Philos. Technol. (2019) 32:155–171
https://doi.org/10.1007/s13347-018-0319-1

RESEARCH ARTICLE

The Digital Phenotype: a Philosophical and Ethical Exploration

Michele Loi[1]

Received: 31 October 2016 / Accepted: 31 May 2018 / Published online: 11 June 2018

Abstract The concept of the digital phenotype has been used to refer to digital data prognostic or diagnostic of disease conditions. Medical conditions may be inferred from the time pattern in an insomniac's tweets, the Facebook posts of a depressed individual, or the web searches of a hypochondriac. This paper conceptualizes digital data as an extended phenotype of humans, that is as digital information produced by humans and affecting human behavior and culture. It argues that there are ethical obligations to persons affected by generalizable knowledge of a digital phenotype, not only those who are personally identifiable or involved in data generation. This claim is illustrated by considering the health-related digital phenotypes of precision medicine and digital epidemiology.

Keywords Information technologies · Innovation · Policy making · Risk, biomedical data · Big data · Algorithms · Discrimination · Genotyping · Microbiomics · Digital epidemiology · Infoveillance · Infodemiology, feedback loop, holism

1 Introduction

This paper proposes a conceptual revision of the concept of the digital phenotype, which has recently been introduced in epidemiology to indicate human digital footprints with diagnostic and prognostic value (Jain et al. 2015). For example, the content and time pattern in a person's tweets can indicate whether she suffers from insomnia; Facebook posts are considered to be depression symptoms (Jain et al. 2015) and could

Part of this work was done while ML was a scientific collaborator at ETH Zurich.

✉ Michele Loi
michele.loi@uzh.ch

[1] University of Zurich, Zurich, Switzerland

possibly be used in predicting suicidal tendencies (Kelion 2017). It is common to classify some information comprising the digital phenotype as health-related even if it is not health information in the customary sense.[1]

The goal of this paper is to explore the ethics of the digital phenotype, which is relevant for the ethical assessment of data governance in personalized medicine and public health. Due to its normative focus, it speaks to an intended audience of bioethicists, political philosophers, and scholars dealing with the regulation of new technologies. As it develops a concept that was drawn from evolutionary biology and repositioned at the boundary between epidemiology and information studies, it intersects the interests of philosophers of technology, philosophy of science, and scholars of science and technology studies. The paper goes beyond the concept of digital phenotype in the literature by considering how data shapes the human environment and is involved in feedback loops. For example, Facebook software architectures affect real-world friendship, and Twitter metrics are perceived to be symbols of popularity and prestige, thus affecting actual popularity and prestige.[2] Thus, we need to ask how persons are influenced by health-related digital phenotypes (indirectly, through the beliefs, behaviors, and norms the data supports), even if they are not identifiable in the data or involved in its production. Moreover, we need to consider differences between digital phenotypes emerging in different ways.

The paper is structured as follows:

In Section 2, I define my purported revision of the original concept of the digital phenotype, introduced in the literature as an extension of Dawkins' theory of the extended phenotype. I explore several facets of the analogy between digital data and the extended phenotype of non-human organisms, such as termite nests or beaver dams, while discussing similarities and differences between the two realms. I argue that the analogy with the extended phenotype of social animals can be insightful in ways that have not been fully appreciated by the scholars responsible for introducing the concept in the literature. In Section 3, I discuss two moral frameworks on the issue of control of digital information: one revolves around personal data protection as an aspect of dignity and the other one on data ownership as an extension of self-ownership. I identify the limits of each approach in dealing with certain types of data-ethics questions. In Section 4, I describe some governance problems concerning the data generated in the context of so-called personalized medicine and data from internet platforms used in digital epidemiology.

This article's contribution to philosophy and ethics is twofold: it criticizes and friendly amends the only conception of the *digital phenotype* in the literature and analyzes two ethical issues that are relevant to the governance of health-related data.

[1] I wish to thank Prof. Ernst Hafen, who inspired me to work on this topic, participants to the ethics panel of the MyData 2017 Conference (Tallin-Helsinki) and Paul-Olivier Dehaye, for co-organizing that panel and providing feedback on a previous draft of this article. Special gratitude is owed to the two anonymous reviewers of this journal, who enriched the paper with their inputs, and in several rounds of review very patiently helped me to give shape to these views and to remove at least the worst sources of unclarity. All remaining problems in the paper are solely the author's responsibility.

[2] Twitter followers include *bots* (software programmed by paid professionals) whose goal is to enhance the perception of popularity of politicians, celebrities and companies (Freelon 2014). Freelon cites the so-called "Karpf's rule": "any metric of digital influence that becomes financially valuable, or is used to determine newsworthiness, will become increasingly unreliable over time". See David Karpf, "Social Science Research Methods in Internet Time," *Information, Communication & Society* 15, no. 5 (June 1, 2012): 650.

2 What Is the Digital Phenotype?

2.1 The Definition of the Digital Phenotype

The concept of digital phenotype has been introduced in the literature as an extension of Richard Dawkins' theory of the extended phenotype (Dawkins 1999, chap. 11). The extended phenotype of an organism can be understood as the transformations it produces in its environment that influence its Darwinian fitness (i.e., the likelihood of transmitting its genes to the next generation). For example, beavers build dams, spiders build nets, bees build beehives, termites build their nests, and earthworms modify the soil in which they live, in ways beneficial to their survival and reproduction.

The digital phenotype—as understood here—is fundamentally a human phenomenon. Humans modify their online environment, leaving traces of themselves while interacting with an ever-expanding network of digital sensors placed by themselves or other humans. They produce the infrastructure by virtue of which all digital information is generated. They use digital data as a means to, or as a way of, exercising their species-typical capabilities (not only biological ones like survival and reproduction, but also those enabled and defined by human culture). Finally, (most) humans aim to control digital information in order to derive benefits from it.

I focus on the ethological meaning of Dawkins' concept—the importance of considering environments as extensions of individual phenotypes—while bracketing its genetic reductionism. The range of online behaviors that can be explained by invoking genetic influences is, plausibly, quite limited. Most behaviors, beliefs, and social norms that influence digital information are too complex and sophisticated to be explained by exclusively biological, or even more narrowly genetic, causes. Likewise, the behaviors, beliefs, and social norms affected by digital information are not reducible to the biological or genetic level of explanation. My emphasis here will be on the co-evolution of cultural entities (in the broadest possible sense) and data-shaped environments.

