

Chapter 1

The Ethics of Digital Well-Being: A Multidisciplinary Perspective

Christopher Burr and Luciano Floridi

Abstract This chapter serves as an introduction to the edited collection of the same name, which includes chapters that explore digital well-being from a range of disciplinary perspectives, including philosophy, psychology, economics, health care, and education. The purpose of this introductory chapter is to provide a short primer on the different disciplinary approaches to the study of well-being. To supplement this primer, we also invited key experts from several disciplines—philosophy, psychology, public policy, and health care—to share their thoughts on what they believe are the most important open questions and ethical issues for the multi-disciplinary study of digital well-being. We also introduce and discuss several themes that we believe will be fundamental to the ongoing study of digital well-being: digital gratitude, automated interventions, and sustainable co-well-being.

Keywords Artificial intelligence · Automated interventions · Digital ethics · Digital well-being · Sustainable design

1.1 Introduction

Recently, digital well-being has received increased attention from academics, technology companies, and journalists (see Burr et al. 2020a, b). While a significant amount of this interest has been focused on understanding the psychological and social impact of various digital technologies (e.g. Orben and Przybylski 2019), in other cases the interest has been much broader. For instance, the International Network for Government Science Advice (INGSA)—a forum that advises on how scientific evidence can inform policy at all levels of government—claims that

C. Burr (✉) · L. Floridi
Oxford Internet Institute, University of Oxford, Oxford, UK

The Alan Turing Institute, London, UK
e-mail: cburr@turing.ac.uk

“to understand wellbeing in the 21st century requires an understanding of transformative digital technologies as drivers of change not just in human material circumstances, but also in *human values* and *organisational systems* that support wellbeing” (Gluckman and Allen 2018, p.10). The digital transformation of society, it seems, requires a more thorough investigation of how our conceptual understanding of well-being may have been altered by emerging technologies and the new *modes of being* they enable.

One may rightfully wonder why this surge in interest has happened now. After all, digital technologies have been around for decades and our well-being has been dependent on technology for far longer (Floridi 2014). What, if anything, is different this time around? The short answer is that the function, use, effects, and even experience of digital technologies has been altered significantly by the widespread implementation of ubiquitous computing (e.g. wearables, smartphones), machine learning, and more recently artificial intelligence (AI). These technological developments have resulted in drastic changes to our environment, including social domains such as healthcare, education, employment, policy, and entertainment, and have also been accompanied by drastic shifts in media consumption and lifestyle habits (Ofcom 2018). Combined, these developments are exposing humans to an environment that is increasingly *adaptable to them*, either as individuals or as members of segmented groups, by monitoring and analysing digital traces of their interactions with intelligent software agents (Burr et al. 2018). This is an important shift. Whereas humanity has refined its ability to engineer and reconstruct its environmental niche over the course of our evolutionary history (Sterelny 2003), we are now at a stage where the design and construction of our environmental niche can be automated, thereby reducing the need for human agency and oversight. For example, recommender systems, due to their ability to operate at scale and speed, are deployed to control the architecture of our online environments, making split-second decisions about the design elements of web pages (e.g. colour of fonts), placement and content of links (e.g. advertisements), appropriate pricing for products (e.g. dynamic pricing of holiday packages), and much more (Milano et al. 2020). Such a change is unprecedented and demands that we consider the ethical implications for our individual and social well-being. This is the primary purpose of this edited collection: to explore the ethics of digital well-being from a multi-disciplinary perspective, in order to ensure that the widest possible aperture is employed without losing focus on what matters most.

The purpose of this introductory chapter, more specifically, is to provide an informative foundation to ground and contextualise the subsequent discussion, while also offering some initial suggestions about where to head next. That said, we do not consider it necessary to provide a precise definition or theory of ‘digital well-being’ that can serve as a universal placeholder for each of the subsequent chapters. This would be inappropriate for a number of reasons. First, as a multi-disciplinary (and often interdisciplinary) collection, each chapter will emphasise different aspects of digital well-being, conditional on the explanatory goal they wish to achieve. Second, it is unclear at present whether we need a new concept of ‘digital well-being’ that is distinct from ‘well-being’ in a meaningful way. And, finally, the

purpose of this collection is to generate further interdisciplinary interest in the topic, in the hope that greater conceptual clarity may arise from subsequent discussions. Therefore, for present purposes, ‘digital well-being’ can be treated as referring loosely to the project of studying the impact that digital technologies, such as social media, smartphones, and AI, have had on our well-being and our self-understanding of what it means to live a life that is *good for us* in an increasingly digital society.

