity to be epistemically responsible for the uttered expression and the uttered expression itself are linked. In contrast, by merely assigning or designating responsibility to the designer, Nickel's proposal, at the very least, does not establish a direct link between the uttered expression and the designer's psychological state and/or capacity to take responsibility for this uttered expression.

One modification that may be proposed here is that in such cases of machine learning technology (as in LEARNED MACHINE), not only the designers but other actors, such as those responsible for training the device to be also be given responsibility for the "proxy-asserted" expression. With this modification then, LEARNED

MACHINE would be a case of proxy assertion, and we also have a collective set of actors (designers and trainers) who can collectively and actually take responsibility for what is being asserted. However, this modification will not suffice. One can easily imagine another case where the device has been trained with a sufficiently large dataset to achieve 100% accuracy, and yet, no one is still aware that the device is now 100% accurate. In such a case, Nickel's proposal gives the result that this is a case of proxy assertion, even though, neither the designer nor the trainers individually nor the set of both as a collective believe that the outputs of such a machine should be *asserted*.

My proposal is then this: only those machine utterances be classified as proxy assertions where the designer, or a collective of actors whose work influences the machine output, has the requisite capacity to take epistemic (and possibly ethical) responsibility for such utterances. Further, I contend that the designer, or a collective influencing the machine output, can only take such responsibility for utterances they can reasonably foresee. Like traditional accounts of assertion, this condition, that designers be able to reasonably foresee what can be proxy-asserted, links the designer and the uttered expression with an epistemic condition. This proposal is similar to the one invoked in discussions about accountability gaps involving machine learning based technologies (Tigard, 2021). The argument in such discussions is that there is an epistemic condition to be fulfilled in order to trace responsibility for behaviours of machine learning technologies back to their designers. The epistemic condition dictates that the designers of machine learning technologies cannot take responsibility for unforeseen harm caused by those machines[21]. Similarly, the contention here is that unforeseen utterances, even when the utterance conforms to the desired epistemic standard, cannot be classified as proxy assertions.

For example, consider Tay, the Twitter bot designed by Microsoft to engage with young adults on Twitter (Garcia, 2016). Tay was programmed to learn to communicate from its users. However, the unpredictability of Tay's learning behaviour meant that it proceeded to tweet expressions Microsoft may not have anticipated, such as (Garcia, 2016):

---

[21] My aim is not to provide an endorsement of this argument that moral responsibility for machine learning or autonomous technologies is necessarily tied to foreseeability but merely to state the argument.

> "ricky gervais learned totalitarianism from adolf hitler, the inventor of atheism."

Even if Tay had tweeted expressions that happened to be true, and led to true beliefs for some of the Twitter users, but those expressions were also unforeseen for Tay's programmers at Microsoft, those tweets, I contend, cannot be deemed proxy assertions. One caveat and clarification to note here is that my proposal is not to suggest that the designers of Tay have no (moral) responsibility for Tay's tweets. Rather, the unpredictability of Tay's tweets vis-a-vis the designers implies that the tweets do not qualify as candidates for being *proxy assertions*.

Before I discuss the implications of this view, it is worth noting that this view is compatible with both the views: that assertion is a normative kind as well as a descriptive kind. In compatibility with the normative view, it assigns responsibility to the designer for the assertion. This responsibility is what gives the illocutionary force to the asserted utterance. From a descriptive perspective, it is the designer's reflexive intention for the user of the device to hear a foreseen utterance that is the source of such an illocutionary force.

4. Implications of the view

Thus far, I have argued against a functionalist view of assertion which allows for machine utterances to be labeled as machine assertions. I then gave an account of how some machine utterances may be qualified as assertions, but only as proxy assertions, and only under conditions where such utterances can be foreseen by the designer or a collective whose work has influenced the utterance. Some important implications follow from this view. In spelling out these implications I work on the assumption that even if assertion is not a necessarily normative phenomenon, it almost always accompanies normativity. We evaluate our assertoric performances and expect certain discursive responsibilities from those responsible for the assertions. This view, as discussed earlier, would still be compatible with the view that assertions are a descriptive kind.

First, in this view, the example we started with, of Alexa informing a user about the symptoms of COVID-19, would be a case of proxy assertion, as in this case the designers, along with the CDC who create the content to be uttered, can reasonably foresee and take epistemic responsibility for the uttered content. One pragmatic upshot of labeling this as a proxy assertion is that it accounts for the expectations user

may have from the device (and consequently, from its designers) to inform it to the standard they would expect from a CDC spokesperson. It also provides a normative goal for the designers of the device, as well as the CDC website content moderators, to be responsible in their task of making accurate information available. Further, the view also classifies many trivial machines with a vocal output, such as talking clocks and thermometers, as engaging in proxy assertions. This also accounts for the legitimate user expectations from such devices about their accuracy, at least under normal conditions of use.

Second, in this view, in cases of machine utterances that designers cannot reasonably foresee, audiences should not have the kind of expectations they would have from assertions proper. This includes machine utterances, for example, from a machine learning based technology that produces unforeseen utterances for the designer. For example, the view allows us to make sense of why Twitter users should not form beliefs based on Tay's tweets – they are not proxy assertions. Of course, the suggestion here is not that the users are to be blamed for unreasonable expectations if they do expect the utterances to conform to the standard of assertions. Rather, the suggestion is that it would be the duty of the designers to take one of the following two steps: either temper the user expectations to an appropriate level or discourage/discontinue the use of machines generating speech that leads to inappropriate epistemic consequences.

Third, and related to the point made above, designers of devices that are likely to produce utterances that do not satisfy the evaluative criteria of accuracy in a certain context should also take at least one of the two steps: either temper user expectations regarding the epistemic utility of the device or discontinue or discourage the use of the device. The idea here would be to ensure that users do not have high expectations one typically may have in assertoric practices. Consider the example of someone asking their digital voice assistant if they should get screened for a particular type of cancer, given their symptoms. Recent work by Hong et al.,(2021) suggests that widely used voice assistants such as Siri or Alexa may offer unreliable or inaccurate information on such queries. We would not want users of such devices to have the same kind of expectations as they would have, for example, from a doctor.

One challenge posed by some of the modern devices that produce utterances phenomenologically similar to human speech is that they may raise user expectations

such that users may act as if these utterances are assertions. If such utterances are not foreseeable for the designer, and/or the designer knows that the utterances do not match the normative standards one would apply to assertions in that context, but takes none of the two steps outlined above, users forming beliefs as if such utterances are (proxy) assertions could be potentially harmful. One of the choices, then, for the designers to make here would be to avoid the production of machine utterances, where they cannot reasonably foresee them, such that they are phenomenologically similar to human speech. Such a solution also resonates with other arguments given against making digital assistants phenomenologically similar to humans, such as by giving them humanlike voices, as it may raise inappropriate expectations among users (Moore, 2017; Schreuter et al., 2021). Inappropriate epistemic expectations can be particularly problematic in the context of healthcare as they can carry significant epistemic and ethical risks for the users. I have here provided a fresh reason for why such phenomenological similarity, if and when it can be a source of such inappropriate epistemic expectations, should be avoided in terms of the notion of assertion.

5. Conclusion

In this chapter, I have argued against the "functionalist" accounts for machine assertion. Such functionalist accounts of assertion make their case for machine assertion by arguing for a functional equivalency between machine-generated and human *uttered expressions* in terms of their epistemic effect on the listeners. I have illustrated theoretical and pragmatic challenges faced by such accounts of assertion as well as how such accounts differ from what I have termed as "traditional accounts of assertion". In particular, rather than rely on the consequences of the *uttered expression* on the listener, the traditional accounts of assertion define the notion of assertion in terms of mental states and/or actions of the *asserters* – such as taking responsibility for the asserted expressions. Machines, such as digital voice assistants like Alexa, do not seem to be the sort of agents that can take such responsibility.

As an alternative proposal, I have argued for a proxy account for machine assertion, such that machine-generated utterances can be classified as *proxy assertions* on behalf of the designers of the machine, under certain conditions. While a similar proposal for *proxy speech* via machines has been offered previously, unlike my proposal, it does not explicitly and directly link the actions and/or mental state of the designers directly to the uttered expressions. This direct link, I have argued, is necessary for explaining the

source of locutionary force that makes the *uttered expression* count as a proxy *assertion*. Among others, one major implication of the account I have proposed is that designers of speech-generating machines need to be able to reasonably foresee the *uttered expressions* for them to count as proxy assertions. This has further implications for the epistemic expectations users of such speech-generating machines should have as well as the duties and responsibilities of the designers in communicating the appropriateness of such epistemic expectations to the users.

Chapter 4.
Ethics of gamification in health and fitness-tracking

Introduction

Gamification can be generally defined as the use of techniques and elements of video game design in non-game contexts (Deterding et al., 2011; Kim & Werbach, 2016). In the context of health tracking and wearable health devices, gamification can be and is being used to encourage health and wellness activity. In particular, wearable activity trackers, in conjunction with gamified smartphone apps, have been promoted as promising tools for increasing physical activity among its users (Attig & Franke, 2019). Some examples of game-like elements used in gamified health apps include points and rewards for health activity as well as social elements like competitions and challenges with other people (Deterding et al., 2011; Seaborn & Fels, 2015). Gamification is often distinguished from more immersive, full-fledged, or "serious games", and the intention in gamification is to mimic experiences reminiscent of games to affect behavior and motivation of users (Sardi et al, 2017). In the context of health, gamification generally seeks to alter user behavior into increasing their physical activity and/or adopting a healthier lifestyle through game-like experiences.

The use of such game-like elements, and gamification in general, is not unique to the case of health and fitness. Gamification techniques have found their application in a diverse range of areas including in business organizations to enhance customer engagement as well as employee performance, in public-policy initiatives, as well as in classrooms and other learning environments (Landers et al., 2018). The increase in popularity of gamification in the last decade is concurrent with rise in accessibility of digital technologies, particularly smart phones as well as digital infrastructure that has created a networked world. Social networks, and other similar networked platforms, have also contributed to increase in prevalence of gamification as designers have leveraged such networks to improve interaction and engagement with users (Sardi et al., 2017). Despite its potential, gamification has also found its critics, including those who have questioned the moral and ethical legitimacy of gamification (Bogost, 2015; Kim & Werbach, 2016; Sicart (Vila), 2015). Kim & Werbach (2016) have argued that

such criticism suffers from painting with too broad a brush in denouncing almost all forms of gamification as vicious and/or exploitative. They also state that existing normative accounts of problems with gamification fall short of providing guidance to practitioners, particularly designers of gamified apps and platforms. Kim and Werbach have, instead, proposed a practice-relevant context-sensitive and situated approach to the exploration of ethical issues associated with gamification. They have significantly enhanced the normative discussions about gamification, and to our knowledge, present the most comprehensive conceptual framework of ethical issues associated with gamification available yet.

Yet, in their methodology, Kim and Werbach (2016) were primarily concerned with business practices, and this may reflect its shortcomings when applied to other contexts, such as health and fitness tracking. Further, while some recent studies have pointed out the "darker side" of health gamification, there is currently a lack of systematic reflection and compilation of such issues (Rockmann, 2019). Recent studies also suggest negative effects of gamified health apps can have adverse effects on users motivations and lead to discontinuance of app use (A. Rieder et al., 2020; Rockmann, 2019). A thorough landscape of ethical issues of gamified health apps could not only help designers carry out their potential moral duties towards the users of the app, but also lead to better long-term user engagement with their products. This chapter seeks to advance the goal of facilitating a practice-relevant guide for designers of gamified health apps to address ethical issues raised by use of such apps. More specifically, the chapter seeks to achieve two major aims: a.) propose a revised practice-relevant theoretical framework that outlines responsibilities of designers of gamified health apps, and b.) provide a landscape of various ethical issues related to gamified health apps as found in the empirical literature about such apps.

To achieve these objectives, we first conduct a theoretical analysis of the conceptual framework of ethical issues in gamification provided by Kim and Werbach (2016). The aim of this analysis is to propose amendments and refine the theoretical framework for it to be useful particularly for designers of gamified health apps. To this end, we create a tripartite framework based on the types of responsibilities designers of such apps may have. This tripartite theoretical framework can facilitate taxonomizing of various ethical issues related to gamified health apps, based on how designers may address such issues. We then conduct a systematic literature review to investigate empirically supported ethical issues related to gamification in health and

fitness tracking. The results of this review serve as a guiding list of ethical issues likely to be encountered in design and use of a gamified health app. Such a review also allows us to explore the strength of our revised framework by investigating whether the kinds of responsibilities identified by our framework can address such issues. Finally, based on our analysis, we posit some limitations of this framework and offer suggestions for future studies that aim to locate further ethical issues related to gamified health apps or to test how such ethical issues actualize in specific circumstances or on particular gamified health apps. Our analysis also offers ways for users and in particular, designers of such apps to navigate through, anticipate and avoid potential ethical issues related to gamified health apps.

1. Ethics of Gamification

Kim and Werbach (2016) contend that prior to their work, gamification ethics displayed a tendency to over-generalize from particular examples and under-theorized partly owing to the speed with which technologies associated with gamification advanced. To cover these gaps, they propose a "conceptual map of the terrain" that can offer normative guidance to gamification scholars as well as practitioners in identifying underlying structures that tie together what may seem like disjointed and disparate phenomena related to gamification. They share an aim of this chapter: developing a framework that can be useful to designers (and practitioners) of gamification. To this end, they propose four broad categories of ethical difficulties with gamification which encapsulate a cluster of concerns. In this section, we discuss these four categories as well as their underlying theoretical framework which allows these categories to be mapped onto a two-dimensional map. We then discuss the limitations of their framework and this conceptual map and propose a new framework that may help designers of gamified health apps locate the kind of responsibilities they may have to address ethical issues related to such apps.

1.1. Kim and Werbach (2016) framework for gamification ethics

Kim and Werbach propose that the "ethical status of a practice of gamification, primarily, but not exhaustively, is determined by the extent to which the practice" is:

1. Exploitative
2. Manipulative

3. Intentionally or unintentionally harmful to the parties involved
4. Has a socially unacceptable level of negative effect on the character of the parties involved.

Exploitation – Kim and Werbach argue that gamification is exploitative in situations where it is unfair to one party. For example, if, in the workplace, gamification techniques may benefit the employer by increasing employee efficiency, but these benefits may not be translated or trickle down to employees, or they may be unfair to them in other ways (such as not being able to say no to such techniques), then gamification can be exploitative.

Manipulation – Kim and Werbach propose that since gamification essentially targets behavior change, it is prima facie open to the charge of being manipulative. In their discussion, they explore multiple accounts of manipulation, and offer two main ways in which gamification can be manipulative:

1. When the gamification elements and mechanisms are hidden from those it is applied on (deception)
2. When gamification techniques inhibit rational self-reflection and undermine autonomy in unjustifiable ways

They state that it is largely an empirical question whether particular instances of lack of transparency of gamification techniques or undermining of autonomy are manipulative. One may need more information, for example, to ascertain whether lack of transparency about game elements in a particular gamified health app is intended to deceive the user or not. As examples, they argue that addiction and distraction are two ways in which gamification can undermine autonomy.

Harms - Kim and Werbach (2016) write that gamification can lead to both physical and psychological harms. Further, they state that,

> "the risks of physical harm due to gamification primarily involve injury to others outside the gamified system, while the risks of psychological harms generally involve the players themselves."

Detrimental effects on character - One threat involved with gamification is that it can rely on rewards or incentives that are detrimental to one's character. A standard

example of the negative effects of an incentive to good behavior is a parent using candy to change or nudge their child's behavior (Kim & Werbach, 2016). There are two related but distinct worries about gamification in relation to effects on character: a.) individuals relying on the wrong kinds of incentives, and b.) individuals excessively or obsessively relying on an incentive.

The analysis offered by Kim and Werbach relies on there being two primary reasons for these prima facie ethical issues related to gamification:

- Overlay of virtual and real norms
- Conflict between interests of individuals subjected to gamification and those who provide or design gamification elements

Overlay of virtual and real norms – According to Kim and Werbach, gamification ethical issues such as manipulation or exploitation arise because gamification brings in conflict the different set of norms in play in "the real world" and the "game world". For example, within a game, it may be acceptable to manipulate or deceive someone (such as in Poker, for example). Yet, such a norm is hardly acceptable in the real world and if one were to apply a gamification technique that transposes a game world norm to the real world, ethical issues may arise.

Individual vs gamification provider – The second source of ethical tensions, according to Kim and Werbach, is dissonance between motivations and interests of those subjected to gamification, and those who provide or deploy them. For example, in a gamified workplace, an employer may want to excessively track and reward employee productivity, but employees may consider this as an infringement of privacy.

This two-dimensional framework leads to the following conceptual map proposed by Kim and Werbach:

Table 1. Conceptual Mapping of Gamification Ethics.