Hence, I propose to characterize the human digital phenotype as *an assemblage of information in digital form, that humans produce intentionally or as a by-product of other activities, and which affects human behavior*. More succinctly (but less precisely), the human digital phenotype consists of digital information *produced by humans* and *affecting humans*.

2.2 Similarities and Differences Between the Human Digital Phenotype and the Extended Phenotypes of Other Organisms

Let us begin by considering a similarity between the way humans shape their digital environment, and soil invertebrates as ecosystem engineers. The ecological significance of invertebrates such as earthworms, termites, and ants is not reducible to their contribution to the food chain; rather, it also involves their responsibility "for altering ecosystem dynamics through the modification, maintenance, and/or creation of habitats for other organisms in the ecosystem" (Jouquet et al. 2006, 154). Social insects such as termites and ants, for example, create structures (e.g., the *termitarium* and anthill) which are key to their adaptation to the external environment. For example, the *termitarium* is formed by very cohesive soil, built in such a way to prevent water flux

and ants into the nests, and designed to maintain appropriate levels of moisture. On the other hand, all their biological characteristics are adapted to their extended phenotype. The biological success (Darwinian fitness) of social insects—the likelihood of transmitting their genes to the next generation—would be different for the organism considered in isolation from the nest. The creation of such specialized environments and the transmission of genetic information form a feedback loop: different selective forces would be in place if these organisms could no longer produce nest structures, and the evolution of these species would take a different course, genetically and morphologically (Jouquet et al. 2006, 160).

Consider now digital data as a human extended phenotype. First, digital information also causes *transformations of the immediate environment* in which humans pursue their goals. Information collected, aggregated, linked, and made accessible in digital form affects cognition and motivation, facilitating the spread of certain beliefs, behaviors, and social norms, and hindering or slowing down the spread of others. Behaviors, beliefs, and norms, in turn, affect the ways in which humans pursue their goals and at the same time redefine the nature of these goals.[3] Secondly, as in social invertebrates (but unlike, say, spiders), a significant part of the human digital phenotype arises from social coordination, not from the solitary action of isolated individuals.

Let us now focus on the disanalogies between humans and much simpler animals (Fig. 1). The level of evolution influenced by the digital phenotype is mainly *cultural* evolution. The capabilities by virtue of which humans shape their digital environment would not exist without the information contained in *both* genes and in cultural forms (e.g., culturally transmitted know-how). The motivation to generate digital information is not explained exclusively by genetically determined goals, such as survival and reproduction. Culturally defined goals (such as friendship, creative self-expression, political ideas) are equally if not more important. In spite of the higher complexity of human evolution, the analogy with other extended phenotypes retains its heuristic value: humans are also involved in evolutionary feedback loops with their data on both a faster (cultural) evolutionary scale and a slower (genetic) one. In both humans and many complex non-human organisms, evolution involves different levels. These levels are not independent but interdependent, since genetics affects culture and culture affects genetics. This interdependence has been explored in both biology and anthropology (Cavalli-Sforza and Feldman 1981; Jablonka and Lamb 2005), but it is too complex to be considered here in relation to data, so I will focus on the cultural level in isolation from the evolution of the genome. (Notice that only the feedback loop of human cultural evolution is represented on the right-hand side in Fig. 1.)

A biological metaphor that serves a similar heuristic function as that of a (digital) extended phenotype is Deborah Lupton's concept of digital assemblages as a "digital companion species" (Lupton 2016a), in Donna Haraway's sense (Haraway 2008). For Haraway, technologies are a companion species of humans in the sense that they co-evolve through mutual influences. As Lupton observes, such exegesis applies to digital data, where mutual influences between biological life and the data it generates are also the norm:

[3] These feedback loops take place across the offline-online boundary. The philosopher Luciano Floridi uses the concept of *onlife* to indicate the porous nature of the offline-online boundary (The Onlife initiative 2015).

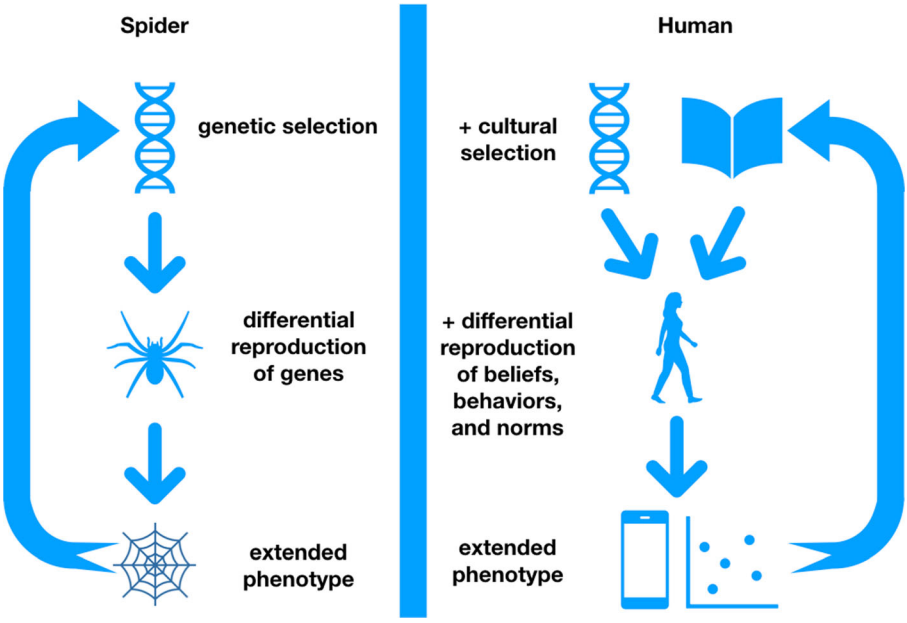

Fig. 1 Analogies between the extended phenotypes of spiders and humans. Built with Apple Keynote 8.0.1, except for web image, by Denis Frezzato, from The Noun Project, CC BY 3.0, https://commons.wikimedia.org/w/index.php?curid=24126123

> Just as digital data assemblages are comprised of specific information points about people's lives, and thus learn from people as algorithmic processes manipulate this personal information, people in turn learn from the assemblages of which they are a part (Lupton 2016a, 2).