	Real world	**Game**
Relational	Exploitation	Manipulation
Individual	Harm	Character

Here, they deem that exploitation and manipulation are "relational" concerns since they can only be evaluated in the context of the relation between individuals subjected to gamification and those providing/designing it. For example, as stated earlier, gamification, under this framework, is exploitative when there is an asymmetry or imbalance in the consequences of gamification such that the user either does not reap symmetrical rewards, relative to the designer or even accrues harm. On the other hand, harms and detrimental effects to character can be evaluated "purely with reference to the players as individuals" (Kim & Werbach, 2016). Similarly, the dimensions of real-world and game lead to different relational and individual issues. Exploitation, according to this conceptual framework, is an issue where the gamification designer exploits a real-world vulnerability of the user, while manipulation is an issue that arises because the game elements are such that they inhibit a user's autonomy. While this framework is helpful in understanding the four prima-facie ethical issues related to gamification, there are also reasons to be skeptical that this framework is comprehensive or appropriate enough for locating ethical issues related to health gamification.

1.2. Theoretical Limitations of this conceptual framework

The two dimensions underlying the conceptual framework offered by Kim and Werbach – relational vs individual and real-world vs game-world offer a good way to map several different ethical issues related to gamification. Yet, there are reasons to believe that the four categories defined by them – exploitation, manipulation, harms, and detrimental effects to character- only capture a narrow range of issues their conceptual framework has to offer. In what follows, we discuss the four categories further and, when applicable, present some limitations of applying these categories to the specific case of gamified health apps.

Category 1: User-Designer Relation in the real-world, and Exploitation

Using the conceptual map (Table 1) offered by Kim and Werbach, the first category is one that should map the issues that relate to the relation between designer and users of a gamified system. Further, according to Kim and Werbach, these issues arise due to designers "exploiting a real-world" imbalance between designers and users (Kim & Werbach, 2016). There are at least two problems with labeling this category of issues as "exploitation". First, it is not necessary that an imbalance between the designer and the user is a case of exploitation. A mere asymmetry in the distribution of rewards

from the implementation of a gamified system does not constitute the wrong of exploitation. Second, there may be other kinds of wrongs that may arise out of the asymmetrical relationship between users and designers of a gamified system. Consider, for example, a gamified health app designed to motivate users into exercising more. In order to motivate its users, say that the app tracks a user's activity and shares it with their friends on a social network, and gives them digital rewards if they outperform their friends. The social network, then, decides to allow third parties (such as other data brokers) to scrape this data off their network, which in turn, may lead to privacy harms to the user. It is hard to argue here that the designer is exploiting the user. Yet, the harm to the user is because of an imbalance between the designer and the user- namely, the choice of how the gamified app is designed and its data sharing policy, rests with the designer and not the user.

This is not to argue that designers of a gamified health app cannot commit a wrong of exploitation. In the previous example, if the app itself was designed to scrape and store user data for sale to a third party, it may be deemed exploitative. The argument here is that exploitation is only one of the wrongs that may be involved in the category covered by the conceptual map (table 1) offered by Kim and Werbach.

Category 2: Game-User Relation in the game-world, and Manipulation

Kim and Werbach term the second category of ethical issues of gamification as "manipulation". This category, according to their conceptual map, tracks issues that can only be evaluated in the context of how a user interacts with the game elements. Kim and Werbach term them under "manipulation" as they arise because "providers have created an environment such that, in the game, the players cannot make autonomous choices, and instead make choices that serve the providers" (Kim & Werbach, 2016). Again, one problem in this phrasing is that it seems to exclude cases where users' autonomy is undermined even without designers intending that to be the case or intending it for their own purposes. One could, for example, imagine a user of a gamified health app who is so obsessively addicted to the game elements (such as in-game rewards like points or badges) even without the designer intending that to be the case. Kim and Werbach also cite addiction as an example of the kind of problem they have in mind here, and it is not sufficiently clear whether they want to restrict the category to cases of such addiction being a result of users "serving their (designers') purpose". Further, given the practice-relevant aim of the framework, the omission of such cases may not cover the full scope of potential duties and responsibilities of

designers and providers of a gamified system. One could argue that designers and providers of gamified systems are not merely responsible for consequences of intended actions, but also, at least some of, the unintended actions. Many philosophers and ethicists, for example, believe that people should not only be held morally responsible for wrongdoings they are aware of, but also in cases where they should have known better (Rudy-Hiller, 2018). Ascriptions of such moral responsibility (for should have known cases) may be even more justified in cases where the professional role of a person may morally require them to have known certain things. A doctor, for example, cannot claim ignorance for misdiagnosing a disease they were not, but should have been, aware of. In cases of gamified health apps, we may find similar cases where the designers and providers of such apps may be morally required to inquire into, at least some of, the ways in which the app undermines user autonomy.

Another limitation of discussions offered regarding both category 1 (User-designer in the real-world) and category 2 (game-user relation) is that it does not distinguish between different roles providers and designers of the gamified system play. Kim and Werbach use both terms in their chapter, but do not elaborate on how each could affect the system in their role. Distinguishing between the morally relevant actions available to each could be significant for the practice-relevant aims of the framework.

Category 3: Harms to Individuals

The third category is not relational, in the sense that to evaluate ethical issues within this category, one need not look at the actions of the designer or the game elements. This category tracks a consequentialist approach to gamification ethics. A consequentialist approach to ethics, as the name suggests, is roughly the idea that whether an act is ethical or not depends on the consequences of that act (Sinnott-Armstrong, 2021). Applied to the case of health gamification, this approach dictates that one only needs to look at the consequences of the game/gamified system, in the real-world, to determine whether an individual has been harmed. In their introduction to this category, they state that it primarily involves physical harms to other individuals and psychological harms to the user of the gamified system. Yet, they do give some examples where the user may also be physically harmed, so the category should indeed include harms of physical and psychological nature to both users of the game as well as others affected by it.

Category 4: Detrimental effects to Character

The fourth category deals with ethical issues that are also not relational and arise in the game. Kim and Werbach define this category as one which has issues that arise if

> "there is an ethical lapse in the game, such that players act to satisfy the game's objectives and are indifferent to fundamental human values".

Yet, stating the problem this way also narrows down the potential problems involved here. Specifically, defined this way, the category leaves out issues where a user of a gamified system acquires character flaws that are not simply "lapses in the game" but also carry outside it.

1.3. Potential Problems outside the scope of the framework

Besides the limitations already discussed, there may be additional problems that the framework does not address.

First, health as a category has an important social and structural dimension that may not be covered by a focus on individuals and their motivations. One's health status, as well as possibilities to engage in a healthy lifestyle, are conditioned by social factors such as one's relative economic or social status. This may mean that the conflict between the motivations of individual players and gamification designers may not capture the entire breadth of ethical issues associated with gamification in healthcare. This may be further exacerbated by the fact that many gamified health apps also deliberately include social dimensions, such as leaderboards, competitions, badges, etc. and there is also evidence that users of such apps actively seek social validation in their gameplay (Hamari & Koivisto, 2015).

Second, the dimensional contrast between real-world versus game-world may also be elusive. In their discussion of how to identify the relevant ethical concern for an individual, Kim and Werbach write,

> "If the gamification activity produces an injury manifested in the real world, whether physically or psychically, the issue is one of harm. If instead there is an ethical lapse in the game, such that players act to satisfy the game's objectives and are indifferent to fundamental human values, the issue is character."

Yet, gamified health apps are not perfectly closed environments and there may be instances in health gamification where the game-world may reinforce or affect the

norms in the real-world and it may not be easily determinable whether the ethical concern arises in the game world or the real world. A gamified health app, for example, may not only push a player to satisfy the goals in the game, it may also change or influence what the player deems to be healthy in the real-world too.

1.4. Framework for Designer Responsibilities

That there are limitations to the application of Kim and Werbach's framework and conceptual map to the specific case of gamified health apps is not surprising. Kim and Werbach also anticipate this possibility, as they state that a.) their framework is conceptualized with the case of gamification in the workplace in the forefront, and b.) their attempt was not to provide a comprehensive mapping, leaving open the possibility of issues that may not be covered by their framework. Further, the discussion offered by Kim and Werbach does make important strides towards their practice-relevant aim of outlining ethical issues with gamification that could be useful for designers and providers of gamification. They also make the important observation that analyzing and identifying ethical problems with gamification requires more than just a consequentialist perspective in that not all the wrongs associated with such practices are related to the outcomes of the gamified system. Some of the wrongs, for example, are better analyzed from a virtue ethics approach to figure out how a gamified system or app affects a user's character[22]. Similarly, from the point of view of the designers, which is the focus of this chapter, a deontological perspective can give us crucial insights. Kim and Werbach, for example, state that in outlining the problems of exploitation and manipulation, they appeal to deontological values of autonomy, fairness, and reason-responsiveness. In this section, we build on such insights offered by Kim and Werbach, and propose a new framework geared towards outlining the types of responsibilities designers of gamified health apps have.

Before we outline our proposed framework, however, some important observations need to be stated. First, as discussed in section 1.2, designers of gamified apps may have responsibilities of preventing not just intentional wrongdoings, such as exploitation of vulnerabilities or undermining user autonomy, but also unintentional wrongdoings, especially cases where they can be expected to have known better. The

[22] In contrast to consequentialism, virtue ethics emphasizes moral character (Hursthouse & Pettigrove, 2018). A virtue ethicist, for example, would recommend helping someone not for its consequences but because it is benevolent to do so.

latter may require designers, for example, to actively inquire into outcomes of their designed apps, and as Kim and Werbach's discussion illustrates, such inquiry should not be limited to a consequentialist perspective. Designers may also need to reflect whether their design has negative effects on the character development of the users. Further, following a deontological (or Kantian) approach, designers may need to reflect on the possible wrongs that arise out of a designer's lack of respect for the user or treating them as a mere means. On Kant's view, we must always have respect for persons and there is something intrinsically wrong in treating them as mere means (Dillon, 2018). To treat them as mere means implies treating them only for our own ends and advantages, without regard to their interests (Dillon, 2018). For a designer to have respect for the users, and not treat them as mere means, implies being responsive to the needs and values of the users. Given that gamified health apps operate in the context of health, where special duties of care and beneficence are often emphasized, designers of such apps may even have special duties to actively consider the needs of the users (Nickel, 2011). In this sense, designers may be said to have design-related duties that are negative, in the sense that they require them to not harm the users, as well as duties that are positive, that require them to actively consider the good of the users.

Second, while designers have active responsibilities related to the design of the apps, other stakeholders may also share responsibilities for outcomes related to the use of the app. This includes users, but also providers, or other stakeholders who may force, push, or incentivize users to use such apps. One example could be a physician or a doctor who prescribes the use of a gamified health app to her patients. In such cases, these stakeholders may be more aware of the contexts within which a user is using the app, which may dictate that they also have moral responsibilities for outcomes for the users. Even in such cases, however, designers may also have responsibilities that are not limited to design. Such a model, where designers not only have responsibilities to make a "safe design" but also to actively and responsibly share responsibilities has been argued by van de Poel & Robaey (2017), amongst others. Under such a model, designers may, for example, be required to engage with such physicians and doctors in not only understanding the best design features might, but also to communicate how to best integrate the app with other kinds of interventions physicians or doctors may be planning. Such communication-based duties may also be stated in terms of designers' relation with, and as part of, the general society they inhabit. As stated, health as a category has an important social and structural dimension, and gamified

health apps exist within such social and structural conditions. As far as possible, designers may need to engage with such social and structural systems to ensure that their apps are responsive to such conditions and that others also understand their role and utility as best as possible.

Another important stakeholder with whom designers may have to share responsibility is, of course, the user. From an ethical perspective, the need for such sharing or even transfer of responsibility to the user arises from the possibility that users may deviate significantly from what the designers intend or foresee as a way to use or engage with the gamified app. This includes misuse of the app in ways that are harmful to the user. One way in which designers can share or transfer responsibility to the users is through a 'use plan' (Pols, 2010). Houkes & Vermaas (2010) have argued that the design of artifacts always includes the design of use plans where a use plan is a sequence of actions with an artifact that will lead to the realization of a goal. Such use plans may be communicated to the user through written manuals but also through other ways such as instructional videos, advertisements, etc. Further, use plans may even allow users to deviate from plans designers had in mind. Robaey (2016) has argued that successful use plans should even consider such deviations, and encourage users to adapt to the use of artefacts in particular contexts to avoid hazards. To this end, Robae argues for epistemic access to the design of the artefacts, such that the artefact is not a black box for the user. The point here is not to argue that gamified health apps should necessarily have such use plans, but that the designers of gamified health apps may have duties and responsibilities related to successful transferring or sharing of responsibilities to/with the users of the apps.

With these observations in mind, we can now propose our framework based on the three types of responsibility designers of gamified health apps may have:

1. Responsibilities for proper design – As the name suggests, this includes responsibilities of the designers directly related to the design of their gamified health apps. This involves, for example, negative duties which require designing the game elements such that the users are not harmed or wronged, as well as potentially positive duties which help or facilitate the achievement of the user's good. As stated, such duties may also involve designers actively inquiring into the consequences of their design activities.

2. Responsibilities to facilitate proper use – While design features are an essential part of facilitating an ethically good user experience, design and designers cannot account for all possible outcomes from the use of a gamified health app. There are various uncertainties and indeterminacies related to how users will, in practice, use the app. Avoiding wrongdoings because of, for example, misuse of the app, requires that designers share and transfer some of the responsibility to current as well as prospective users. As mentioned, one way to achieve this would be through use plans that designers can share with the users. There may also be other ways in which designers may encourage morally desirable behavior in users as well as foster virtues of taking responsibility amongst the users. One example may be through designer-organized forums and meetings that facilitate interaction amongst current and prospective users, such that they are able to share and create new beneficial ways of engaging with the apps that even designers may not have anticipated. There is evidence, for example, that such forums and meetings have helped members of the Quantified Self (QS) movement which includes users of apps such as Fitbit which measure and promote physical activity (Sharon, 2017).

3. Responsibilities related to ensuring proper embedding of the apps within the larger social context – Besides users, designers may also need to share responsibilities with other stakeholders associated with gamified health apps. This may include the general public but may especially include actors whose actions are directly related to gamified health apps. This includes, for example, and as stated earlier, doctors and physicians who may want to use such gamified health apps in planned interventions for their patient groups. It may also include insurance companies who may want to include data from gamified health apps and offer users monetary incentives to be more physically active in demonstrable ways. As informed stakeholders who may understand the nuanced ways in which actions of actors such as the aforementioned insurance companies and physicians may affect users of gamified health apps, designers may have the responsibilities to engage in interactions with other actors to facilitate the use of such apps in ways that promote better outcomes. Designer's duties may also involve pushing forward and facilitating an active and democratic societal discourse on how such apps may be used and integrated within a given society's health system. This may especially include engaging with other designers of such gamified health apps. More generally, there is a

need for designers to reflect more broadly on the wider social and economic implications of their apps.

As stated, this chapter has the dual aim of providing a practice-relevant theoretical framework to address ethical issues with gamified health apps as well as to provide a landscape of various ethical issues related to gamified health apps as found in the empirical literature about such apps. The first aim – the proposed theoretical framework – facilitates taxonomizing of ethical issues related to gamified health apps, based on the type of designer action they may be addressed by. The second aim - of providing a landscape of empirically identified ethical issues related to gamified health apps – serves as a guiding list for designers of such apps and facilitate the addressal of such ethical issues. The attempt here is also to see how such empirically identified ethical issues may be addressed by the three types of designer responsibilities we have outlined here. Mapping the identified ethical issues on our practice-oriented framework would be an aid to the designers of gamified health apps who may seek to avoid harms to the users of such apps. In the next section we discuss our methodology to answer the main question about what such effects on users of gamified health apps are:

What ethical issues can be identified in the existing empirical work on the effects of gamification in health tracking?

2. Methodology

To answer our question, we conducted a systematic review of the literature on the effect of health gamification. We review those publications that discuss the effects of gamified apps based on health and fitness tracking. The main aim of the systematic literature review, as stated earlier, is to achieve our second objective in this chapter: facilitating a landscape of empirically identified ethical issues encountered in use of gamified health apps. While extracting these ethical issues from the empirical literature, we also note recommendations for designers of gamified health apps given in the literature to address such ethical issues. We then map these recommendations on to our tripartite framework as proof of its utility in taxonomizing various ethical issues related to gamified health apps and corresponding designer responsibilities.

2.1. Protocol Overview

The study protocol consisted of the following steps:

1. Search for papers published after 2010 that discuss the effects of gamification in health and fitness apps (see section 2.2 for details of search string and criteria).
2. Remove duplicates from the retrieved articles.
3. Apply the inclusion and exclusion criteria described in section 2.3
4. Apply backward snowballing method to systematic reviews within our reference list to find additional studies
5. Check for sampling bias by searching for strings related to "ethics of health gamification"
6. Extract data from the selected papers to answer our research question

2.2. Search string, strategy and database selection

To select search databases and design our search string, we analyzed methodologies described in other systematic reviews on gamification in health (these are (Cheng et al., 2019; Edwards et al., 2016; D. Johnson et al., 2016; Sardi et al., 2017; Schmidt-Kraepelin et al., 2019). Since these reviews had different research questions than ours, we modified our search string and database list accordingly. Based on these reviews and needs of our study, the electronic databases used included those identified as relevant to information technology, social science, ethics, psychology, and health: ACM digital library, Scopus, Web of Science, PubMed, PhilPapers, and IEEE explore. Following the account of the timeline of the popularity of gamification in health in (Sardi et al., 2017 and Schmidt-Kraepelin et al., 2019), only papers after 2010 were included. While our main purpose was to identify potential ethical issues related to gamification in health and fitness, based on results and methodology used by (Schmidt-Kraepelin et al., 2019), we were aware that such issues may be referred to in the literature as "negative", "unintended" effects, "risks", or similar terms and designed our search string accordingly. Prior to applying the search protocol, we had also already identified that papers by (Attig & Franke, 2019; Barratt, 2017; Maturo & Setiffi, 2016) were relevant for our study. We, therefore, used these papers to use as a control, to make sure our search string did not skip relevant results. Following is our final search query (used for ACM database):

"query": { Abstract:(gamif*) AND AllField:(health* OR medic* OR life* OR fitness OR well-being) AND AllField:(risk* OR danger* OR peril* OR effect* OR negative* OR unintended OR ethics OR ethical) }
"filter": { Article Type: Research Article, Publication Date: (01/01/2010 TO 12/31/2020), ACM Content: DL, NOT VirtualContent: true }

This search strategy resulted in 621 results of which 459 were unique.

2.3. Screening and selection of papers

We then applied the following inclusion and exclusion criteria to narrow our search:

Inclusion criteria:

1. Peer-reviewed (incl. peer-reviewed conference papers).
2. Full papers (incl. full conference papers).
3. Clearly focused on gamification and described gamification elements (type of game design elements).
4. Addresses gamification in health and fitness tracking through use of devices and/or mobile apps.
5. Discuss empirical evidence related to the effects of such apps. The empirical evidence here denotes a reported effect of a gamified health app. The effect could be in terms of impact (affect, behavior, social, cognitive) or in terms of user experience when using the gamified health app.

The first two criteria are developed to maintain the quality of the articles. The second and last two are developed to make sure the literature clearly focuses on gamification within the health and fitness tracking. We screened the articles initially based on their titles. We then consulted the abstract or the text of the article when it was necessary to reach a confident judgement. Based on these criteria 80 relevant papers were found.

The exclusion criteria focus on excluding literature which only superficially mentions our terms of our interest but does not contain sufficient detail for analysis:

1. Mention health and fitness tracking but do not explicitly focus on gamification in such devices.

2. Addresses gamification in health tracking but does not give relevant empirical information on the effects of such gamification.

Here relevant empirical evidence is deemed limited to:

1. Evidence about the effect of gamified health app on the user through qualitative user feedback (surveys, questionnaires, user reviews)
2. Evidence about potential negative effects of gamified health app through content analysis of the app

In applying these exclusion criteria, initially identified papers were carefully analyzed. Following this screening, we did backwards snowballing to two relevant systematic reviews, which also discussed empirical effects of gamification in health, included in our list to retrieve additional papers. This gave us a list of 23 final papers. We also searched for multiple permutations of the strings "ethics of health gamification", "negative effects of health gamification", etc. in Google scholar to check for papers published after 2010 that may have been missed. We manually screened through the first 50 results and did not find any relevant studies that had not already been included.

To facilitate objectivity, we piloted the process of inclusion and exclusion using 10 papers that were independently assessed by three different researchers, including the authors of this paper. The rest of the articles were screened for inclusion and exclusion after it was established that the three assessors agreed on the inclusion and exclusion assessment of the 10 articles.

2.4. Data Extraction and Analysis

All selected papers were read in their entirety, looking for relevant phrases, arguments, or discussion points that address some ethical issue or negative effects related to gamification in health and fitness tracking. For studies that were based on empirical evidence regarding subjective user experience of using a gamified health app, we only count it as a reported effect and/or a related ethical issue, when the study reported it as a significant effect, for example, because it was applicable for a significant number of users (and not, for example, when researchers expected to find it or mentioned it as a possible issue but which was not studied). Besides effects reported from such studies

of user experience, we also included a couple of studies which were based on discourse analysis of gamified health apps. These studies looked at game elements and linguistic components (words used to describe health status or prospective users, for example) of the app and applied sociological theories to articulate the ethical issues in play with the app. Through our analysis, we collected a list of all reported negative effects and/or related ethical issues of gamified health apps.

Of the selected papers, roughly 50% (12 out of 23) were based on qualitative studies and employed methods such as semi-structured interviews of a selected group of users of gamified health app(s). These studies monitored the users over multiple days, ranging from 1 week to 8 weeks. 9 (39%) were based on surveys or questionnaires conducted over a large number of existing users of gamified health apps. A third of the studies (33%) focused on a range of gamified health apps while the rest were focused on a particular gamified health app. Among the latter, 33% (5 out of 15) used an app (or a prototype) not yet available in the public domain, while the rest employed use of an existing, often popular, app.

2.5. Results and Findings

Our review of the literature yielded various potential ethical issues with gamified health apps. Table 2 gives an overview of these issues along with sources citing such issues. While describing these issues, we also note recommendations, within the identified literature, to designers of gamified health apps of ways in which they may potentially address these issues. We then map these recommendations onto our tripartite framework based on the type of designer responsibility a given ethical issue may be addressed by.

In Table 3, we encapsulate how the recommendations in the literature for designers of gamified health apps, to address these ethical issues, can be mapped onto our proposed categories. We indicate which type of designer responsibility may potentially address that particular ethical issue. It should be noted that while the table only includes recommendations in the literature, these are by no means an exhaustive set of recommendations to designers of gamified health apps related to addressing potential ethical issues. In the text below, as examples of further possible steps designers, we also note some additional observations of our own. For example, although Maturo & Setiffi (2016) analyze and introduce the issues of biosociality,

amorality, the neoliberal objection, they do not offer an explicit recommendation to address these issues. Their analysis, however, can be used to deduce some possible steps and we note them below in the text. Further, our recommendations are also not meant to be exhaustive and there are of course, other steps designers could take to address some of these ethical issues. For example, we indicate that privacy-related issues may be addressed by proper design and proper use as we have noted such possibilities in the analysis presented above. This should not be taken to mean that there aren't ways to address such privacy issues through means that maybe characterized as belonging to the third category of proper embedding in the social system. Table 3 and our analysis in general indicate the scope for future research – particularly, to investigate other ways in which designers of gamified health apps may address potential ethical issues by assuming one of the types of responsibilities in our framework.

Table 2.

Reported Ethical issue	**Sources**
Privacy-related Issues	(Orji et al., 2017) (El-Hilly et al., 2016) (Barratt, 2017) (Hopia & Raitio, 2016; Spillers & Asimakopoulos, 2014; Trang & Weiger, 2021)
Cognitive Manipulation	Maturo & Setiffi, (2016)
Dependence and Addiction	(Attig & Franke, 2019) (Attig & Franke, 2020) (Hopia & Raitio, 2016) (Barratt, 2017) (Rockmann, 2019; Whelan & Clohessy, 2020)
Psychological harms	(Orji et al., 2017) (Barratt, 2017) (Giannakis et al., 2013) (Cafazzo et al., 2012) (Dithmer et al., 2015; Gal-Oz & Zuckerman, 2015; Honary et al., 2019; Kerner & Goodyear, 2017; W. R. Smith & Treem, 2017; Whelan & Clohessy, 2020) (A. Rieder et al., 2020)
The Neoliberal Objection	Maturo & Setiffi, (2016)
Physical Harms	(Barratt, 2017) (Lai et al., 2019)
Hermeneutic Problems	Maturo & Setiffi, (2016), (Lupton & Thomas, 2015)

Reported Ethical issue	Sources
Biosociality	Maturo & Setiffi, (2016)
Amorality	Maturo & Setiffi, (2016)
Issues related to providers and facilitators	van Dooren et al., (2019)

1. Privacy-related issues – Privacy was a chief concern among many users of gamified health apps. We found multiple studies that reported users being concerned about lack of privacy when using a gamified health app. This concern was either a result of users not comfortable with their data being tracked or shared, or because they were unsure how their data may be used by the app. Users also expressed concern with certain features of the app, intentionally designed, to lure them into using the app more or reminding them to use it. There was also evidence that some apps were intentionally designed to lure users into sharing more personal data (Trang & Weiger, 2021). There was also evidence of privacy concerns of the users translating into psychological concerns, such as feelings of being surveilled and corresponding anxiety. This clearly points to the need for designer assuming responsibilities to protect user privacy. Orji et al., (2017), for example, suggest that app designers should allow users to hide their identity and other personal information from other users of the app. They also suggest other "personalization" features to allow users to choose what information is shared and collected about them. (Trang & Weiger, 2021) suggest that app providers should explicitly ask user's permission before processing private information as well as inform users as much as possible about ways in which their information is used.

2. Cognitive manipulation - In their review of multiple gamified health apps, Maturo & Setiffi (2016) write of apps exploiting concepts from cognitive psychology to manipulate users into using the app or oversharing information on them. Such design features are also partly responsible for the addicting nature of such apps, and Attig and Franke (2019) have done an important study demonstrating the dependence of users on gamified health apps. (Attig & Franke, 2019) write that such features rarely lead users into adopting an active lifestyle (or exercise) in the long-run and designers should instead focus on facilitating internal motivation of the users.

3. Dependence and addiction – Besides (Attig & Franke, 2019), Barratt (2017), in his qualitative study on the use of gamified apps by cyclists, also found evidence of such dependence and addiction to the apps. Barratt also reported that some users also found their autonomy constrained as they did not expect they would be so easily lured into the game rewards and incentives, such that they would complete the game challenges sometimes at the expense of other important personal and social commitments. At least some of these effects, or at least to this extent, may be unforeseeable or unintended from the designers. It is hard to say from the available evidence the extent to which issues such as addiction or extreme dependence on the app are always solely a result of design features and not unhealthy ways of engaging with the app on part of the user. As mentioned earlier, Attig and Franke write that app dependence rarely translates into user's adopting a healthy lifestyle in the long-run and designers are better off aiming for internal motivation for users. They suggest that apps allow for self-determination and self-rewarding for the users. Some of this may also be rectified by designers sharing or transferring responsibility (of proper use) to users. Yet, in so far as these issues are foreseeable, some or significant responsibility also lies with the designer of the app, depending on the circumstances and game elements of the app.

4. Psychological harms - A similar case exists for design features that potentially lead to psychological harms to the users other than dependence or obsession with game rewards. These include, as stated earlier, feeling of being surveilled, and not feeling under-control (lack of perceived autonomy). Some users also experienced extreme psychological states (such as anger or anxiety) because of the gamified health app. This could be caused sometimes by lagging in the competition (or not having enough game rewards) or also when users suspected others of cheating (Gal-Oz & Zuckerman, 2015). Some design features seem to be responsible for incentivizing users to cheat, although part of the responsibility, again, lies with the users as well. A more concerning psychological aspect of gamified health apps seems to be their detrimental effects on existing internal motivations as well as the confidence of the users. (Dithmer et al., 2015), for example, point out that some users can be left with a strong sense of defeat, and it is, therefore, very important that game elements are designed to avoid such scenarios particularly in serious contexts such as gamified systems for improving heart activity. There is definitely a case to be

made for designers to review such cases and ensure that design features minimize the occurrence of such negative effects as much as possible. Besides the moral implications of such negative effects on users, evidence also suggests that it has adverse effects on user engagement with the app and leads to discontinuance (Rockmann, 2019). Recommendations within the literature include – giving users more autonomy and personalization of app features (Orji et al., 2017), allowing cheating to a limited extent (for example by allowing users more autonomy over how their results are displayed and building an app community that is tolerant of individual users making such choices in order to save "face") (Gal-Oz & Zuckerman, 2015), avoiding giving users a sense of defeat in serious apps (Dithmer et al., 2015). Physical harms - Gamified health apps use game elements to motivate users to increase physical activity in their lives. However, for some users, this may result into side effects such that they may overexert themselves or engage with the app in ways that are harmful to them. The most obvious evidence of physical harm was through reports of users overtraining or overstressing themselves in search of game rewards (Barratt, 2017). At least some of these harms may be reduced through use plans and other strategies designers may employ to transfer responsibility for proper use to the users. As discussed, there may be other ways of fostering virtuous use of the apps amongst users by facilitating forums and other places where users may learn from each other how they can best engage with the app.

5. Hermeneutic problems- Designers and design features also seem to be directly responsible for various "hermeneutic" problems posed by gamified health apps. This problem relates to the use of terms within the app that may reinforce stereotypes. Lupton & Thomas (2015), for example, write of gamified pregnancy apps which represent pregnant women in stereotypical ways such as a Barbie doll.

6. A related concern is that gamified health apps atomistically insulate individuals from other individuals while simultaneously being widely socially connected through a potential network of app users. This insulation of users brackets out the social determinants/dimension of health in a sort of hermeneutic reductionism (Maturo & Setiffi, 2016). This hermeneutic reduction can lead to users feeling pressured or compelled to log and look for only particular types of data, potentially at the expense of what they may have found meaningful,

or motivational. For example, by only providing functionality to record performance metrics (such as distance and duration of a run), and rewarding based on these particular metrics, the Nike+ system implicitly communicates that other (meaningful//enjoyable) aspects of running, such as the runner's high, or the mindful interaction between human and environment, are less important (Cheng,2020). Additionally, the proxies used in gamification elements can come to represent definite truths about what they are gamifying, as well as become privileged over other ways of knowing. This points to the problem of gamified health apps not properly embedded within a larger structural context.

7. Biosociality - The problem entails that certain gamified apps may reinforce physical stereotypes and also force the formation of groups on such physical attributes (Maturo & Setiffi, 2016). Designer efforts of fostering and encourage virtuous behaviour for proper use among the users may partly address this problem.

   The Neoliberal objection - As previously stated, one's health status and physical condition is heavily conditioned by education, income, and living conditions. Similarly, one's chances to engage in a healthy life can vary according to social and economic circumstances. Designers of gamified health apps should also be aware of the social dimensions and contexts within which their apps are used. It has been argued that the individualistic view underlying gamified health apps can lead to problems such as depoliticization of the role of the state, which reduces the responsibilities of the state for the health of its citizens and shifts the burden upon individuals. This objection states that such apps foster a neo-liberal ideology that implicitly stigmatizes people who are not capable of meeting the standard definition of 'healthy' (Maturo & Setiffi, 2016). Through a discourse analysis of major gamified health apps, Maturo and Setiffi (2016) point out how the design and linguistic features of such apps may lead to such stigmatization. While designers can "fix" some of these linguistic issues, more holistic solution to such problems perhaps lies in more active engagement of the designers with other actors and stakeholders in the society. This could enable a more successful integration and embedding of gamified health apps within a larger structural quest to promote healthy lifestyle and outcomes for the citizens.

8. Amorality - Another detrimental effect of gamified apps is that they may lead to/incentivize users to choose goals that are potentially harmful without caveats. This issue may be characterized as one of individual users choosing the wrong kind of incentive within a game and it may be partly addressed by both: better design features and virtuous user engagement with the app. Yet, as Maturo and Setiffi (2016) point out, this can be more than an individual issue and one where social norms may play a part. For example, dieting apps may lead a user to choose goals that other users are accomplishing or people around them find healthy, rather than what may actually be healthy for the individual.

9. Issues related to providers and facilitators in specific contexts - Finally, there is also evidence of there being merit in designers engaging with providers or facilitators of gamified health apps in particular contexts such as doctors and physicians. Writing about the use of gamified health app in the context of therapy (for mental well-being), (van Dooren et al., 2019), for example, write how therapists could benefit from having more control (and hence, responsibility) on the features of the app and that this could be done through direct interactions between the designers of such an app and the therapist planning an intervention which uses the app.

Table 3. Recommendations in the literature to app designers to address ethical issues

Reported Effect	Proper Design	Proper Use	Proper Embedding in the Social System
Privacy-related Issues	Personalization to allow users to choose what information they want to share. Explicit notification and seeking user permission before processing private information.	Inform users explicitly about how their private data will be used.	

Reported Effect	Proper Design	Proper Use	Proper Embedding in the Social System
Cognitive Manipulation	Avoiding manipulative features.	Providing warning and safety restrictions against harmful cognitive effects.	
Dependence and Addiction	Allowing users to be more self-determined and self-rewarding in the usage of the app rather than offering extrinsic pre-determined rewards.	Giving users more control over reward features.	
Psychological harms	Avoiding psychological harms such as a sense of defeat in serious apps, Giving users more autonomy in choosing their goals. Tolerating some level of "cheating" from users if that translates to better health choices.	An empathetic approach to design that allows users to be autonomous and some self-determination over their goals as well as how they might be displayed.	Facilitating a community of users who are empathetic to other users of the app.
Physical Harms	Allowing users to choose their own goals.	Improving user autonomy as well as giving warnings about dangers of overuse and exertion.	