Both concepts can serve the same heuristic function, as they both highlight the co-evolution and feedback loops in which data are involved. The concept of a companion species is clearly applied *metaphorically* to digital data. Arguably, however, also the concept of the digital phenotype involves a *non literal* extension of the original (gene-centric) concept found in Dawkins. The "companion species" metaphor is a post-human metaphor which treats data as if it were itself a living species (Haraway 2008), while the digital phenotype idea places humans at the center, both ontologically and ethically, of a data-reality ecosystem.

2.3 Why the Extended Phenotype Analogy Is a Better Analogy than a Family of Other (Widely Abused) Biological Analogies

If the above account is correct, digital information can be considered an extended phenotype not only because, irrespective of its original purpose, it may reveal the (for instance, health-related) conditions of the persons it is about, but also in virtue of how it retroacts on them and other people. The metaphor of a termite nest is more suitable than others such as *footprints*, *traces*, *and tracks* to convey this message. The footprint metaphor suggests a one-to-one correspondence between, for instance, a bear's footprint and a bear's foot. In these abused metaphors, there is no hint to the fact that the

human digital phenotype is a collective creation, involved in social feedback loops. The following section explains why personal data and libertarian moral frameworks are silent on some social implications of generating a digital phenotype. Section 4 presents governance issues independent from personal data protection and ownership and related to the effects of creating a digital phenotype.

3 The Ethics of the Digital Phenotype

One of the most important *normative* questions concerning the digital phenotype concerns who *ought morally to* control it. Different moral frameworks attribute importance to different relations between agents and data. Here, I will analyze two influential moral frameworks, one focused on the protection of the dignity of identifiable individuals and the second on libertarian ownership of the data by individuals. I will explain why these frameworks have limits—there are ethical considerations outside the scope of these theories which are nonetheless ethically important. This analysis helps one to identify the new governance problems examined in Section 4, arising in personalized medicine and digital epidemiology.

The first influential framework analyzed here focuses on the *dignity* of the person *identified* by data. The idea of human dignity is particularly influential in human rights discourse (Griffin 2008, chap. 2), including the right to privacy (Floridi 2016a). Arguably, it received its first modern expression in the work of Pico della Mirandola (an early Renaissance philosopher) and later, a transcendental reformulation through the philosophy of Immanuel Kant (Griffin 2008, chap. 2). The second considers data ownership as a natural extension of self-ownership. The libertarian conception of (moral) property rights as grounded in self-ownership found its most influential expression in John Locke's political philosophy and is still widely influential in that branch of philosophy.[4]

In what follows, I will highlight features of the digital phenotype—in particular, its ability to impinge on the lives of persons who are not identifiable in the data and do not contribute to its production—as a lens to highlight the shortcomings of these two ways of thinking about data.

3.1 The Limits of the Personal Data Protection Approach

The personal-data approach is most influential in data protection and privacy law. The EU's General Data Protection Regulation (GDPR), for example, affords special protection only to *personal* data, defined as data about an *identified or identifiable individual*. This is reasonable in so far as an identifiable individual is, generally speaking, more likely to suffer as a result of misuses of the data, for instance, in that she can become the target of discrimination. However, a person's dignity can also be affected by the generation of digital phenotypes that are not personal data of that individual, as I will now show.

[4] They do not exhaust the range of possible moral theories (or even of European Enlightenment-inspired moral theories). For example four-principlism (Beauchamp and Childress 1994) is influential in bioethics and can be stretched to develop an ethics of health-related data (Mittelstadt 2017a).

Consider the genotyping of an ethnic group, for example the Maori, the indigenous Polynesian people of New Zealand. In 2006, a team of geneticists analyzed the presence of known variants of a gene for MAO (monoamine oxidase), as a marker for alcohol and tobacco dependence, in a group of Maori research subjects. A genetic variation in the metabolism of MAO is also statistically associated with aggressive behavior, lack of self-control, and risk-taking. The science writer Ann Gibbon tagged the MAO gene as the "warrior gene" (Perbal 2013). It turns out that that specific gene was present in 56% of the Maori's sample, with an even greater frequency in a sample of people who have Maori great-grandparents. This is a sensitive issue in so far as it reinforces the stereotype of the violent Maori, that is already widespread (Perbal 2013).

In the genetic case, a digital phenotype (genomic data in digital form) can become a threat to the dignity of Maori individuals, who are potentially negatively affected by stigmatization or discrimination, by virtue of an association with a stereotype of violence. In the worst-case scenario, some individuals may think that these genetic data provide a justification for discrimination on an ethnic basis. Among the potential victims, one may find individuals whose personal data were not used in the research in question, who are not identifiable in the data and thus outside the scope of the data protection framework. Yet, their dignity is arguably threatened by the risk of discrimination.

It may be observed that a lot of instruments exist to empower genetically related groups (or entire genetically related populations) to control their data.[5] But notice that this does not detract from my main claim in this section, which is that, *for the sake of data protection,* such instruments are not called for.

The objection, moreover, invites a useful clarification between the *rationale* sometimes offered for the collective governance of genomic resources and the more general argument used here. It is often pointed out that genetic information is shared (in different degrees) across bloodlines, which entails that the genetic information about a person may also tell us something about his family members and, more broadly, ethnic group. A version of this argument can also be used to justify the collective governance of genetic databases for populations sharing a common ancestry.[6] One is tempted to conclude that the justification of *non individualistic* forms of governance is *essentially* connected with the allegedly special nature of genetic information, namely its being shared (in variable degrees) between individuals with the same genetic ancestry. This may suggest that collective governance is not needed for non-genetic information.

The genetic relatedness argument is, however, a red herring, because virtually any kind of generalizable knowledge (results that can be generalized to a larger population beyond the site of data collection or population studied), genetic or not, exposes individuals who are not research subjects to risks of harm. For example, knowing that smoking increases cancer risk (generalizable knowledge) enables an observer to infer sensitive features of individuals (e.g., cancer risk), from their public features (e.g., smoking). Thus, generalizable knowledge exposes *all* smokers to higher prices by

[5] For example, the "biotrust model" has been proposed to govern genetic biobanks (David E. Winickoff and Winickoff 2003; D. E. Winickoff and Neumann 2005), the solidarity model for research biobanks (Prainsack and Buyx 2013).

[6] This is suggested by indicating shared traits of the genome as part of the reasons to adopt a biotrust model (D. E. Winickoff and Neumann 2005, 9). A similar line of argument appears in Widdows's work on the Connected Self (2013, chap. 3).

insurance companies. One ought to take notice of the fact that generalizable knowledge exposes smokers to be harmed by insurance discrimination, independently from their contributing the data. The analysis of the digital phenotype can uncover new relations between data, creating several new groups of potentially affected parties:

> Think of the owners of such and such kind of car, shoppers of such and such kinds of goods, people who like this type of music, or people who go to that sort of restaurant, cat owners, dog owners, people who live in a specific postal code, carriers of a specific gene, people affected by a particular disease, team fans (Floridi 2016b, 97).

The general phenomenon here is that generalizable knowledge from digital phenotypes enables some agents to discriminate against (or in favor of) some individuals.[7] Here, I am using "discrimination" as a morally neutral term, meaning that benefits are offered, or penalties imposed, on some people but not others, on account of some differences between them. And clearly, not all forms of discrimination are equally morally objectionable—some are arguably justified all things considered. So, for example, even if a minority of smokers (those who are unable to quit smoking) are harmed by higher insurance prices which can be imposed on them, generally the information that smoking causes cancer is beneficial to them, because, for example, it enables them to make informed decisions, for example, to quit smoking or buy life insurance. On the other hand, in the case of research about a "warrior gene," "homosexuality gene" or "race and IQ", the balance between the social value of generalizable knowledge and the risk of discrimination for individuals may be tilted against the generation or diffusion of such knowledge (more on this in Section 4.1).

This section has identified a fundamental moral limit of the personal data protection framework: the possibility of discrimination harm due to generalizable knowledge, for persons who are not identified by the data, which the data protection framework is not equipped to solve. The general implication of this problem is that persons can be affected (in some cases, in dignity-violating ways) by generating new digital phenotypes, irrespective of whether they are identifiable in it. It is important to notice that the moral problem of generalizable knowledge concerns all scientific knowledge, but it arises in a peculiar form for analyses of the digital phenotype. As discussed in Section 4, this is the case for at least two reasons: first, the number and granularity of the affected groups can be higher, for example, specific risks may be identified for the profile of a mother, age group 30–40, of Native American ancestry, commuting at night hours. Second, this knowledge may only be captured as a complex mathematical function describing the behavior of nodes in a neural network, not in a more transparent knowledge form.

3.2 The Limits of the Libertarian Approach

One feature of the personal data protection framework is that the person who produced the data matters morally *only if* that person is identifiable in the data. One plausible

[7] Floridi's account of this issue describes the problem as one of *group privacy* (Floridi 2014, 2016b). My treatment makes no such commitment.

alternative is that she matters *because* of her data contribution, irrespective of identifiability. This ethical belief can be justified by appealing to (politically) libertarian moral principles. Libertarian philosophers have argued that initial (moral) ownership rights are grounded in (moral) *self*-ownership: the natural right that each individual has to use his own body the way one wishes, except in violent ways, against others (Cohen 1995, 292). The purpose of this section is to highlight some difficulties in the application of a libertarian political framework to the digital phenotype.

Libertarian ideas of self-ownership can be extended to data based on the idea of information as an extension of the self. This intuition is nicely captured by Floridi (who, however, does not defend a libertarian view of moral rights to data), when he points out that:

> "my" in "my data" is not the same "my" as in "my car", it is the same "my" as in "my hand", because personal information plays a constitutive role of who I am and can become (2016a, 308).[8]

Alternatively, libertarian rights could be extended to data, based on the analogy between data production and human labor. In a libertarian perspective, acting with one's body on reality (labor) creates moral entitlements to resources that are not antecedently owned (Nozick 1974, 150–53). Suppose that an astronomer records information about the movements of a distant star. These will be *her data* because, first of all, no one had an antecedent title on the information (the data about the star did not exist before the astronomer recorded them); and, second, the astronomer created data with resources (let us suppose) that she was morally entitled to use: her own body and legitimately owned scientific instruments. The astronomer morally owns the data she has produced and has the moral right to do whatever she wants with it.

The libertarian theory applied to digital phenotypes faces (at least) two problems of internal consistency. The first is that there are two equally plausible *rationales* for initial property in information, which can conflict: self-extension and labor. The person who generates the data (relevant for the ownership via labor analogy) can be different from the one the data are about (relevant for the self-extension analogy). Their self-ownership rights may conflict. Suppose that a shop owner had unrestricted ownership and control rights on the video recordings of his customers, deriving from the fact that he has recorded these videos (labor *rationale*). Some conceivable uses of the shop keeper's libertarian ownership rights *qua* creators of the data (e.g., watching recordings of preferred female customers for the purpose of private entertainment) may be morally objectionable because they are violations of the customers' privacy, which can also be interpreted as a violation of self-ownership (self-extension *rationale*).

The second difficulty is that joint production of digital phenotypes, which is widespread, leads to joint ownership in them. Consider digital phenotypes that represent relationships between persons, e.g., online conversations between anonymous individuals, that clearly show how they relate to each other (e.g., politely, aggressively,

[8] See also (Floridi 2011). Notice that Floridi here talks about *personal information*, e.g. a person's Facebook profile is an extension of the self, but this is, arguably, not essential to the argument. I can recognize my informational extensions produced by participating to internet conversations in an anonymous form as *mine*, even though others cannot recognize me.

etc.). *Relations* represented in the data are not *about* any individual in particular—they are about all participants to the conversation simultaneously and about no one specifically. From the point of view of the self-extension theory, the information about different individuals often overlap. If we are *constituted by* our data—if our data are not just *our*, but *us*—then we do not have mutually exclusive identities, even though we are materially distinct individuals. Rather, as informational entities (Floridi 2011), we are, as it were, conjoined twins. Joint ownership and control models however create control problems for digital phenotypes derived from thousands of individuals. Joint ownership arrangement between thousands of, or more, individuals dilute control to such a degree that the value of ownership for the right holder is significantly reduced.