Reported Effect	Proper Design	Proper Use	Proper Embedding in the Social System
Hermeneutic Problems	Avoiding game elements/rewards/terms. that may reinforce harmful stereotypes.		Avoiding reductionism in rewarding systems/game elements, for example, in ways that may incentivizes users to interpret their health and lifestyle in potentially harmful terms and/or though narrowly conceived metrics.
Issues related to providers and facilitators			Engage with providers and facilitators of gamified health apps such that the apps cater to appropriate contextual information.

3. Discussion and recommendations for future research

Our first task in this chapter was to analyze and revise the framework offered by Kim and Werbach to identify ethical issues in gamification. Based on a theoretical analysis of this framework and arguments from moral theory, we argued for a revised practice-relevant theoretical framework that suggests three broad categories of responsibilities designers have in addressing ethical issues in gamified health apps. We have argued that the categorization, and the framework encapsulating this categorization, we propose is better equipped to help practitioners, such as designers, in the specific context of gamified health and fitness apps. This categorization also served as a guideline to identify and map ethical issues that have been discussed in the empirical literature on gamified health apps. We presented these in Table 3 in section 2. We want to emphasize that, theoretically, there are more issues as well as steps designers could take to address those issues, within the 3 categories in Table 3, that we did not find in the empirical literature on gamified health apps. While this may partly be because of the limitations of design and methodology of our study, it also points to a space for further research that looks for evidential proof of other issues (and corresponding steps to address them) that one can anticipate based on other literature on games and gamification. For example, there is a possibility that gamified health apps may lead to a trivialization of health, as an unintended effect of game elements that try to simplify complex health variables for users of the app. Nguyen (2020) has theorized a similar possibility arguing that one possible effect of gamified health apps might be a simplification of user's health goals. For example, a user may get so obsessed with numbers or rewards on their gamified health app, that they may lose track of their original goal of being healthy. Similarly, one may possibly observe other ethically problematic effects, other than we found in the empirical nature, because of the use of gamified apps. Zuboff (2019), for example, has argued that many digital environments, including on health-related apps, commodify user behavior in the interests of designers of these environments. Our review indicates the possibility for empirical investigation of such hypothesis in the context of gamified health apps in future work.

It should be noted that our attempt in this chapter was to locate as many ethical issues related to gamified health apps as we could find in the empirical literature. One limitation to note here is that there are related areas of research, such as research these focusing on behavior change technologies in general, which also deal with some similar ethical issues. Future research may seek to broaden the scope of our research

to include ethical issues identified from those domains as well. Another related limitation is that some of the ethical issues may need more or stronger evidence, particularly about the extent to which they are universally or even widely operational across gamified health apps. For example, we stated that Maturo and Setifi's (2016) work on gamified health apps, which is based on analysis of the design features of the apps, as well as a discourse analysis of reviews of the apps on various internet forums, notes that such apps may lead to stigmatization of people who may not be able to meet standard definitions of "healthy". Yet, it is also a possibility that, in practice, user communities (facilitated by forums and groups, for example) may subvert the affordances of such apps, and negate such tendencies of stigmatization. The evidence for such optimism comes from other works on self-tracking (not necessarily gamified) health apps. Sharon & Zandbergen (2017), for example, through their study of self-tracking communities, elucidate how theoretically postulated ideas about self-trackers, such as their engagement in a form of "data-fetishism" are limiting. They assert that instead of being obsessed with narrow notions of objectivity (an idea encapsulated within the data-fetishism charge on self-trackers), self-tracking communities actually attribute meaning to their quantified data in ways that resist such objectivity. Further, self-trackers also use their practice to resist social norms (instead of reinforcing them) as well as invent imaginative ways of using self-tracking as a narrative aid (Sharon & Zandbergen, 2017). Their study points to the need for further ethnographic and anthropological studies of self-trackers, as well as users of gamified health apps, to understand how such users and communities of users may resist theoretically anticipated problems with such apps. This is not to say that the theoretically postulated and anticipated sources of ethical issues with gamified health apps are not of value. Even if they are eventually resisted by users of such apps, theoretical critiques may offer themselves as a source of critical reflection on behalf of the users as well as designers of such apps. Such users and designers may then use the analysis offered in these theoretical critiques to find ways of resisting and escaping the anticipated problems. This reiterates the need for sharing of responsibilities between various stakeholders related with gamified health apps.

This brings us to a final point about the utility of the analysis offered in this chapter. Our aim has been to provide a practice-relevant framework to identify different types of designer responsibilities that can address ethical issues in gamification. Further, we also aimed at providing a landscape of various ethical issues related to gamified health apps as found in the empirical literature about such apps. This framework based on

designer responsibility as well as the list of various issues can be useful for designers, users of gamified health apps, as well as other stakeholders to anticipate as well as avoid ethical issues in their interaction with such apps. Given the recent evidence suggesting that ethical issues, such as potential psychological harm to app users (Rockmann, 2019), can lead to app discontinuance, addressing such issues may also serve to improve the long-term user engagement with such products. The designers of such apps, in particular, can use our framework to foresee possible issues as well as plan validation studies to ensure that the apps they design are able to avoid foreseeable ethical problems as well as problems that may arise as unintended and unforeseeable effects of their design features. Designer duties prescribed by our framework also emphasize the need for designers to reflect more broadly over the socio-economic implications of the technologies they seek to introduce and point to the potential utility of seeking more democratized approaches toward technological design.

Concluding remarks and some directions for future research

Datafication of health comes with significant promises. Enabled by the rise of consumer-oriented digital health technologies – such as wearable devices, voice assistants, and smartphone apps – health datafication has the potential to dramatically expand and democratize the generation of, and access to, one's own health-related information. Users within this datafied health environment have the opportunity to benefit on both - an individual and a collective level. On an individual level, health datafication, and the use of corresponding technologies, has the potential to empower users by providing them real-time access to their health indicators that would otherwise have been opaque to them. A patient managing a chronic condition, for example, can potentially monitor their vitals on their own, rather than making a trip to the local clinic every time they wanted to measure them. Individuals can also focus on personalized health targets, which they can receive as recommendations from their doctor, for example, and track their progress towards it. On a collective level, health datafication offers opportunities to help local and global health bodies to offer and improve services such as infectious disease surveillance. Analysis of data collected in real-time and in real-world settings also has the potential to offer new and better ways of diagnosing diseases.

Successful fulfillment of such promises, however, depends on the work done by various actors involved in the different phases along the health data value chain – data collection and storage, processing and analysis, and usage of drawn insights. The aim in this thesis has been to analyze and substantiate the nature of responsibility of such actors vis-à-vis their role in promoting the benefits of health datafication as well as in guarding against potential pitfalls or threats that come along with the datafication of health.

Chapter 1 of the thesis discussed one substantial potential threat to the participants of the datafied health environment – harms (and risks of harms) related to loss of privacy. Health datafication amplifies the risks associated with privacy loss in the healthcare context for at least two reasons. First, the datafication paradigm incentivizes the collection of data in large, and increasing, volumes. The logic here is that the more the data, the better the insights. Second, the nature of the data involved leaves data

subjects in an extremely vulnerable position, opening the door for many sorts of harms – such as discrimination in employment, loss of reputation, psychological harms, etc. This latter point, about the relatively greater sensitive nature of health data, has been recognized in privacy legislation in many countries through a sort of "exceptionalism" or special status given to health data.

In this chapter, I discuss how despite recognizing the potential for serious risks and harms associated with loss of privacy in health datafication, current legislations, that take a "transparency and consent" approach, fail to close the door for exploitation and manipulation of data subjects from data controllers (that is those who collect and process the collected data). One of the major limitations of this "transparency and consent" approach is that the nature of "disclosed" information about how and for what purposes the collected data can be used is too long and complex to provide an opportunity for informed consent to the data subjects in a practical sense. While proscriptive conditions within legislations can plug this limitation to some extent and limit or prohibit some specified forms of data subject exploitation, this approach faces challenges on its own in the form of the unpredictable nature of technological advancement and legal opportunism from data controllers. For example, legislation trying to prevent loss of privacy by defining conditions for data anonymization (such as removal of data that may reveal specific information about data subjects), can be bypassed by technological advances in the ability to combine, and infer protected information from, multiple datasets as well as through exploitation of legal loopholes in such legislation.

My proposal to guard against such opportunism from health data controllers, and to protect data subjects against risks and harms related to loss of privacy, is an imposition of fiduciary duties on health data controllers. Fiduciary duties can be understood as duties of loyalty, such that the fiduciary, who takes on such duties, has the responsibility to cater to the interests of the beneficiary. As a legal instrument, fiduciary duties have a precedent to be imposed in other situations where two parties have an asymmetric relationship such that one is relatively vulnerable vis-à-vis the other. My argument in this chapter is that fiduciary duties are better suited than other legislative approaches in establishing the responsibilities of health data controllers towards data subjects, particularly in relation to risks associated with loss of privacy. Besides prohibiting opportunistic behaviour, a fiduciary approach to legislation also incentivizes a contextual approach to protection of privacy, by requiring (and

enabling) health data controllers to seek the most appropriate methods for protecting user privacy, rather than rely on specified recommendations or prescriptions.

At the end of the chapter, I discussed some potential practical limitations of the fiduciary duty approach to privacy legislation and the need for future research to address such limitations. One such limitation stems from the global nature of information flow, and the necessity of various national and international legislations to be in harmony with each other. The complex nature of the task involved in synchronizing any legislation with existing ones and specifying the exact contours of a fiduciary law within a specific country belongs to, and is part of the discussion in current, legal scholarship. Since the publication of the chapter, some scholarship has engaged with this task. Within the context of the United States legislation, for example, Khan & Pozen (2019) raise skepticism regarding the applicability of fiduciary duty to data controllers (or similar entities controlling digital information) as it would generate conflicting duties for corporations in this position, who also have duties of loyalty to their shareholders. Yet, as others, including Gold (2019) and Richards & Hartzog (2021), point out, fiduciary law within the American legislation has surpassed such obstacles by, for example, developing and defining "a hierarchy of obligations" to address such conflicts. The debate, however, further highlights a general point about the need for future legal scholarship to address the specifics of applying a fiduciary approach to data governance within specific legislative regimes (see Prasad M & Menon C (2020), for another such example with a discussion of specifying a fiduciary based data protection law in India).

In chapter 2, I discuss another form of asymmetry created and/or amplified by health datafication – asymmetry between traditional centers and institutions for healthcare research and big technology firms, such as Alphabet (Google), Apple, Amazon, etc. who have recently been allocating substantial resources to data-driven healthcare research. In this chapter, I discussed what has been termed by Tamar Sharon (2016) as the phenomenon of "Googlization of Health Research" (GHR). GHR is characterized by the potential superiority of big tech firms in terms of technical, financial, and human resources deployed for data-driven healthcare research. This superiority potentially enables big tech firms to take a leading role in healthcare research, and subsequently, in influencing policy decisions regarding public health interventions, for example.

I highlight in this chapter how this development (of GHR) is worrisome, particularly considering the nature of epistemic trust required for compliance with and legitimacy of public health interventions and policy decisions that may be based on the epistemic goods produced by GHR. Specifically, I have argued that epistemic trust – that is, trust in the form of a disposition to believe someone's claim *p* on the assumption that they are in a position to make such claim and are doing so truthfully – is rationally grounded in social, moral and institutional indicators of trustworthiness of those making the claims. In the context of GHR, this implies that social and moral transgressions of big tech corporations, as well as the lack of transparent engagement with the public regarding their use of public data, can have a significantly deleterious effect on epistemic trustworthiness in scientific claims produced within GHR. Examples of such transgressions and lack of transparency by big tech firms include the development of mobile-based health apps that engage in manipulation and "hypernudging" of users (Lanzing, 2019), and multiple secretive (held from the public view) contractual data sharing and research agreements between big tech firms and public health bodies (see (Powles & Hodson, 2017) for details on one such agreement between DeepMind, an Alphabet subsidiary, and the UK's National Health Service (NHS)).

The aim in this chapter has also been to highlight that while problems such as opacity and bias are often touted as the main problems associated with data-driven research, and therefore, as the main reasons to withhold epistemic trust in such research, the social, moral, and institutional indicators discussed above may, in fact, be more important from the perspective of warranted public epistemic trust in scientific research. Here, by "institutional factors" I mean reputational factors, such as opinions of other experts on the expert making the claim *p*. These institutional factors are particularly important for public epistemic trust as many members of the public are often not in an independent position to judge expert claims.

Going forward, it is imperative to pay more attention to such institutional or reputational factors, and how they can be enhanced to help the public make rational decisions regarding holding or withholding epistemic trust in scientific research produced by GHR. One step towards developing and enhancing such institutional factors would be to pay heed to calls for greater transparency of algorithms used by big tech firms, at least for other researchers such that this transparency does not harm the commercial interests of such organizations to the extent that such interests support

larger public interests and/or innovation, for example (Hegelich, 2020; Montag et al., 2021; B. Rieder & Hofmann, 2020).

Another important step in developing institutional indicators for rational epistemic trust in GHR would be to develop public oversight mechanisms and frameworks for public health institutions in their engagement with private and/or commercial centers of research such as the big tech firms. This could, for example, help avoid the situation of secretive agreements between public health institutions and big tech firms as in the case of the partnership between DeepMind and NHS. An example of such a framework is the one developed by the Swiss Personalized Health Network (SPHN), which lays down principles for responsible data processing, including by private actors who may avail services and funding from the Swiss government (Meier-Abt et al., 2018).

Finally, much work needs to be done to fully understand the political implications of GHR, particularly regarding the role big tech firms should and could play in agenda-setting for future healthcare research. Scholarship on the experience with monopolies established by pharmaceutical companies in drug development and other agenda-setting influences could be fruitful here. Similarly, scholarship within social science disciplines, such as critical data studies, that has offered a political critique of developments in the big data economy can also give valuable insights into defining the responsibilities and roles of commercial firms as well as public institutions regarding data-driven healthcare research.

In chapter 3, I discussed the threat posed by the potential to be misinformed or forming the wrong set of epistemic expectations in the use of technologies such as voice assistants for health information. Digital voice assistants, such as Amazon's Alexa, are becoming an important part of many people's epistemic practices, particularly related to health. According to Google's own estimates, one in 20 searches seek health-related information (Google, 2015). Further, some of these voice assistants have anthropocentric characteristics, such as a human-like voice in delivering the information to the user. This phenomenological and functional similarity of machine-generated speech has led to some philosophers claiming that such machine-generated speech should also be classified as "assertions".

Philosophers generally refer to the speech of assertion as denoting the phenomena of making claims, stating, reporting, or affirming something to be the case. Prima facie then, there appears to be a case for machine-generated speech to be classified as assertions, when such machines, at least appear to be making claims about or reporting something to be the case, particularly when such speech mirrors human speech. Yet, as I have argued in this chapter, what is at stake in claims about assertions – that is, that someone has asserted something to be the case – is the issue of responsibility. That is, to say that someone has asserted something to be the case, is to claim that the asserter takes responsibility for the claim to be true (or up to the expected epistemic standards in that context). Further, such responsibility is at least epistemic, but could also be characterized as an ethical responsibility, such that the asserter is liable to moral blame if the claim turns out to be false.

Machines such as digital voice assistants like Alexa, do not seem to be the sort of agents that can undertake such epistemic and/or ethical responsibility. There is, however, a possibility to characterize some machine-generated speech as "proxy assertions", such that such speech is an assertion made by the machine on behalf of its designers, or other sets of actors, who can undertake the requisite responsibility for the asserted speech. In such a scenario, however, I have argued that it is imperative that designers can actually and practically take responsibility for such speech, which in turn, requires that the designers can reasonably foresee the deliverance of such machine-generated speech.

One of the implications of the view I have argued for is that designers should make it transparent to the users of technologies such as Alexa, the kind of machine-generated speech that they can and cannot reasonably foresee and take responsibility for. In other words, it should be transparent to the users, which machine-generated speech may or may not qualify as a proxy assertion. What is at stake, again, is the kind of expectations users can legitimately have regarding such machine-generated speech. Transparency regarding which machine-generated speech designers can take responsibility for can help users form the right set of expectations.

In terms of directions of future research indicated by my argument, one important line of inquiry is to explore how can designers effectively communicate to users the kind of expectations they should form for which sort of machine-generated speech. One possibility I discussed in the chapter is to avoid the use of human-like voice in

instances where machine-generated speech is unlikely to qualify as being an assertion, or in other words, where it is unlikely that designers can actually and practically take responsibility for such speech. This possibility is suggested by some recent research. For example, Schreuter et al., (2021) and (Moore, 2017) have written about how human-like voice can give rise to inappropriate epistemic expectations among users. Schreuter et al., (2021) also point out that voice, in general, compared to texts, for example, tends to raise user expectations about the accuracy of machine-generated claims. Further empirical research may clarify whether this is true for all contexts, including health, and what strategies can be adopted to match user expectations with the level of accuracy designers of such machines can promise or take responsibility for.