Summing up: the personal data framework leaves persons helpless against the harmful effects of digital phenotypes formed by non-identifiable data. The libertarian theory cannot remediate this flaw because it often fails to provide clear cut criteria to distribute initial property rights or dilutes the value of these titles in a way that leads to ineffective control. In addition to these internal challenges, the libertarian theory does not solve the knowledge problem, the fact that knowledge may have benefits or harm persons who are not involved in the data generation. The conclusion is that we need to consider *the consequences of digital information on all persons affected*, including those affected by the beliefs, behaviors and norms supported (directly or indirectly) by digital phenotypes. The next section deals with this problem in the context of health.

4 Governing the Health-Relevant Digital Phenotype

Having identified the limits of the two moral frameworks of data protection and libertarian ownership, in this section, I explore the implications of these limits for the *health domain*. The first case-study deals with governance of the digital phenotype of personalized medicine, focusing on the problem of harm due to generalizable knowledge (examined in general terms in Section 3.1). The second deals with the ethical obligations emerging from having a large power of *creating and shaping* a digital phenotype, a power that large internet corporations like Google and Facebook have. These examples illustrate the impact of technologies generating a digital phenotype, involved in evolutionary feedback loops (Section 2.2), on persons who are not identifiable in it and have not contributed to its creation. This raises a governance problem about which personal data protection (Section 3.1) or libertarian principles (Section 3.2) offer no guidance.

Before turning to the case studies, I now clarify the *scope* of this section of the paper, which deals with the *health-relevant* digital phenotype. A provisional and non-exhaustive list of data comprising the *health-related digital phenotype* includes (at least):

- data produced in the contexts of disease surveillance, immunization records, public health reporting, vital statistics, and registries (Vayena et al. 2016);
- Web data (including social media data, e.g., Twitter data), e.g., the time pattern in an insomniac's tweets, the Facebook posts of a depressed individual, or the web searches of a hypochondriac (Jain et al. 2015);

- lifestyle data (e.g., as data coming from loyalty cards, location data from computer apps, and mobile phones) (Vayena et al. 2015; European Commission DG Health 2014);
- data parameters such "as heart rate, respiration, blood oxygen saturation, skin temperature, blood glucose, blood chemistry, and body weight […] collected alongside behavioural parameters (e.g., motion, acceleration, mood)" (Mittelstadt 2017b), by health monitoring smart phones application "that can track exercise or movement and share data with health professionals and other third parties" (Mittelstadt 2017a, 17) or by other Internet of Things (IoT) devices, including ambient assisted living devices ("smart home" or "ageing at home" applications, including smart pillboxes and fall detectors), "wearable, embedded and implantable sensors" and "semi-permanent sensors [which] can also be woven into clothes or wristbands for physiological measurements, including detection of specific molecules in perspirations […] or motion metrics to monitor Parkinson's disease" (Mittelstadt 2017a, 17);
- data generated by IT platforms for disease surveillance (e.g., fluenearyou.org) (Vayena et al. 2015, 7);
- molecular data from research in genomics and other so-called—omics (metabolic, proteomic, microbic information), as well as medically applied translational genomics (Vayena et al. 2016; Hood et al. 2015; National Academy of Sciences 2011).

Notice that the health-relevant digital phenotype is not limited to *health data,* even on a broad construction of the latter notion. It clearly encompasses information that is not medical information in the narrow sense of information "gathered by physicians or other members of the medical team" (WMA—The World Medical Association 2016). The *health-related* digital phenotype also encompasses information produced by "individuals themselves via social media, fitness trackers, remote sensors, and the Internet of things", which, it has been argued, will be very relevant for health-care (Aicardi et al. 2016a, see also Aicardi et al. 2016b). A broader concept than that of medical information is that of "data concerning health", as defined by Article 4(15) of the GDPR, which refers (according to recital 35) to information able to "reveal information about […] health status" (General Data Protection Regulation 2016/ 679 2016). This includes information from social media, fitness trackers, and remote sensors, whenever it is processed for health purposes (Mittelstadt 2017a, 2). The concept of the health-relevant digital phenotype is even broader than the concept of data concerning health. The health-related digital phenotype, for example, includes anti-vaccination sentiments on social media. This is a health-related digital phenotype, because it can be studied by public health scholars to understand opposition to vaccination, which is useful for planning vaccination campaigns (Salathé and Khandelwal 2011). But, in many cases, it is not made of "data concerning health", in the GDPR sense, because such data may not be processed to infer the health status of any natural person. The distinction between medical data, data concerning health, and the health-related digital phenotype is a meaningful one and worth preserving, in that there might be ethical arguments that apply specifically to medical data, in the narrow sense, e.g., confidentiality in the doctor-patient relationship, but not to other instances of the broader concepts. The health-related digital phenotype is the broadest concept,

and the ethical questions concerning it are more general than questions concerning medical data or data concerning health.

4.1 Case Study 1. P4 Medicine

Section 3.1 has identified the limits of the personal data and ownership frameworks in relation to the human digital phenotype, namely the fact that these frameworks do not protect persons who are not identifiable in the digital phenotype or involved its generation. Here, I explore some ethical issues related to a digital phenotype produced in a specific medical context (P4 medicine). Although this digital phenotype includes medical data in the narrow sense and data concerning health, here, I consider only the ethical problems arising at the level of analysis of the health-related digital phenotype.

So-called personalized, or P4 (predictive, preventative, participatory, and personalized) medicine aims to combine data from genomics, other molecular biomarkers (for example, epigenetic ones, see Relton and Smith 2010), data from mobile sensors (e.g., GPS on patients smartphones), and self-monitoring (e.g., answers to online surveys), to obtain more fine-grained estimations of risk and tailored therapies for patients, often labeled "participants" (Hood et al. 2015). Some proponents of this approach, more recently rebranded precision medicine, also include internet data and public health surveillance data about environmental or social risk factors (the "exposome"), among the possible sources from which a "Google map" of drivers of health and disease could be derived (National Academy of Sciences 2011, 19). The digital phenotype can be expanded further through patient/citizen participation, by eliciting data generation in a positive feedback loop involving virtual health/fitness coaches, that is, apps running on individual smartphones (Lupton 2015, 2016b).

In spite of the appellative "personalized" often used to describe such applications, the improved ability to detect biomarkers and other data predictive of disease constitutes *generalizable knowledge*, knowledge that goes beyond the individual dimension. Potentially, patterns involving food consumption, cultural consumption, transportation, behavior on social networks, biometrics, and psychometrics can be correlated with medical data, to make predictions about disease, health, and wellness. The knowledge produced in this way is generalizable because it is applicable to other patients/clients beyond the ones contributing the data. Hence, the knowledge from data mining raises the ethical question of discrimination.

This is, again, a general discrimination risk potentially generated by all forms of scientific discovery about humans, as shown by the example of the knowledge that smoking increases cancer risk, which leads to higher insurance prices for smokers (see Section 3.1). As already discussed, this general discrimination risk alone does not make scientific discovery morally objectionable, all things considered. The moral reasons in favor of a discovery (e.g., the overall benefit for humanity of knowing that smoking causes cancer) can be much stronger than reasons against it, when there are only limited negative consequences for some groups (e.g., higher insurance prices for smokers unable to quit). In other cases, balancing the benefits of augmented knowledge with risks for specific group will be more difficult, as in the case of research on race and IQ, or research involving traits that are stigmatized in certain communities, such as sexual orientation. Even valid statistical results, extrapolated from their context, are liable to be misunderstood, and to reinforce bigoted prejudices that might harm such groups

(Kitcher 2001, chap. 10; Buchanan 2007). As both cases illustrate, generalizable knowledge can be both beneficial and harmful for individuals who are not identifiable in the data or, more broadly, involved in their production (Section 3). All generalizable knowledge has, potentially, these features. Yet, as I now briefly discuss, the ramifications of this problem in the context of P4 medicine pose special challenges for governance.

First, in comparison to traditional, hypothesis-driven research, it is harder to *identify* groups harmed or benefited by knowledge from digital phenotypes. Thus, it is harder to identify the stakeholders. Normally big data are not studied for the sake of testing a limited set of hypotheses, identified prior to the data collection. Instead, algorithms are used to discover statistical correlations indicative of new hypotheses (Boyd and Crawford 2012). Moreover, predictions based on statistical correlations may be used in support of decision making. The study of hybrid digital phenotypes in precision medicine may provide decision makers outside health-care with more tools to make health predictions based on non-medical information. For example, companies may be tempted to use information from CVs or employee digital surveillance, which they legitimately collect, to predict the health prospects of applicants and employees and discriminate on that basis. In some cases, such predictions may place special burdens on candidates of already disadvantaged ethnic groups and social backgrounds, including groups not involved in the initial data used to discover the correlations. In others, the discriminated groups may be individuals whose only meaningful association is the statistical one identified by the algorithm (Floridi 2016b, 88), lacking consciousness of themselves as a group. Hence, governance frameworks for digital phenotypes cannot be modeled after governance frameworks for genomic databases or biobanks relating to an antecedently well-defined populations (frameworks for genetic research on endangered populations or small nations).

A second challenge derives from the opacity of some algorithms, such as neural nets, which are being increasingly used to discover relations between data (Mittelstadt et al. 2016; Danaher et al. 2017). When statistical generalizations about ethnic, gender, or sexual preference groups are published in scientific papers, these claims tend to be carefully scrutinized during the peer-review process. Controversial claims tend to attract the attention of the scientific community, which helps society to reduce the impact of the results affected by methodological problems and biases (Kitcher 2001). But with neural nets, no expert may actually be able to explain the scientific grounds behind the traits that are used to make health predictions (Mittelstadt et al. 2016). When the science behind the algorithm is a trade secret and the training data is not publicly accessible, it is more difficult to discuss human prejudices and flawed methods behind a predictive model.

Summing up, data governance should consider not only the populations generating the digital phenotype, but all those that may be harmed or *benefited* by generalizable knowledge from the digital phenotype, for example, in support of public health goals.[9] Yet, such governance framework, involving stakeholders beyond digital phenotype creators, may be difficult to design and implement. There are trade-offs, for example, between the privacy (and ownership) rights of patients providing the data used to train algorithms and the public interest of society of understanding how predictive models

[9] This recommendation corresponds to ethical principle #1 in (Mittelstadt 2017a, 2).

are generated. Moral frameworks that are blind to the latter, or assign absolute priority to the former, may not be ethically adequate after all.

4.2 Case Study 2. Digital Epidemiology and the Use of Internet Data in Public Health

This case study deals with the ethical obligations for those who have unusual powers to generate and control digital phenotypes. By using the digital phenotype as a conceptual frame of reference, I argue that Internet giants have ethical obligations, which derive from the potential of the data they collect, to generate knowledge that is valuable for public health purposes.

Digital epidemiology (Salathé et al. 2012), also known as "infodemiology," has been defined as "the science of distribution and determinants of information in an electronic medium, specifically the Internet, or in a population, with the ultimate aim to inform public health and public policy" (Eysenbach 2009, e11). Digital epidemiology produces knowledge by recycling the "data exhaust" from web activities, such as Internet surfing or tweeting, that are anonymously released by their aggregators, or intended to be public. The knowledge produced in this way is an instance of how digital phenotypes can affect humans by influencing their beliefs, for example, the beliefs of public health authorities about population health and health risks.