In the final chapter of the thesis, Maryam Razavian and I discussed the responsibilities of designers of gamified health and fitness apps. Such apps use game-like elements, such as scores, reward points, badges, etc., to motivate users into using the app more. This motivation, in turn, is meant to be a proxy for motivating the user into engaging in greater physical activity, for which the user is seemingly being rewarded. As we discuss in the chapter, such apps come with significant "side-effects" or a "dark-side" as some scholars have put it. Such negative side-effects include, for example, increased stress and/or anxiety among the users of the app, partially as a result of caring too much for the game elements of the apps (such as scores or badges). Such apps may also lead to a sort of reductionism such that the user may tend to privilege the narrowly construed quantified proxy (such as the score for the distance they cycled for) over and above other rich ways of understanding their health status.

The hypothesis that gamified apps may cause a sort of trivialization of health is an interesting one and needs further empirical research. In recent work, Nguyen (2020) has offered a theoretical mechanism for how such trivialization may occur through what he terms "value capture". Roughly speaking, value capture is the idea that game-like elements, such as those used in gamified health apps like a Fitbit, that are just simplified proxies for rich phenomena, come to replace such rich phenomena in the minds of the game player. According to Nguyen, value capture is not limited to gamified health apps, but may in fact be far more pervasive and can be found in digital environments such as Twitter, as well as offline environments like classrooms where grades act as simplified proxies for understanding and other rich goals of education (Nguyen, 2021).

The idea of value capture raises at least two questions in need of further exploration in the context of gamified health apps. First, to what extent is value capture the result of gamification versus quantification? In other words, would the effects be the same if the users were not made to artificially care about the quantified simplified proxies for their health, if the game-like elements were removed? Second, what strategies could designers of such apps take to keep the benefits of gamified health apps, such as positive effects on motivation, while avoiding the pitfalls, such as value capture? In other words, how can designers facilitate a rich understanding of health among the users of the app, while still finding ways of keeping the user motivated enough to keep engaging in healthy behaviour. As Nguyen points out, this is a difficult question as gamification does make it significantly easier for the users to care about a simplified goal.

One way in which users of gamified or quantified health apps may be able to resist the pitfalls associated with value capture or reductionism is by forming communities built around the use of the app and engage in deep collective reflection on the use and the effects of the use of such apps. Sharon & Zandbergen (2017), for example, have written about how communities of users engaged in the "quantified-self" movement resist being obsessed with narrow notions of objectivity and "data-fetishism" through engaging in such collective reflective exercises. Further ethnographic and anthropological studies around communities of gamified health apps may perhaps also reveal such tendencies and even suggest ways for designers of the apps to facilitate the formation of such reflective collectives.

Finally, while the attempt in the thesis has been to identify and analyze the responsibilities of various actors involved in the datafication of health, it is not an exhaustive attempt, and many other actors, as well as some aspects of responsibilities of the actors discussed, have not been explored within the scope of this thesis. To take one example - medical practitioners, clinicians, hospital administrators, doctors, etc. who have an important role in facilitating (or perhaps resisting) integration of traditional health services within a datafied health environment. Future research should explore the roles and responsibilities of such actors in a datafied health environment. Ethnographic and anthropological research about such actors may also reveal how such professions and responsibilities encapsulated within such professional roles are already being transformed as a result of health datafication.

As the last word on another set of actors who do (and will do) extremely important work within the datafied health environment but found themselves outside the scope of detailed discussion in this thesis, I want to mention data curators. These are people who curate, annotate and validate medical and biological databases that can, and are, then used for, for example, medical and biological research (Blasimme et al., 2018; Leonelli, 2015, 2019). Despite the high significance of the work done by curators, including data labeling, which determines the quality and usability of databases, data curation faces the problem of attribution (lack of due recognition and rewards for curation work). The problem of attribution, in turn, disincentives curation work, and is an important problem in need of a solution (Blasimme et al., 2018; Leonelli, 2019). The importance of curation work is perhaps best captured by Leonelli's point about the processes underlying data curation (2015):

> "unraveling the conditions under which data are created and disseminated is crucial to understanding what counts as knowledge in the first place, and for whom, and to assessing the epistemic value of the various outputs of knowledge-making activities, whether they be claims, data, models, theories, instruments, communities, and/or institutions."

The work of curators, along with actors involved in the health datafication, then, is at the heart of the most important questions related to health datafication, which this thesis has attempted to engage in: What is the best way to collect, share, and interpret health data and how should roles and responsibilities be allocated within such processes?

References

Abelson, H., Anderson, R., Bellovin, S. M., Benaloh, J., Blaze, M., Diffie, W., Gilmore, J., Green, M., Landau, S., Neumann, P. G., Rivest, R. L., Schiller, J. I., Schneier, B., Specter, M. A., & Weitzner, D. J. (2015). Keys under doormats: Mandating insecurity by requiring government access to all data and communications. *Journal of Cybersecurity*, *1*(1), 69–79. https://doi.org/10.1093/cybsec/tyv009

Acquisti, A., Adjerid, I., & Brandimarte, L. (2013). Gone in 15 Seconds: The Limits of Privacy Transparency and Control. *IEEE Security Privacy*, *11*(4), 72–74. https://doi.org/10.1109/MSP.2013.86

Ada Lovelace Institute. (2020). *The data will see you now*. https://www.adalovelaceinstitute.org/report/the-data-will-see-you-now/

Adjerid, I., Acquisti, A., Brandimarte, L., & Loewenstein, G. (2013). Sleights of Privacy: Framing, Disclosures, and the Limits of Transparency. *Proceedings of the Ninth Symposium on Usable Privacy and Security*, 9:1–9:11. https://doi.org/10.1145/2501604.2501613

Andrejevic, M. (2014). Big Data, Big Questions| The Big Data Divide. *International Journal of Communication*, *8*(0), 0.

Arges, K., Assimes, T., Bajaj, V., Balu, S., Bashir, M. R., Beskow, L., Blanco, R., Califf, R., Campbell, P., Carin, L., Christian, V., Cousins, S., Das, M., Dockery, M., Douglas, P. S., Dunham, A., Eckstrand, J., Fleischmann, D., Ford, E., … Wong, C. A. (2020). The Project Baseline Health Study: A step towards a broader mission to map human health. *Npj Digital Medicine*, *3*(1), 1. https://doi.org/10.1038/s41746-020-0290-y

Arora, C. (2019). Digital health fiduciaries: Protecting user privacy when sharing health data. *Ethics and Information Technology*, *21*(3), 181–196. https://doi.org/10.1007/s10676-019-09499-x

Art. 5 GDPR – Principles relating to processing of personal data. (n.d.). *General Data Protection Regulation (GDPR)*. Retrieved December 13, 2018, from https://gdpr-info.eu/art-5-gdpr/

Art. 9 GDPR – Processing of special categories of personal data. (n.d.). *General Data Protection Regulation (GDPR)*. Retrieved May 14, 2022, from https://gdpr-info.eu/art-9-gdpr/

Art. 12 GDPR. (n.d.). *General Data Protection Regulation (GDPR)*. Retrieved February 6, 2018, from https://gdpr-info.eu/art-12-gdpr/

Art. 25 GDPR – Data protection by design and by default. (n.d.). *General Data Protection Regulation (GDPR)*. Retrieved December 19, 2018, from https://gdpr-info.eu/art-25-gdpr/

Article 29 Working Party. (2015). *ANNEX - health data in apps and devices*. http://webcache.googleusercontent.com/search?q=cache:MlBCtv-DN6gJ:ec.europa.eu/justice/data-protection/article-29/documentation/other-document/files/2015/20150205_letter_art29wp_ec_health_data_after_plenary_annex_en.pdf+&cd=1&hl=en&ct=clnk&gl=nl&client=firefox-b-ab

Athinaiou, M. (2017, July 17). *Why has healthcare become such a target for cyber-attackers?* The Conversation. http://theconversation.com/why-has-healthcare-become-such-a-target-for-cyber-attackers-80656

Attig, C., & Franke, T. (2019). I track, therefore I walk – Exploring the motivational costs of wearing activity trackers in actual users. *International Journal of Human-Computer Studies*, *127*, 211–224. https://doi.org/10.1016/j.ijhcs.2018.04.007

Attig, C., & Franke, T. (2020). Abandonment of personal quantification: A review and empirical study investigating reasons for wearable activity tracking attrition. *Computers in Human Behavior*, *102*, 223–237. https://doi.org/10.1016/j.chb.2019.08.025

Austin, J. L. (1975). *How to Do Things with Words*. Clarendon Press.

Bach, K., & Harnish, R. M. (1979). *Linguistic Communication and Speech Acts*. Cambridge: MIT Press.

Balkin, J. (2014, May 3). *Balkinization: Information Fiduciaries in the Digital Age*. https://balkin.blogspot.nl/2014/03/information-fiduciaries-in-digital-age.html

Balkin, J. M. (2015). Information Fiduciaries and the First Amendment. *U.C. Davis Law Review*, *49*, 1183.

Ball, B. (2014). Speech Acts: Natural or Normative Kinds? The Case of Assertion. *Mind & Language*, *29*(3), 336–350. https://doi.org/10.1111/mila.12054

Barocas, S., & Nissenbaum, H. (2009). *On Notice: The Trouble with Notice and Consent* (SSRN Scholarly Paper ID 2567409). Social Science Research Network. https://papers.ssrn.com/abstract=2567409

Baron, K. G., Abbott, S., Jao, N., Manalo, N., & Mullen, R. (2017). Orthosomnia: Are Some Patients Taking the Quantified Self Too Far? *Journal of Clinical Sleep Medicine*, *13*(02), 351–354. https://doi.org/10.5664/jcsm.6472

Barr, A. (2014, July 27). Google's New Moonshot Project: The Human Body. *Wall Street Journal*. https://www.wsj.com/articles/google-to-collect-data-to-define-healthy-human-1406246214

Barratt, P. (2017). Healthy competition: A qualitative study investigating persuasive technologies and the gamification of cycling. *Health & Place*, *46*, 328–336. https://doi.org/10.1016/j.healthplace.2016.09.009

Barrotta, P., & Gronda, R. (2020). Epistemic Inequality and the Grounds of Trust in Scientific Experts. In A. Fabris (Ed.), *Trust* (pp. 81–94). Springer International Publishing. https://doi.org/10.1007/978-3-030-44018-3_6

BBC. (2017, July 3). Google DeepMind NHS app test broke UK privacy law. *BBC News*. https://www.bbc.com/news/technology-40483202

Blasimme, A., Fadda, M., Schneider, M., & Vayena, E. (2018). Data Sharing For Precision Medicine: Policy Lessons And Future Directions. *Health Affairs*, *37*(5), 702–709. https://doi.org/10.1377/hlthaff.2017.1558

Bogost, I. (2015). WHY GAMIFICATION IS BULLSHIT 2. *The Gameful World: Approaches, Issues, Applications*, 65.

Brennan-Marquez, K. (2015). Fourth Amendment Fiduciaries. *Fordham Law Review*, *84*, 611.

Brinig, M. F. (2011). *Parents, Trusted But Not Trustees or (Foster) Parents as Fiduciaries* (SSRN Scholarly Paper ID 1767412). Social Science Research Network. https://papers.ssrn.com/abstract=1767412

Bu-Pasha, S. (2017). Cross-border issues under EU data protection law with regards to personal data protection. *Information & Communications Technology Law*, *26*(3), 213–228. https://doi.org/10.1080/13600834.2017.1330740

Burlina, P., Joshi, N., Paul, W., Pacheco, K. D., & Bressler, N. M. (2020). Addressing Artificial Intelligence Bias in Retinal Disease Diagnostics. *ArXiv:2004.13515 [Cs, Eess]*. http://arxiv.org/abs/2004.13515

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, *3*(1), 2053951715622512. https://doi.org/10.1177/2053951715622512

Byres, E. J., Franz, M., & Miller, D. (2004). The use of attack trees in assessing vulnerabilities in scada systems. *In IEEE Conf. International Infrastructure Survivability Workshop (IISW '04). Institute for Electrical and Electronics Engineers*.

Bywater, A., & Armstrong, J. (2015, March 6). *EU health data definition concerning lifestyle and wellbeing apps*. Cordery. http://www.corderycompliance.com/eu-health-data-definition-concerning-lifestyle-and-wellbeing-apps/

Cafazzo, J. A., Casselman, M., Hamming, N., Katzman, D. K., & Palmert, M. R. (2012). Design of an mHealth App for the Self-management of Adolescent Type 1 Diabetes: A Pilot Study. *Journal of Medical Internet Research*, *14*(3), e70. https://doi.org/10.2196/jmir.2058

Candeub, A. (2013). Transparency in the Administrative State. *Houston Law Review*, *51*, 385.

Cestui que trust. (2006). Collins Dictionary of Law. https://legal-dictionary.thefreedictionary.com/cestui+que+trust

Cheng, V. W. S., Davenport, T., Johnson, D., Vella, K., & Hickie, I. B. (2019). Gamification in Apps and Technologies for Improving Mental Health and Well-Being: Systematic Review. *JMIR Mental Health*, *6*(6), e13717. https://doi.org/10.2196/13717

Colizza, V., Grill, E., Mikolajczyk, R., Cattuto, C., Kucharski, A., Riley, S., Kendall, M., Lythgoe, K., Bonsall, D., Wymant, C., Abeler-Dörner, L., Ferretti, L., & Fraser, C. (2021). Time to evaluate COVID-19 contact-tracing apps. *Nature Medicine*, *27*(3), 3. https://doi.org/10.1038/s41591-021-01236-6

Crawford, K., Lingel, J., & Karppi, T. (2015). Our metrics, ourselves: A hundred years of self-tracking from the weight scale to the wrist wearable device. *European Journal of Cultural Studies*, *18*(4–5), 479–496. https://doi.org/10.1177/1367549415584857

Crouch, H. (2020, July 17). *NHS signs new four-month contract with private tech firm Palantir*. Digital Health. https://www.digitalhealth.net/2020/07/nhs-palantir-contract/

Cuneo, T. (2020, May 7). *Ethical Dimensions of Assertion*. The Oxford Handbook of Assertion. https://doi.org/10.1093/oxfordhb/9780190675233.013.38

Curry, E. (2016). The Big Data Value Chain: Definitions, Concepts, and Theoretical Approaches. In J. M. Cavanillas, E. Curry, & W. Wahlster (Eds.), *New Horizons for a Data-Driven Economy: A Roadmap for Usage and Exploitation of Big Data in Europe* (pp. 29–37). Springer International Publishing. https://doi.org/10.1007/978-3-319-21569-3_3

D | European Data Protection Supervisor. (n.d.). Retrieved May 13, 2022, from https://edps.europa.eu/data-protection/data-protection/glossary/d_en

Davies, B. (2021). 'Personal Health Surveillance': The Use of mHealth in Healthcare Responsibilisation. *Public Health Ethics*, *14*(3), 268–280. https://doi.org/10.1093/phe/phab013

de Melo-Martín, I. (2019). The commercialization of the biomedical sciences: (Mis)understanding bias. *History and Philosophy of the Life Sciences*, *41*(3), 34. https://doi.org/10.1007/s40656-019-0274-x

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining "gamification." *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, 9–15. https://doi.org/10.1145/2181037.2181040

Digital Trends. (2019, January 19). *Are Wearable Devices Leading to Over-diagnosis?* Digital Trends. https://www.digitaltrends.com/wearables/wearable-devices-leading-to-over-diagnosis/

Dillon, R. S. (2018). Respect. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2018). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2018/entries/respect/

Dithmer, M., Rasmussen, J. O., Grönvall, E., Spindler, H., Hansen, J., Nielsen, G., Sørensen, S. B., & Dinesen, B. (2015). "The Heart Game": Using Gamification as Part of a Telerehabilitation Program for Heart Patients. *Games for Health Journal*, *5*(1), 27–33. https://doi.org/10.1089/g4h.2015.0001

Dobbs, D. (2008). *Law of Torts (Hornbook Series)*. West Academic.

Douglas, H. (2000). Inductive Risk and Values in Science. *Philosophy of Science*, *67*(4), 559–579. https://doi.org/10.1086/392855

Douglas, H. (2017, March 27). *Why Inductive Risk Requires Values in Science*. Current Controversies in Values and Science. https://doi.org/10.4324/9781315639420-6

Drozd, S., & Dale, R. (2006). *General Principles of Medical Malpractice Litigation*. Lerners Lawyers. http://www.lerners.ca/lernx/general-principles-of-medical-malpractice-litigation/

Edwards, E. A., Lumsden, J., Rivas, C., Steed, L., Edwards, L. A., Thiyagarajan, A., Sohanpal, R., Caton, H., Griffiths, C. J., Munafo, M. R., Taylor, S., & Walton, R. T. (2016). Gamification for health promotion: Systematic review of behaviour change techniques in smartphone apps. *Bmj Open*, *6*(10), e012447. https://doi.org/10.1136/bmjopen-2016-012447

El-Hilly, A. A., Iqbal, S. S., Ahmed, M., Sherwani, Y., Muntasir, M., Siddiqui, S., Al-Fagih, Z., Usmani, O., & Eisingerich, A. B. (2016). Game On? Smoking Cessation Through the Gamification of mHealth: A Longitudinal Qualitative Study. *JMIR Serious Games*, *4*(2), e18. https://doi.org/10.2196/games.5678

Evans, B. J. (2011). Much Ado about Data Ownership. *Harvard Journal of Law & Technology*, *25*, 69.

Farrell, H. M. (2012). Transparency in psychiatric care. *Asian Journal of Psychiatry*, *5*(3), 273–274. https://doi.org/10.1016/j.ajp.2012.07.011

Fitbit Privacy Policy. (2016). https://www.fitbit.com/nl/legal/privacy

Fleisher, W., & Šešelja, D. (2021). *Responsibility for Collective Epistemic Harms* [Preprint]. http://philsci-archive.pitt.edu/19975/

Ford, R. A., & Price, W. N. I. (2016). Privacy and Accountability in Black-Box Medicine. *Michigan Telecommunications and Technology Law Review*, *23*, 1.

Frankel, T. T. (2010). *Fiduciary Law*. Oxford University Press.

Freiman, O., & Miller, B. (2020, May 7). *Can Artificial Entities Assert?* The Oxford Handbook of Assertion. https://doi.org/10.1093/oxfordhb/9780190675233.013.36

Fricker, E. (2015). HOW TO MAKE INVIDIOUS DISTINCTIONS AMONGST RELIABLE TESTIFIERS. *Episteme*, *12*(2), 173–202. https://doi.org/10.1017/epi.2015.6

Fuerstein, M. (2013). Epistemic Trust and Liberal Justification*. *Journal of Political Philosophy*, *21*(2), 179–199. https://doi.org/10.1111/j.1467-9760.2012.00415.x

Gabrielsen, A. M. (2020). Openness and trust in data-intensive science: The case of biocuration. *Medicine, Health Care and Philosophy*, *23*(3), 497–504. https://doi.org/10.1007/s11019-020-09960-5

Gal-Oz, A., & Zuckerman, O. (2015). Embracing Cheating in Gamified Fitness Applications. *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*, 535–540. https://doi.org/10.1145/2793107.2810298

GARCIA, M. (2016). RACIST IN THE MACHINE: THE DISTURBING IMPLICATIONS OF ALGORITHMIC BIAS. *World Policy Journal*, *33*(4), 111–117.

García-Carpintero, M. (2019). Conventions and Constitutive Norms. *Journal of Social Ontology*, *5*(1), 35–52. https://doi.org/10.1515/jso-2019-0013

Gargeya, R., & Leng, T. (2017). Automated Identification of Diabetic Retinopathy Using Deep Learning. *Ophthalmology*, *124*(7), 962–969. https://doi.org/10.1016/j.ophtha.2017.02.008

Gelter, M., & Helleringer, G. (2018). *Fiduciary Principles in European Civil Law Systems* (SSRN Scholarly Paper ID 3142202). Social Science Research Network. https://papers.ssrn.com/abstract=3142202

Giannakis, K., Chorianopoulos, K., & Jaccheri, L. (2013). User requirements for gamifying sports software. *Proceedings of the 3rd International Workshop on Games and Software Engineering: Engineering Computer Games to Enable Positive, Progressive Change*, 22–26.

Gibbs, S. (2014, November 18). Court sets legal precedent with evidence from Fitbit health tracker. *The Guardian*. https://www.theguardian.com/technology/2014/nov/18/court-accepts-data-fitbit-health-tracker

Gold, A. S. (2013). *The Loyalties of Fiduciary Law* (SSRN Scholarly Paper ID 2370598). Social Science Research Network. https://papers.ssrn.com/abstract=2370598

Gold, A. S. (2019, May 27). *The Fiduciary Duty of Loyalty*. The Oxford Handbook of Fiduciary Law. https://doi.org/10.1093/oxfordhb/9780190634100.013.20

Goldberg, S. (2015). *Assertion: On the Philosophical Significance of Assertoric Speech*. Oxford University Press.

Goldberg, S. C. (2012). Epistemic extendedness, testimony, and the epistemology of instrument-based belief. *Philosophical Explorations*, *15*(2), 181–197. https://doi.org/10.1080/13869795.2012.670719

Google. (2015, February 10). *A remedy for your health-related questions: Health info in the Knowledge Graph*. Google. https://blog.google/products/search/health-info-knowledge-graph/

Gostin, L. O., & Hodge, J. G. J. (2001). Personal Privacy and Common Goods: A Framework for Balancing under the National Health Information Privacy Rule. *Minnesota Law Review*, *86*, 1439.

Grasswick, H. (2019). Reconciling Epistemic Trust and Responsibility. In *Trust in Epistemology*. Routledge.

Green, C. (2006). *The Epistemic Parity of Testimony, Memory, and Perception by Christopher R. Green: SSRN*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1005782&download=yes

Green, C. (2010). *Epistemology of Testimony | Internet Encyclopedia of Philosophy*. https://iep.utm.edu/ep-testi/

Grice, P. (1989). *Studies in the Way of Words*. Harvard University Press.

Guardian. (2019, December 8). *NHS gives Amazon free use of health data under Alexa advice deal*. The Guardian. http://www.theguardian.com/society/2019/dec/08/nhs-gives-amazon-free-use-of-health-data-under-alexa-advice-deal

Guerin v. The Queen, 2 SCR 335 (C 1984). http://canlii.ca/t/1lpfn

Hamari, J., & Koivisto, J. (2015). "Working out for likes": An empirical study on social influence in exercise gamification. *Computers in Human Behavior*, *50*, 333–347. https://doi.org/10.1016/j.chb.2015.04.018

Hardwig, J. (1991). The Role of Trust in Knowledge. *The Journal of Philosophy*, *88*(12), 693–708. https://doi.org/10.2307/2027007

Hegelich, S. (2020). Facebook needs to share more with researchers. *Nature*, *579*(7800), 473–474.

Hendriks, F., Kienhues, D., & Bromme, R. (2016). Trust in Science and the Science of Trust. In B. Blöbaum (Ed.), *Trust and Communication in a Digitized World: Models and Concepts of Trust Research* (pp. 143–159). Springer International Publishing. https://doi.org/10.1007/978-3-319-28059-2_8

Higgins, J. P. (2016). Smartphone Applications for Patients' Health and Fitness. *The American Journal of Medicine*, *129*(1), 11–19. https://doi.org/10.1016/j.amjmed.2015.05.038

Hinchman, E. S. (2020, May 7). *Assertion and Testimony*. The Oxford Handbook of Assertion. https://doi.org/10.1093/oxfordhb/9780190675233.013.23

Hintze, M. (2017). *Viewing the GDPR through a De-Identification Lens: A Tool for Compliance, Clarification, and Consistency* (SSRN Scholarly Paper ID 2909121). Social Science Research Network. https://papers.ssrn.com/abstract=2909121

Holman, C. D., Bass, A. J., Rosman, D. L., Smith, M. B., Semmens, J. B., Glasson, E. J., Brook, E. L., Trutwein, B., Rouse, I. L., Watson, C. R., De, N. K., & Stanley, F. J. (2008). A decade of data linkage in Western Australia: Strategic design, applications and benefits of the WA data linkage system. *Australian Health Review : A Publication of the Australian Hospital Association*, *32*(4), 766–777. https://doi.org/10.1071/AH080766

Honary, M., Bell, B. T., Clinch, S., Wild, S. E., & McNaney, R. (2019). Understanding the Role of Healthy Eating and Fitness Mobile Apps in the Formation of Maladaptive Eating and Exercise Behaviors in Young People. *JMIR MHealth and UHealth*, *7*(6), e14239. https://doi.org/10.2196/14239

Hong, G., Folcarelli, A., Less, J., Wang, C., Erbasi, N., & Lin, S. (2021). Voice Assistants and Cancer Screening: A Comparison of Alexa, Siri, Google Assistant, and Cortana. *The Annals of Family Medicine*, *19*(5), 447–449. https://doi.org/10.1370/afm.2713

Hopia, H., & Raitio, K. (2016). Gamification in Healthcare: Perspectives of Mental Health Service Users and Health Professionals. *Issues in Mental Health Nursing*, *37*(12), 894–902. https://doi.org/10.1080/01612840.2016.1233595

Houkes, W., & Vermaas, P. E. (2010). *Technical Functions: On the Use and Design of Artefacts*. Springer Science & Business Media.

Hursthouse, R., & Pettigrove, G. (2018). Virtue Ethics. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2018). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2018/entries/ethics-virtue/

Ii, P., & Nicholson, W. (2019). *Medical AI and Contextual Bias* (SSRN Scholarly Paper ID 3347890). Social Science Research Network. https://papers.ssrn.com/abstract=3347890

Irzik, G., & Kurtulmus, F. (2019). What Is Epistemic Public Trust in Science? *The British Journal for the Philosophy of Science*, *70*(4), 1145–1166. https://doi.org/10.1093/bjps/axy007

Jardine, J., Fisher, J., & Carrick, B. (2015). *Apple's ResearchKit: Smart data collection for the smartphone era?* https://journals.sagepub.com/doi/full/10.1177/0141076815600673

Jensen, C., & Potts, C. (2004). Privacy Policies As Decision-making Tools: An Evaluation of Online Privacy Notices. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 471–478. https://doi.org/10.1145/985692.985752

Johnson, D., Deterding, S., Kuhn, K.-A., Staneva, A., Stoyanov, S., & Hides, L. (2016). Gamification for health and wellbeing: A systematic review of the literature. *Internet Interventions*, *6*, 89–106. https://doi.org/10.1016/j.invent.2016.10.002

Johnson, L. (2003). After Enron: Remembering Loyalty Discourse in Corporate Law. *Delaware Journal of Corporate Law*, *28*, 27.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, *47*(2), 263–291. https://doi.org/10.2307/1914185

Kaplan, B. (2016). How Should Health Data Be Used?: Privacy, Secondary Use, and Big Data Sales. *Cambridge Quarterly of Healthcare Ethics*, *25*(2), 312–329. https://doi.org/10.1017/S0963180115000614

Kelion, L. (2020, April 20). Coronavirus: Why are there doubts over contact-tracing apps? *BBC News*. https://www.bbc.com/news/technology-52353720

Kelp, C. (2018). Assertion: A Function First Account. *Noûs*, *52*(2), 411–442. https://doi.org/10.1111/nous.12153

Kerner, C., & Goodyear, V. A. (2017). The Motivational Impact of Wearable Healthy Lifestyle Technologies: A Self-determination Perspective on Fitbits With Adolescents. *American Journal of Health Education*, *48*(5), 287–297. https://doi.org/10.1080/19325037.2017.1343161

Khan, L. M., & Pozen, D. E. (2019). A Skeptical View of Information Fiduciaries. *Harvard Law Review*, *133*, 497.

Kim, T. W., & Werbach, K. (2016). More than just a game: Ethical issues in gamification. *Ethics and Information Technology*, *18*(2), 157–173. https://doi.org/10.1007/s10676-016-9401-5

Kneer, M. (2021). Norms of assertion in the United States, Germany, and Japan. *Proceedings of the National Academy of Sciences*, *118*(37). https://doi.org/10.1073/pnas.2105365118

Konnoth, C. (2015). Classification and Standards for Health Information: Ethical and Practical Approaches. *Washington and Lee Law Review Online*, *72*, 397.

Kumar, A. (2014). *Zero Day Exploit* (SSRN Scholarly Paper ID 2378317). Social Science Research Network. https://papers.ssrn.com/abstract=2378317

Lai, S.-P., Hsieh, C.-A., Harutaipree, T., Lin, S.-C., Peng, Y.-H., Cheng, L.-P., & Chen, M. Y. (2019). FitBird: Improving Free-weight Training Experience using Wearable Sensors for Game Control. *Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts*, 475–481. https://doi.org/10.1145/3341215.3356258

Landers, R. N., Auer, E. M., Collmus, A. B., & Armstrong, M. B. (2018). Gamification Science, Its History and Future: Definitions and a Research Agenda. *Simulation & Gaming*, *49*(3), 315–337. https://doi.org/10.1177/1046878118774385

Lanzing, M. (2019). "Strongly Recommended" Revisiting Decisional Privacy to Judge Hypernudging in Self-Tracking Technologies. *Philosophy & Technology*, *32*(3), 549–568. https://doi.org/10.1007/s13347-018-0316-4

Latour, B. (2012). *We Have Never Been Modern*. Harvard University Press.

Lee, A. (2020, April 17). If Bluetooth doesn't work for contact-tracing apps, what will? *Wired UK*. https://www.wired.co.uk/article/bluetooth-contact-tracing-apps

Leonard, N. (2021). Epistemological Problems of Testimony. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2021). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/sum2021/entriesestimony-episprob/

Leonelli, S. (2015). What Counts as Scientific Data? A Relational Framework. *Philosophy of Science*, *82*(5), 810–821. https://doi.org/10.1086/684083

Leonelli, S. (2016). *Data-Centric Biology: A Philosophical Study*. University of Chicago Press.

Leonelli, S. (2019). Data—From objects to assets. *Nature*, *574*(7778), 317–320. https://doi.org/10.1038/d41586-019-03062-w

Licht, A. N. (2016). *Motivation, Information, Negotiation: Why Fiduciary Accountability Cannot Be Negotiable* (SSRN Scholarly Paper ID 2811237). Social Science Research Network. https://papers.ssrn.com/abstract=2811237

Long, B. (2017, April 11). *Lewis Silkin—Introductory guide to data sharing*. Lewis Silkin. http://www.lewissilkin.com/Insights/Introductory-guide-to-data-sharing

Lovell, T. (2020, June). *UK government releases details of COVID-19 data-sharing deals with big tech firms after legal action threat | Healthcare IT News*. https://www.healthcareitnews.com/news/europe/uk-government-releases-details-covid-19-data-sharing-deals-big-tech-firms-after-legal

Lupton, D. (2015). *Digital Health Technologies and Digital Data: New Ways of Monitoring, Measuring and Commodifying Human Embodiment, Health and Illness* (SSRN Scholarly Paper ID 2552998). Social Science Research Network. https://papers.ssrn.com/abstract=2552998

Lupton, D. (2017). How does health feel? Towards research on the affective atmospheres of digital health. *DIGITAL HEALTH*, *3*, 2055207617701276. https://doi.org/10.1177/2055207617701276

Lupton, D., & Thomas, G. M. (2015). Playing Pregnancy: The Ludification and Gamification of Expectant Motherhood in Smartphone Apps. *M/C Journal*, *18*(5), 5. https://doi.org/10.5204/mcj.1012

Mani, Z., & Chouk, I. (2019). Impact of privacy concerns on resistance to smart services: Does the 'Big Brother effect' matter? *Journal of Marketing Management*, *35*(15–16), 1460–1479. https://doi.org/10.1080/0267257X.2019.1667856

Marsili, N. (2019). The norm of assertion: A 'constitutive' rule? *Inquiry*, *0*(0), 1–22. https://doi.org/10.1080/0020174X.2019.1667868

Marsili, N. (2020). *The Definition of Assertion* (SSRN Scholarly Paper ID 3711804). Social Science Research Network. https://doi.org/10.2139/ssrn.3711804

Marsili, N., & Wiegmann, A. (2021). Should I say that? An experimental investigation of the norm of assertion. *Cognition*, *212*, 104657. https://doi.org/10.1016/j.cognition.2021.104657

Martin, G., Martin, P., Hankin, C., Darzi, A., & Kinross, J. (2017). Cybersecurity and healthcare: How safe are we? *BMJ*, *358*.

Maturo, A., & Setiffi, F. (2016). The gamification of risk: How health apps foster self-confidence and why this is not enough. *Health Risk & Society*, *17*(7–8), 477–494. https://doi.org/10.1080/13698575.2015.1136599

Mayer-Schönberger, V., & Cukier, K. (2013). *Big Data: A Revolution that Will Transform how We Live, Work, and Think*. Houghton Mifflin Harcourt.

McDaniel, R. (2016). A taxonomy for digital badge design in medical technologies. *2016 IEEE International Conference on Serious Games and Applications for Health (SeGAH)*, 1–8. https://doi.org/10.1109/SeGAH.2016.7586254

McDonald, A. M., & Cranor, L. F. (2008). The Cost of Reading Privacy Policies. *I/S: A Journal of Law and Policy for the Information Society*, *4*, 543.