The "parable of Google Flu" (Lazer et al. 2014) is a remainder of the fact that these platforms are not optimized to produce knowledge that serves the purposes of public health. Google Flu was the first widely influential project which repurposed Internet data to serve public health goals. After 1 year of operation, it became obvious that it was overestimating flu prevalence by a large margin. Part of the problem was that it was discounting for media-induced flu panics, often related to the outbreak of related epidemics (such as the bird flu or the swine flu). Furthermore, the analysis did not properly discount for a statistical artifact deriving from the way Google had collected the data (recording the search key selected after auto-complete suggestions by Google). This was amplified by changes in the Google auto-completion algorithm, aimed at presenting recommendations for flu treatments to users searching for typical flu symptoms (which are, however, compatible with common colds) (Lazer et al. 2014). The latter is an exemplary case of a platform effect (Malik and Pfeffer 2016), introducing biases in the data.[10]

There are two main lessons which can derived from this case. First, Google algorithms are optimized to serve Google's goals as a company, not to collect information relevant to epidemiology. Second, the fact that Google kept its algorithm secret and Google researchers had access to non-public data (Ruths and Pfeffer 2014), by virtue of protecting the algorithm from external scrutiny, led to errors of greater proportions.

Let us now consider Google's data assets as a digital phenotype, that can only be produced thanks to the company and its successful product, but which can also affect the life of anyone affected by better public health measures. This view suggests that

[10] Another example of platform effect is the "People You May Know" function in Facebook, which caused a spike in the observed rate of triadic closures (the phenomenon by which two nodes in a network are more likely to establish a link between each other if they each already share a link with another node) in Facebook friendships (Zignani et al. 2014). Platform effects are especially important for the epistemology of online sociology but they are problematic also for epidemiology.

Google cannot focus solely on generating profit for its shareholders. Google has—it may then be argued—an ethical obligation to all persons potentially affected by its digital phenotype. This obligation can be conceptualized as a form of corporate social responsibility (Carroll 1991; Heal 2005); in this way—given Google's favored position in the rank of internet powers—the idea of *noblesse oblige* is carried over to the ethics of big data regulation. In response to the problems analyzed here, large companies should strive, for example, to make data more openly accessible to researchers who could build the appropriate algorithms, so as to enhance the beneficial value of the data.

Once again, notice that this ethical issue is not a concern from the standpoint of data protection or of the (libertarian) data ownership framework: first, the digital epidemiology uses anonymous data (e.g., frequency of aggregate searches of the word "flu") where the likelihood of re-identifying individuals is very low, so the first is irrelevant; second, making data usable for public health is not something that Google owes to data co-generators (those using its search engine to make flu searches), but to all the persons potentially benefitted or harmed by this knowledge.

5 Conclusions

This essay contributes to philosophy and ethics by conceptualizing digital data as an extended phenotype and exploring the ethical implications of this idea. The digital phenotype is not an individual passive entity, as suggested by the digital footprint metaphor. It is a socially created causal factor in an evolutionary feedback loop, that can harm or benefit humans, by influencing their beliefs, behaviors, and norms.

The data protection and ownership frameworks assign a special moral importance, respectively, to identifiable individuals and those contributing data. By contrast, ethical reflection on the digital phenotype should highlight the risks and potential benefits of digital information for humans in general, including individuals and future generations, who are not identifiable in the data and did not contribute to it.

Admittedly, Section 4 does not provide ready-made solutions for the governance of the digital phenotype. As shown by the case study of P4 medicine, identifying groups potentially affected by generalizable knowledge can be challenging in this context. Empirical sociology and informatics can help society to build methods to identify the at-risk groups and predict discrimination risks. Research in ethics and political philosophy should contribute to clarifying the balance between the social benefits and ills of generalizable knowledge; empirical testing and social experiments may lead to governance frameworks responsive to such concerns. A broader conversation with stakeholders, including but not limited to the clients of large Internet companies, is needed to define the corporate social responsibility of large generators of digital phenotypes.

References

Aicardi, C., Savio Del L, Dove, E. S., Lucivero, F., Mittelstadt, B., Niezen, M., Prainsack, B., Reinsborough, M., and Sharon, T. (2016a). Shortcomings of the Revised 'Helsinki Declaration' on Ethical Use of Health Databases. The Hastings Center. November 2, 2016. http://www.thehastingscenter.org/shortcomings-world-medical-associations-revised-declaration-ethical-use-health-databases/.

Aicardi, C., Del Savio, L., Dove, E. S., Lucivero, F., Tempini, N., & Prainsack, B. (2016b). Emerging ethical issues regarding digital health data. On the world medical association draft declaration on ethical considerations regarding health databases and biobanks. *Croatian Medical Journal, 57*(2), 207–213.

Beauchamp, T. L., & Childress, J. F. (1994). *Principles of biomedical ethics* (4th ed.). USA: Oxford University Press.

Boyd, D., & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society, 15*(5), 662–679.

Buchanan, A. (2007). Institutions, beliefs and ethics: Eugenics as a case study. *Journal of Political Philosophy, 15*(1), 22–45.

Carroll, A. B. (1991). The pyramid of corporate social responsibility: toward the moral management of organizational stakeholders. *Business Horizons, 34*(4), 39–48.

Cavalli-Sforza, L., & Feldman, M. W. (1981). *Cultural transmission and evolution: a quantitative approach*. Princeton: Princeton University Press.

Cohen, G. A. (1995). *Self-ownership, freedom, and equality*. Cambridge: Cambridge University Press.

Danaher, J., Hogan, M. J., Noone, C., Kennedy, R., Behan, A., De Paor, A., Felzmann, H., et al. (2017). Algorithmic governance: developing a research agenda through the power of collective intelligence. *Big Data & Society, 4*(2), 2053951717726554.

Dawkins, R. (1999). *The extended phenotype: the long reach of the gene* (2nd ed.). Oxford: Oxford University Press.

European Commission DG Health. (2014). *The Use of Big Data in Public Health Policy and Research*. Brussels: European Commission Directorate General for Health an Consumers eHealth and Health Technology Assessment.

Eysenbach, G. (2009). Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the internet. *Journal of Medical Internet Research, 11*(1), e11.

Floridi, L. (2011). The informational nature of personal identity. *Minds and Machines, 21*(4), 549–566.

Floridi, L. (2014). Open data, data protection, and group privacy. *Philosophy & Technology, 27*(1), 1.