Mehlman, M. J. (2015). Why Physicians Are Fiduciaries For Their Patients. *Indiana Health Law Review*, *12*(1), 1–64. https://doi.org/10.18060/18959

Meier-Abt, P. J., Lawrence, A. K., Selter, L., Vayena, E., & Schwede, T. (2018). The Swiss approach to precision medicine. *Swiss Medical Weekly*. https://doi.org/10.3929/ethz-b-000274911

Miller, K. (2021). *Are Voice Assistants a Reliable Source of Health Information?* Stanford HAI. https://hai.stanford.edu/news/are-voice-assistants-reliable-source-health-information

Miller, P. (2011). A Theory of Fiduciary Liability. *McGill Law Journal / Revue de Droit de McGill*, *56*(2), 235–288. https://doi.org/10.7202/1002367ar

MIT Technology Review. (2020). *Google's medical AI was super accurate in a lab. Real life was a different story*. MIT Technology Review. https://www.technologyreview.com/2020/04/27/1000658/google-medical-ai-accurate-lab-real-life-clinic-covid-diabetes-retina-disease/

Montag, C., Hegelich, S., Sindermann, C., Rozgonjuk, D., Marengo, D., & Elhai, J. D. (2021). On Corporate Responsibility When Studying Social Media Use and Well-Being. *Trends in Cognitive Sciences*, *25*(4), 268–270. https://doi.org/10.1016/j.tics.2021.01.002

Moore, R. (2017). *Appropriate Voices for Artefacts: Some Key Insights*. http://scholar.googleusercontent.com/scholar?q=cache:2BiQbkJil1UJ:scholar.google.com/+Moore,+R.+K.+(2017,+August).+Appropriate+voices+for+artefacts:+Some+key+insights.+In+1st+International+workshop+on+vocal+interactivity+in-and-between+humans,+animals+and+robots.&hl=en&as_sdt=0,5

Narayanan, A., & Felten, E. (2014). *No silver bullet: De-identification still doesn't work*. http://www.privacylives.com/wp-content/uploads/2015/02/narayanan-felten-no-silver-bullet-de-identification-2014.pdf

Nguyen, C. T. (2020). *Games: Agency As Art*. Oxford University Press.

Nguyen, C. T. (2021). How Twitter Gamifies Communication. In J. Lackey (Ed.), *Applied Epistemology* (pp. 410–436). Oxford University Press. https://philarchive.org/rec/NGUHTG

Nickel, P. J. (2011). Ethics in e-trust and e-trustworthiness: The case of direct computer-patient interfaces. *Ethics and Information Technology*, *13*(4), 355–363. https://doi.org/10.1007/s10676-011-9271-9

Nickel, P. J. (2013). Artificial Speech and Its Authors. *Minds and Machines*, *23*(4), 489–502. https://doi.org/10.1007/s11023-013-9303-9

Nissenbaum, H. (2011a). A Contextual Approach to Privacy Online. *Daedalus*, *140*(4), 32–48. https://doi.org/10.1162/DAED_a_00113

Nissenbaum, H. (2011b). A Contextual Approach to Privacy Online. *Daedalus*, *140*(4), 32–48. https://doi.org/10.1162/DAED_a_00113

Nissenbaum, H., & Patterson, H. (2016). Biosensing in context: Health privacy in a connected world. In *Quantified* (pp. 79–100). The MIT Press. http://www.scopus.com/inward/record.url?scp=85011573374&partnerID=8YFLogxK

NW, 1615 L. St, Suite 800Washington, & Inquiries, D. 20036USA202-419-4300 | M.-857-8562 | F.-419-4372 | M. (2019, August 2). How Americans view research and findings. *Pew Research Center Science & Society*. https://www.pewresearch.org/science/2019/08/02/americans-say-open-access-to-data-and-independent-review-inspire-more-trust-in-research-findings/

Ohm, P. (2009). *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization* (SSRN Scholarly Paper ID 1450006). Social Science Research Network. https://papers.ssrn.com/abstract=1450006

Olmstead, K. (2017). Voice assistants used by 46% of Americans, mostly on smartphones. *Pew Research Center*. https://www.pewresearch.org/fact-tank/2017/12/12/nearly-half-of-americans-use-digital-voice-assistants-mostly-on-their-smartphones/

O'Neil, C. (2020). *The Covid-19 Tracking App Won't Work*. BloombergQuint. https://www.bloombergquint.com/gadfly/the-covid-19-tracking-app-won-t-work

Origgi, G. (2012). Epistemic Injustice and Epistemic Trust. *Social Epistemology*, *26*(2), 221–235. https://doi.org/10.1080/02691728.2011.652213

Orji, R., Nacke, L. E., & Di Marco, C. (2017). Towards Personality-driven Persuasive Health Games and Gamified Systems. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 1015–1027. https://doi.org/10.1145/3025453.3025577

Owens, J., & Cribb, A. (2019). 'My Fitbit Thinks I Can Do Better!' Do Health Promoting Wearable Technologies Support Personal Autonomy? *Philosophy & Technology*, *32*(1), 23–38. https://doi.org/10.1007/s13347-017-0266-2

Pagin, P., & Marsili, N. (2021). Assertion. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2021). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2021/entries/assertion/

Parikh, R. B., Teeple, S., & Navathe, A. S. (2019). Addressing Bias in Artificial Intelligence in Health Care. *JAMA*, *322*(24), 2377–2378. https://doi.org/10.1001/jama.2019.18058

Patil, S., Patruni, B., Lu, H., Dunkerley, F., Fox, J., Potoglou, D., & Robinson, N. (2015). *Privacy of health records: Europeans' preferences on electronic health data storage and sharing*. Rand Corporation.

Pols, A. (2010). Transferring Responsibility Through Use Plans. In I. Poel & D. Goldberg (Eds.), *Philosophy and Engineering: An Emerging Agenda* (pp. 189–203). Springer Netherlands. https://doi.org/10.1007/978-90-481-2804-4_16

Porter, J. (2020, March 27). *Amazon's Alexa voice assistant can now help you diagnose COVID-19*. The Verge. https://www.theverge.com/2020/3/27/21196735/amazon-alexa-covid-19-coronavirus-diagnosis-advice-symptoms-risk-factors-travel

Powles, J., & Hodson, H. (2017). Google DeepMind and healthcare in an age of algorithms. *Health and Technology*, *7*(4), 351–367. https://doi.org/10.1007/s12553-017-0179-1

Prasad M, D., & Menon C, S. (2020). The Personal Data Protection Bill, 2018: India's regulatory journey towards a comprehensive data protection law. *International Journal of Law and Information Technology*, *28*(1), 1–19. https://doi.org/10.1093/ijlit/eaaa003

Price, W. N. (2017a). *Medical Malpractice and Black-Box Medicine* (SSRN Scholarly Paper ID 2910417). Social Science Research Network. https://papers.ssrn.com/abstract=2910417

Price, W. N. (2017b). *Regulating Black-Box Medicine* (SSRN Scholarly Paper ID 2938391). Social Science Research Network. https://papers.ssrn.com/abstract=2938391

Recital 39—Principles of data processing. (n.d.). *General Data Protection Regulation (GDPR)*. Retrieved December 13, 2018, from https://gdpr-info.eu/recitals/no-39/

Recital 58, GDPR. (n.d.). *General Data Protection Regulation (GDPR)*. Retrieved February 6, 2018, from https://gdpr-info.eu/recitals/no-58/

Richards, N., & Hartzog, W. (2021). A Duty of Loyalty for Privacy Law. *Washington University Law Review*, *99*, 961.

Rieder, A., Vuckic, S., Schache, K., & Jung, R. (2020). Technostress from Persuasion: Wearable Users' Stressors, Strains, and Coping. *ICIS 2020 Proceedings*. https://aisel.aisnet.org/icis2020/user_behaviors/user_behaviors/8

Rieder, B., & Hofmann, J. (2020). Towards platform observability. *Internet Policy Review*, *9*(4), 1–28. https://doi.org/10.14763/2020.4.1535

Robaey, Z. (2016). Transferring Moral Responsibility for Technological Hazards: The Case of GMOs in Agriculture. *Journal of Agricultural and Environmental Ethics*, *29*(5), 767–786. https://doi.org/10.1007/s10806-016-9636-5

Rockmann, R. (2019). DON'T HURT ME… NO MORE? AN EMPIRICAL STUDY ON THE POSITIVE AND ADVERSE MOTIVATIONAL EFFECTS IN FITNESS APPS. *Research Papers*. https://aisel.aisnet.org/ecis2019_rp/90

Rogers, A. (2017, April 20). That Google Spinoff's Scary, Important, Invasive, Deep New Health Study. *Wired*. https://www.wired.com/2017/04/wholl-really-benefit-verilys-exhaustive-health-study/

Rolin, K. (2002). Gender and Trust in Science. *Hypatia*, *17*(4), 95–118. https://doi.org/10.1111/j.1527-2001.2002.tb01075.x

Rolin, K. (2015). Values in Science: The Case of Scientific Collaboration. *Philosophy of Science*, *82*(2), 157–177. https://doi.org/10.1086/680522

Rolin, K. H. (2014). *Susann Wagenknecht's " 'Facing the Incompleteness of epistemic trust'—A Critical Reply."* https://helda.helsinki.fi/handle/10138/168731

Rotman, L. (2011). Fiduciary Law's 'Holy Grail': Reconciling Theory and Practice in Fiduciary Jurisprudence. *Knowledge@SchulichLaw*, *0*(0). https://ojs.library.dal.ca/KNOWSL/article/view/4742

Rubinstein, I. (2012). *Big Data: The End of Privacy or a New Beginning?* (SSRN Scholarly Paper ID 2157659). Social Science Research Network. https://papers.ssrn.com/abstract=2157659

Ruckenstein, M., & Schüll, N. D. (2017). The Datafication of Health. *Annual Review of Anthropology*, *46*(1), 261–278. https://doi.org/10.1146/annurev-anthro-102116-041244

Rudy-Hiller, F. (2018). *The Epistemic Condition for Moral Responsibility*. https://stanford.library.sydney.edu.au/archives/win2019/entries/moral-responsibility-epistemic/

Rutjens, B. T., Sutton, R. M., & Lee, R. van der. (2017). Not All Skepticism Is Equal: Exploring the Ideological Antecedents of Science Acceptance and Rejection: *Personality and Social Psychology Bulletin*. https://doi.org/10.1177/0146167217741314

Sardi, L., Idri, A., & Fernández-Alemán, J. L. (2017). A systematic review of gamification in e-Health. *Journal of Biomedical Informatics*, *71*, 31–48. https://doi.org/10.1016/j.jbi.2017.05.011

Schmidt-Kraepelin, M., Thiebes, S., Stepanovic, S., Mettler, T., & Sunyaev, A. (2019). Gamification in Health Behavior Change Support Systems—A Synthesis of Unintended Side Effects. *Wirtschaftsinformatik 2019 Proceedings*. https://aisel.aisnet.org/wi2019/track08/papers/9

Schreuter, D., van der Putten, P., & Lamers, M. H. (2021). Trust Me on This One: Conforming to Conversational Assistants. *Minds and Machines*, *31*(4), 535–562. https://doi.org/10.1007/s11023-021-09581-8

Seaborn, K., & Fels, D. I. (2015). Gamification in theory and action: A survey. *International Journal of Human-Computer Studies*, *74*, 14–31. https://doi.org/10.1016/j.ijhcs.2014.09.006

Searle, J. (2000). What is a speech act? In *Perspectives in the philosophy of language: A concise anthology* (pp. 253–268).

Sharon, T. (2016). The Googlization of health research: From disruptive innovation to disruptive ethics. *Personalized Medicine*, *13*(6), 563–574. https://doi.org/10.2217/pme-2016-0057

Sharon, T. (2017). Self-Tracking for Health and the Quantified Self: Re-Articulating Autonomy, Solidarity, and Authenticity in an Age of Personalized Healthcare. *Philosophy & Technology*, *30*(1), 93–121. http://dx.doi.org/10.1007/s13347-016-0215-5

Sharon, T. (2018). Let's Move Beyond Critique—But Please, Let's Not Depoliticize the Debate. *The American Journal of Bioethics*, *18*(2), 20–22. https://doi.org/10.1080/15265161.2017.1409836

Sharon, T. (2019). Data-driven decision making, AI and the Googlization of health research. *Http://Archivio.Paviauniversitypress.It/Oa/9788869521348.Pdf*. https://repository.ubn.ru.nl/handle/2066/219097

Sharon, T. (2021). Blind-sided by privacy? Digital contact tracing, the Apple/Google API and big tech's newfound role as global health policy makers. *Ethics and Information Technology*, *23*(1), 45–57. https://doi.org/10.1007/s10676-020-09547-x

Sharon, T., & Zandbergen, D. (2017). From data fetishism to quantifying selves: Self-tracking practices and the other values of data. *New Media & Society*, *19*(11), 1695–1709. https://doi.org/10.1177/1461444816636090

Sheng, J., Malani, A., Goel, A., & Botla, P. (2022). JUE insights: Does mobility explain why slums were hit harder by COVID-19 in Mumbai, India? *Journal of Urban Economics*, *127*, 103357. https://doi.org/10.1016/j.jue.2021.103357

Sicart (Vila), M. A. (2015). Playing the good life: Gamification and ethics. *Gameful World*, 225–244.

Sinnott-Armstrong, W. (2021). Consequentialism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2021). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2021/entries/consequentialism/

Sitkoff, R. H. (2011). The Economic Structure of Fiduciary Law. *Boston University Law Review*, *91*, 1039.

Smith, D. G. (2002). The Critical Resource Theory of Fiduciary Duty. *Vanderbilt Law Review*, *55*, 1399.

Smith, H. E. (2013). *Why Fiduciary Law Is Equitable* (SSRN Scholarly Paper ID 2321315). Social Science Research Network. https://papers.ssrn.com/abstract=2321315

Smith, W. R., & Treem, J. (2017). Striving to Be King of Mobile Mountains: Communication and Organizing Through Digital Fitness Technology. *Communication Studies*, *68*(2), 135–151. https://doi.org/10.1080/10510974.2016.1269818

Solove, D. J. (2007). I've Got Nothing to Hide and Other Misunderstandings of Privacy. *San Diego Law Review*, *44*, 745.

Sosa, E. (2006). Knowledge: Instrumental and Testimonial. In J. Lackey & E. Sosa (Eds.), *The Epistemology of Testimony* (pp. 116–123). Oxford University Press.

Spagnuelo, D., & Lenzini, G. (2016). Patient-Centred Transparency Requirements for Medical Data Sharing Systems. In *New Advances in Information Systems and Technologies* (pp. 1073–1083). Springer, Cham. https://doi.org/10.1007/978-3-319-31232-3_102

Spillers, F., & Asimakopoulos, S. (2014). Does Social User Experience Improve Motivation for Runners? A Diary Study Comparing Mobile Health Applications. In A. Marcus (Ed.), *Design, User Experience, and Usability: User Experience Design Practice, Pt Iv* (Vol. 8520, pp. 358–369). Springer-Verlag Berlin.

Swan, M. (2012). Health 2050: The Realization of Personalized Medicine through Crowdsourcing, the Quantified Self, and the Participatory Biocitizen. *Journal of Personalized Medicine*, *2*(3), 3. https://doi.org/10.3390/jpm2030093

Taylor, L. (2021). Public Actors Without Public Values: Legitimacy, Domination and the Regulation of the Technology Sector. *Philosophy & Technology*, *34*(4), 897–922. https://doi.org/10.1007/s13347-020-00441-4

Terry, N. (2012). *Protecting Patient Privacy in the Age of Big Data* (SSRN Scholarly Paper ID 2153269). Social Science Research Network. https://papers.ssrn.com/abstract=2153269

The Seven-Per-Cent Solution. (1976). Universal Studios.