Floridi, L. (2016a). On human dignity as a foundation for the right to privacy. *Philosophy & Technology, 29*(4), 307–312.

Floridi, L. (2016b). Group Privacy: A Defence and an Interpretation. In L. Taylor, L. Floridi, & B. van der Sloot (Eds.), *Group Privacy: New Challenges of Data Technologies* (pp. 83–100). Cham, Springer.

Freelon, D. (2014). On the interpretation of digital trace data in communication and social computing research. *Journal of Broadcasting & Electronic Media, 58*(1), 59–75.

Griffin, J. (2008). *On human rights*. Oxford: Oxford University Press.

Haraway, D. J. (2008). *When species meet*. Minneapolis: University of Minnesota Press.

Heal, G. (2005). Corporate social responsibility: an economic and financial framework. *The Geneva Papers on Risk and Insurance - Issues and Practice, 30*(3), 387–409.

Hood, L., Lovejoy, J. C., & Price, N. D. (2015). Integrating big data and actionable health coaching to optimize wellness. *BMC Medicine, 13*, 4.

Jablonka, E., & Lamb, M. J. (2005). *Evolution in four dimensions: genetic, epigenetic, behavioral, and symbolic variation in the history of life*. Cambridge: MIT Press.

Jain, S. H., Powers, B. W., Hawkins, J. B., & Brownstein, J. S. (2015). The digital phenotype. *Nature Biotechnology, 33*(5), 462–463.

Jouquet, P., Dauber, J., Lagerlöf, J., Lavelle, P., & Lepage, M. (2006). Soil invertebrates as ecosystem engineers: intended and accidental effects on soil and feedback loops. *Applied Soil Ecology, 32*(2), 153–164.

Kelion, L. 2017. Facebook Uses AI to Spot Suicidal Users. BBC News, March 1, 2017, sec. Technology. http://www.bbc.com/news/technology-39126027.

Kitcher, P. (2001). *Science, truth, and democracy*. Oxford: Oxford University Press.

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google flu: traps in big data analysis. *Science, 343*(6176), 1203–1205.

Lupton, D. (2015). Health promotion in the digital era: a critical commentary. *Health Promotion International, 30*(1), 174–183.

Lupton, D. (2016a). Digital companion species and eating data: implications for theorising digital data–human assemblages. *Big Data & Society, 3*(1), 1–5.

Lupton, D. (2016b). Foreword: lively devices, lively data and lively leisure studies. *Leisure Studies, 35*(6), 709–711.

Malik, M M., Pfeffer, J. (2016). Identifying Platform Effects in Social Media Data. In *Tenth International AAAI Conference on Web and Social Media*. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM16/paper/view/13163.

Mittelstadt, B. (2017a). Designing the health-related Internet of things: ethical principles and guidelines. *Information* 8 (3). http://www.mdpi.com/2078-2489/8/3/77htm.

Mittelstadt, B. (2017b). Ethics of the health-related Internet of things: a narrative review. *Ethics and Information Technology, 19*(3), 157–175.

Mittelstadt, B., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: mapping the debate. *Big Data & Society, 3*(2), 2053951716679679.

National Academy of Sciences. (2011). *Toward precision medicine: nuilding a knowledge network for biomedical research and a new taxonomy of disease*. Washington D.C.: National Academies Press.

Nozick, R. (1974). *Anarchy, state, and utopia*. New York: Basic Books.

Perbal, L. (2013). The 'warrior gene' and the Māori people: the responsibility of the geneticists. *Bioethics, 27*(7), 357–410.

Prainsack, B., & Buyx, A. (2013). A solidarity-based approach to the governance of research biobanks. *Medical Law Review, 21*(1), 71–91.

Regulation on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation 2016/ 679). (2016). http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2016.119.01.0001.01.ENG&toc=OJ:L:2016:119:TOC.

Relton, C. L., & Smith, G. D., (2010). Epigenetic epidemiology of common complex disease: prospects for prediction, prevention, and treatment. *PLoS Med, 7*(10), e1000356.

Ruths, D., & Pfeffer, J. (2014). Social media for large studies of behavior. *Science, 346*(6213), 1063–1064.

Salathé, M., & Khandelwal, S. (2011). Assessing vaccination sentiments with online social media: implications for infectious disease dynamics and control. *PLoS Computational Biology, 7*(10), e1002199.

Salathé, M., Bengtsson, L., Bodnar, T. J., Brewer, D. D., Brownstein, J. S., Buckee, C., Campbell, E. M., et al. (2012). Digital Epidemiology. *PLoS Computational Biology, 8*(7), e1002616.

The Onlife Initiative (2015). The onlife manifesto. In L. Floridi (Ed.), *The onlife manifesto*. Cham: Springer.

Vayena, E., Salathé, M., Madoff, L. C., & Brownstein, J. S. (2015). Ethical challenges of big data in public health. *PLoS Computational Biology, 11*(2), e1003904.

Vayena, E., Dzenowagis, J, Langfeld, M. (2016). The Health Data Ecosystem and Big Data. WHO. 2016. http://www.who.int/ehealth/resources/ecosystem/en/.

Widdows, H. (2013). *The connected self: the ethics and governance of the genetic individual*. Cambridge: Cambridge University Press.

Winickoff, D. E., & Neumann, L. B. (2005). Towards a social contract for genomics: property and the public in the 'biotrust' model. *Genomics, Society and Policy, 1*(3), 8–21.

Winickoff, D. E., & Winickoff, R. N. (2003). The charitable trust as a model for genomic biobanks. *New England Journal of Medicine, 349*(12), 1180–1184.

WMA - The World Medical Association. (2016). WMA Declaration of Taipei on Ethical Considerations Regarding Health Databases and Biobanks. Website of the WMA. October 22, 2016. https://www.wma.net/policies-post/wma-declaration-of-taipei-on-ethical-considerations-regarding-health-databases-and-biobanks/.

Zignani, M, Gaito, S, Rossi, G P, Zhao, X, Zheng, H, Zhao, B Y. (2014). Link and triadic closure delay: temporal metrics for social network dynamics. In Eighth International AAAI Conference on Weblogs and Social Media. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8042.