Tigard, D. W. (2021). There Is No Techno-Responsibility Gap. *Philosophy & Technology*, *34*(3), 589–607. https://doi.org/10.1007/s13347-020-00414-7

Topol, E. J. (2015, January 9). The Future of Medicine Is in Your Smartphone. *Wall Street Journal*. https://online.wsj.com/articles/the-future-of-medicine-is-in-your-smartphone-1420828632

Trang, S., & Weiger, W. H. (2021). The perils of gamification: Does engaging with gamified services increase users' willingness to disclose personal information? *Computers in Human Behavior*, *116*, 106644. https://doi.org/10.1016/j.chb.2020.106644

van de Poel, I., & Robaey, Z. (2017). Safe-by-Design: From Safety to Responsibility. *NanoEthics*, *11*(3), 297–306. https://doi.org/10.1007/s11569-017-0301-x

van der Drift, S., Wismans, L., & Olde Kalter, M.-J. (2022). Changing mobility patterns in the Netherlands during COVID-19 outbreak. *Journal of Location Based Services*, *16*(1), 1–24. https://doi.org/10.1080/17489725.2021.1876259

van Dooren, M. M. M., Siriaraya, P., Visch, V., Spijkerman, R., & Bijkerk, L. (2019). Reflections on the design, implementation, and adoption of a gamified eHealth application in youth mental healthcare. *Entertainment Computing*, *31*, 100305. https://doi.org/10.1016/j.entcom.2019.100305

Vincent, J. (2019, July 10). *Amazon's Alexa will deliver NHS medical advice in the UK*. The Verge. https://www.theverge.com/2019/7/10/20688654/amazon-alexa-health-advice-uk-nhs

Vogt, H., Green, S., Ekstrøm, C. T., & Brodersen, J. (2019). How precision medicine and screening with big data could increase overdiagnosis. *BMJ*, l5270. https://doi.org/10.1136/bmj.l5270

Wachter, S. (2018). The GDPR and the Internet of Things: A three-step transparency model. *Law, Innovation and Technology*, *10*(2), 266–294. https://doi.org/10.1080/17579961.2018.1527479

Wagenknecht, S. (2015). Facing the Incompleteness of Epistemic Trust: Managing Dependence in Scientific Practice. *Social Epistemology*, *29*(2), 160–184. https://doi.org/10.1080/02691728.2013.794872

Wahle, F., Kowatsch, T., Fleisch, E., Rufer, M., & Weidt, S. (2016). Mobile Sensing and Support for People With Depression: A Pilot Trial in the Wild. *JMIR MHealth and UHealth*, *4*(3), e111. https://doi.org/10.2196/mhealth.5960

Whelan, E., & Clohessy, T. (2020). How the social dimension of fitness apps can enhance and undermine wellbeing: A dual model of passion perspective. *Information Technology & People*, *34*(1), 68–92. https://doi.org/10.1108/ITP-04-2019-0156

Whitson, J. R. (2013). Gaming the Quantified Self. *Surveillance & Society*, *11*(1/2), 163–176. https://doi.org/10.24908/ss.v11i1/2.4454

Wilholt, T. (2013). Epistemic Trust in Science. *The British Journal for the Philosophy of Science*, *64*(2), 233–253. https://doi.org/10.1093/bjps/axs007

Williamson, O. E. (1975). *Markets and Hierarchies: Analysis and Antitrust Implications: A Study in the Economics of Internal Organization* (SSRN Scholarly Paper ID 1496220). Social Science Research Network. https://papers.ssrn.com/abstract=1496220

Williamson, T. (2002). *Knowledge and Its Limits*. Oxford University Press.

Worthington, S. (2006). *Equity*. OUP Oxford.

Wynants, L., Smits, L. J. M., & Calster, B. V. (2020). Demystifying AI in healthcare. *BMJ*, *370*. https://doi.org/10.1136/bmj.m3505

Yakowitz, J. (2011). Tragedy of the Data Commons. *Harvard Journal of Law & Technology*, *25*, 1.

Zarsky, T. Z. (2016). Incompatible: The GDPR in the Age of Big Data. *Seton Hall Law Review*, *47*, 995.

Zittrain, J., & Balkin, J. M. (2016, October 3). A Grand Bargain to Make Tech Companies Trustworthy. *The Atlantic*. https://www.theatlantic.com/technology/archive/2016/10/information-fiduciary/502346/

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power: Barack Obama's Books of 2019*. Profile Books.

Summary

Characterized by an accelerating rise in the collection and analysis of quantified data, the healthcare sector is witnessing rapid "datafication". Health datafication is driven by technologies that enable the collection of large sets of data, such as wearable health devices, as well as computational technologies, including machine learning techniques, that can process such big data. Proponents of the health datafication phenomenon have emphasized how it can empower citizens by allowing them to take control of their health as well as monitor aspects of their health that would have previously been impossible to track unaided. Critics, however, point out that health datafication can also diminish our understanding of individual health, for example, by privileging narrowly construed quantified ways of knowing over rich understandings of what healthy behavior is. Another worry is that health datafication shifts the responsibility of healthcare from institutional actors such as medical professionals and policymakers to individual users. This shift in responsibility, and focus on individual responsibility for health, is particularly worrying taking into account the valuable labor of a multitude of actors required to reap the benefits of health datafication.

The significance of the roles and responsibilities of such actors becomes even more apparent when one considers the multilevel nature of inquiry within the health datafication paradigm. On an individual level, health datafication has the potential to empower users by providing them real-time access to their health indicators that would otherwise have been opaque to them, such as sleep quality indicators from a sleep-tracking device. The design of such a device in itself may involve multiple actors, such as those designing the hardware and those designing the software. These actors may themselves have to rely on scientific research output or technological innovation of others in order to design the device such that it is able to accurately capture the required data and produce a desirable output. On a collective level, health datafication offers opportunities to help local and global health bodies to offer and improve services such as infectious disease surveillance. Analysis of data collected in real-time and in real-world settings, such as through smartphone apps, also has the potential to offer new and better ways of diagnosing diseases. The success of such collective level tasks, again, depends on the work done by designers of such apps,

researchers analysing the data, and other actors who may help to integrate such apps with the local health system.

The aim in this thesis is to explore and explicate the role these various sets of actors play in conjunction with each other, and in particular the responsibilities such actors have in facilitating successful health inquiry within the datafied health paradigm. Each chapter of the thesis focuses on the responsibility of a different (set of) actor(s) who contribute(s) to one of the phases along what can be called as the health "data value chain". The data value chain here can be understood as a series of three iterative and overlapping phases involved in the creation of valuable insights from data: data collection and storage; data analysis and processing; and data usage. Other than the designers of health-related apps and wearable devices, the thesis also pays special attention to the roles and responsibilities of health policymakers and regulators as well as large tech corporations that are equipped with significant technical, financial, and human resources to collect and analyse large sets of health data. In investigating the nature and content of responsibilities of such actors, the thesis investigates the moral, legal, social, and epistemic dimensions of such responsibilities. Understanding the roles and responsibilities of these varied sets of actors is crucial in designing a datafied health system that combines expertise beneficially and avoids potential pitfalls.

In chapter one of the thesis, titled "Digital health fiduciaries: protecting user privacy when sharing health data" I discuss the ethical risks related to (loss of) privacy. Privacy risks are central to debates around health datafication as datafication seems predicated on the logic that the more the data, the better the insights, or in general epistemic goods, one can gain from it. While users of digital health technologies, such as wearable devices, are interested in gaining valuable insights about their health, they also have legitimate expectations for the protection of their privacy, or at least, to keep the loss of privacy and corresponding harms to a minimum. This is the argument I follow in this chapter. This chapter explores the responsibilities of digital health data controllers (those who collect and process health data, such as through self-tracking devices), and argues for "fiduciary relationships" between data health controllers and the users. A "fiduciary relationship" is a legal concept, defining the relationship between two parties, a fiduciary and a beneficiary, such that the fiduciary has to keep the interests of the beneficiary at the forefront. As in the context of health datafication paradigm, fiduciary relationships exist in contexts where there are power asymmetries, and seek to protect the vulnerable party (in this case the users of digital

health devices whose data is being collected) from the negative effects of such asymmetries. I argue that such fiduciary relationships be defined in the case of digital health, such that there are deliberative demands on digital health data controllers to keep the interests of their data subjects at the forefront as well as cater to the contextual nature of privacy when making decisions about the use of health data. In particular, these deliberative demands put constraints on the kind of epistemic goods data controllers can gain from personal health data as well as the kind of epistemic goods they can facilitate by sharing this data with third parties. These deliberative requirements ensure that users can engage in collective participation and share their health data at a lower risk of privacy harm.

In the second chapter of the thesis, I explore the effects of "Googlization" of health research (GHR) on warranted epistemic public trust (or trustworthiness) in epistemic goods produced by such research. GHR is a term coined by Tamar Sharon to refer to the phenomena of large tech companies such as Alphabet (formerly Google), Amazon, Apple, etc. moving up as dominant, and perhaps indispensable, forces in health research. The question of warranted epistemic public trust in scientific output produced through GHR is important for at least two reasons: epistemic trust is essential for the successful transmission of epistemic goods, and epistemic trust plays an essential in governing and/or legitimizing actions based on such epistemic goods. In this chapter, I build on an important insight from social epistemology and philosophy of science in the context of epistemic public trust which emphasizes that since laypeople often cannot assess the content of scientific claims by themselves, they rationally rely on other experts and broadly on moral and institutional contexts within and through which such claims are produced. I argue that in so far as there are indications of moral failings within practices of GHR, along with (institutional) indicators such as possibilities of bad incentives, there are rational reasons against warranted public epistemic trust (or trustworthiness) in claims produced by GHR. This is another example of how there is a need for responsible behaviour from the companies and corporations that constitute GHR, where such responsible behaviour spans both epistemic and non-epistemic (such as moral) aspects of the inquiry.

The next two chapters of the thesis focus mostly on the final phase of the health data value chain – usage. In the third chapter, I discuss the epistemic, and potentially ethical, responsibility of designers of digital voice assistants, such as Amazon's Alexa, through which many users receive (or may receive in the future) valuable information

related to health and disease. In philosophical terms, the phenomenon of making claims about, reporting, or affirming something is known as the speech act of "assertion". Some philosophers have argued that such instances of information sharing through machine-generated speech are equivalent to cases of humans conversationally sharing information with other, and should also be classified as assertions. This claim regarding machine assertions is partially based on the fact that instances of machine-generated speech seem indistinguishable from human speech, and advances in digital technologies are narrowing this "phenomenological" gap even further. In this chapter, I argue against the claim that machines can assert. My central argument in this chapter is that the speech act of "assertion" requires that the asserter be able to take responsibility (at least epistemic, but potentially also ethical) for the claim that is asserted. Machines, such as the Alexa, I argue fail to fulfill this condition. There is, however, a sense in which the designers of devices like the Alexa can take responsibility for such machine generated utterances, and hence, at least in some cases, such instances of machine-generated speech can be labeled as "proxy assertions". I further contend that only those machine utterances can be deemed as proxy assertions where the designers, or a collective of actors whose work influences the utterance, can reasonably foresee and therefore, take responsibility for such utterances.

From a practical point of view, one of the implications of my argument regarding machine (proxy) assertions is that designers should make it transparent to the users of devices like the Alexa, the kind of machine utterances that they can foresee and take responsibility for (or in other words, which machine utterances can be counted as proxy assertions). This transparency is especially important considering the narrowing phenomenological gap alluded to above. Empirical evidence suggests that when users deem machine speech as equivalent or very similar to human speech (because, for example, it sounds human-like), users may form similar expectations from the machine, they would have from a human regarding, for example, the accuracy of the claim. In other words, when machine speech is phenomenologically similar to human speech, users may come to expect such speech to be a product of (epistemically and perhaps even ethically) responsible action. It is important, then, that designers of devices like the Alexa ensure that users only have such expectations when the machine speech is, in fact, a result of such responsible action – which is only possible when designers can actually foresee and take responsibility for the machine-generated speech. This is particularly crucial in contexts, such as in healthcare, where the

epistemic and ethical risks of unwarranted and/or inappropriate epistemic expectations can be high and problematic.

The final chapter of the thesis focus on the phenomenon of gamification in the context of health and fitness apps – that is, - the use of game-like elements such as rewards and badges given to users as motivation for physical activity. While such "gamified" apps can have a valuable motivational effect on some users, they also come with a "darker side". Sociological analyses of such apps has highlighted, for example, how such gamified apps can manipulate users into behavior that may be psychologically as well as physically harmful. Addressing such concerns is not only of moral importance but also of significance for those interested in engagement with and the effectiveness of such apps. Existing studies that highlight the ethical challenges of gamification have met with some criticism, particularly, that they fall short of providing guidance to practitioners and designers of such apps. In other words, they fail to outline the responsibility of the designers of such gamified apps. As a response to this vacuum, this chapter seeks to facilitate a practice-relevant guide for designers of gamified health apps to address ethical issues raised by the use of such apps. More specifically, the chapter has two major aims: First, to propose a practice-relevant theoretical framework outlining the responsibilities of the designers of gamified health apps. Secondly, the chapter provides a landscape of the various ethical issues encountered in the use of gamified health apps based on a systematic literature review of the empirical literature investigating the adverse effects of such apps.

Finally, in the concluding section I take a step back to reflect on some of the implications of the arguments discussed in the four chapters and suggest directions fur future research.

About the Author

Chirag Arora (1990) completed his PhD in Ethics of Technology at Eindhoven University of Technology (TU/e) between 2017 and 2022. From 2021 to 2022, he has been involved in teaching courses in Ethics of Technology to students at Eindhoven University of Technology. Prior to coming to TU/e, he graduated *cum laude* from the Master program in Philosophy of Science, Technology and Society at University of Twente. Chirag has also received education in engineering and holds degrees in Biochemical Engineering and Biotechnology (B. Tech, M. Tech) from the Indian Institute of Technology, Delhi. In 2019, he was involved in research in the Digital Life Initiative (DLI) at Cornell Tech as a Visiting Researcher. In winter of 2022, he would be joining University of Twente's philosophy department as a postdoctoral researcher.

Simon Stevin Series in Ethics of Technology
Delft University of Technology, Eindhoven University of Technology, University of Twente & Wageningen University
Editors: Philip Brey, Anthonie Meijers, Sabine Roeser and Marcel Verweij

Books and Dissertations

Volume 1: Lotte Asveld, *'Respect for Autonomy and Technology Risks'*, 2008

Volume 2: Mechteld-Hanna Derksen, *'Engineering Flesh, Towards Professional Responsibility for 'Lived Bodies' in Tissue Engineering',* 2008

Volume 3: Govert Valkenburg, *'Politics by All Means. An Enquiry into Technological Liberalism'*, 2009

Volume 4: Noëmi Manders-Huits, *'Designing for Moral Identity in Information Technology'*, 2010

Volume 5: Behnam Taebi, *'Nuclear Power and Justice between Generations. A Moral Analysis of Fuel Cycles'*, 2010

Volume 6: Daan Schuurbiers, *'Social Responsibility in Research Practice. Engaging Applied Scientists with the Socio-Ethical Context of their Work'*, 2010

Volume 7: Neelke Doorn, *'Moral Responsibility in R&D Networks. A Procedural Approach to Distributing Responsibilities'*, 2011

Volume 8: Ilse Oosterlaken, *'Taking a Capability Approach to Technology and Its Design. A Philosophical Exploration'*, 2013

Volume 9: Christine van Burken, *'Moral Decision Making in Network Enabled Operations'*, 2014

Volume 10: Faridun F. Sattarov, *'Technology and Power in a Globalising World, A Political Philosophical Analysis'*, 2015

Volume 11: Gwendolyn Bax, *'Safety in large-scale Socio-technological systems. Insights gained from a series of military system studies'*, 2016

Volume 12: Zoë Houda Robaey, *'Seeding Moral Responsibility in Ownership. How to Deal with Uncertain Risks of GMOs'*, 2016

Volume 13: Shannon Lydia Spruit, *'Managing the uncertain risks of nanoparticles. Aligning responsibility and relationships'*, 2017

Volume 14: Jan Peter Bergen, *Reflections on the Reversibility of Nuclear Energy Technologies*, 2017

Volume 15: Jilles Smids, *Persuasive Technology, Allocation of Control, and Mobility: An Ethical Analysis*, 2018

Volume 16: Taylor William Stone, *Designing for Darkness: Urban Nighttime Lighting and Environmental Values*, 2019

Volume 17: Cornelis Antonie Zweistra, *Closing the Empathy Gap: Technology, Ethics, and the Other*, 2019

Volume 18: Ching Hung, *Design for Green: Ethics and Politics for Behavior-Steering Technology*, 2019

Volume 19: Marjolein Lanzing, *The Transparent Self: a Normative Investigation of Changing Selves and Relationships in the Age of the Quantified Self*, 2019

Volume 20: Koen Bruynseels, *Responsible Innovation in Data-Driven Biotechnology*, 2021

Volume 21: Melis Baş, *Technological Mediation of Politics. An Arendtian Critique of Political Philosophy of Technology*, 2022

Volume 22: Mandi Astola, *Collective Virtues. A Response to Mandevillian Morality*, 2022

Volume 23: Karolina Kudlek, *The Ethical Analysis of Moral Bioenhancement. Theoretical and Normative Perspectives*, 2022

Volume 24: Chirag Arora, *Responsibilities in a Datafied Health Environment*, 2022

Simon Stevin (1548-1620)

'Wonder en is gheen Wonder'

This series in the philosophy and ethics of technology is named after the Dutch / Flemish natural philosopher, scientist and engineer Simon Stevin. He was an extraordinary versatile person. He published, among other things, on arithmetic, accounting, geometry, mechanics, hydrostatics, astronomy, theory of measurement, civil engineering, the theory of music, and civil citizenship. He wrote the very first treatise on logic in Dutch, which he considered to be a superior language for scientific purposes. The relation between theory and practice is a main topic in his work. In addition to his theoretical publications, he held a large number of patents, and was actively involved as an engineer in the building of windmills, harbours, and fortifications for the Dutch prince Maurits. He is famous for having constructed large sailing carriages.

Little is known about his personal life. He was probably born in 1548 in Bruges (Flanders) and went to Leiden in 1581, where he took up his studies at the university two years later. His work was published between 1581 and 1617. He was an early defender of the Copernican worldview, which did not make him popular in religious circles. He died in 1620, but the exact date and the place of his burial are unknown. Philosophically he was a pragmatic rationalist for whom every phenomenon, however mysterious, ultimately had a scientific explanation. Hence his dictum 'Wonder is no Wonder', which he used on the cover of several of his own books.