While the above outline may serve as a sufficient placeholder for general discussion, there is obviously a risk of it leading to some conceptual confusion. For example, a philosopher could rightfully ask what explanatory or enumerative role the concept offers, over and above ‘well-being’ *simpliciter*. Does the restricted domain, entailed by the inclusion of ‘digital’, offer any useful theoretical constraints, or does it merely impede the philosophical pursuit of identifying the most general conditions for well-being? In addition, psychology and economics have, in recent years, developed new tools that are designed to measure the subjective well-being of individuals or the socioeconomic indicators that are treated as proxies of social well-being. What explanatory or prescriptive role would the concept of ‘digital well-being’ serve in these disciplines? Could it be employed as a theoretical construct to be measured by a range of psychometric tests? Could it offer any useful theoretical constraints to assist with the selection of relevant socioeconomic indicators of social well-being, and in turn help to guide policy decisions that seek to improve human capabilities in an increasingly digital society? These questions require careful consideration, ongoing scrutiny, and thoughtful discussion, and we hope that the current collection may offer a rich starting point for answering them.

This collection includes chapters that explore digital well-being from a range of disciplinary perspectives, including philosophy, psychology, economics, health care, and education. Because of this broad focus, Sect. 1.2 provides a short primer that serves as an introduction for those readers who may be approaching the topic of digital well-being from a particular disciplinary perspective. In each of these subsections the reader will also find a short commentary from invited experts who were asked to provide their own views on how they think their respective disciplines may be affected by ongoing technological innovation (e.g. novel research methodologies, new means to test empirical hypotheses, impacts on policy-making), as a way of pointing to further areas of research for the interested reader. Following this, Sect. 1.3 introduces several themes that we believe will be fundamental to the ongoing study of digital well-being: digital gratitude, automated interventions, and sustainable co-well-being. These topics are not intended to be exhaustive or representative of the literature (see Burr et al. 2020a, b for a more detailed review). Rather, they have been chosen in part because of the connection they have to some key ideas in other chapters. What they offer is merely some initial ideas that are intended to be, in conjunction with the subsequent chapters, a platform and guide for further discussion. Therefore, we hope that this collection as a whole will provide an informative starting point for readers from different disciplines interested in the study of well-being, while also contributing to what we expect will be an exciting and interdisciplinary pursuit of ensuring humanity can flourish in this new digital environment.

1.2 Theories of Well-Being: A Short Primer

Theoretical statements about well-being are typically understood as making either a *descriptive* claim (e.g., whether the implementation of a socio-economic policy typically enhances or decreases some quantifiable measure of well-being), or a *normative* claim (e.g., an evaluation of the goodness or badness of some moral action with regards to whether it maximises welfare). Although this can be a useful heuristic for assessing the nature of a particular well-being claim, it is also conceptually problematic. As Alexandrova (2017, p. xv) argues, empirical (descriptive) claims about well-being rely on an inseparable normative standard: “any standard or method of measurement of well-being is already a claim about the appropriateness of an action or state in the light of some assumed value.” For example, if a policy-maker states that an economic policy (e.g., increasing funding for education) is highly correlated with some measure of social well-being, their descriptive claim is also mixed with a normative element (i.e., increasing funding for education ought to be done to increase social welfare). This is why the study of well-being is an inherently interdisciplinary task.

Many disciplines, including philosophy, psychology, design engineering, economics, law, medicine, and sociology are concerned with well-being, and each discipline has its own distinct theoretical framework. Therefore, it is important to understand the commonalities and differences between the various theoretical perspectives because any digital technology that claims to be promoting or protecting well-being must at the very least implicitly presume some general account of what it is for a life to go well for an individual. By introducing some of the major theoretical perspectives, we will be able to specify more clearly what is at stake. Readers who are already familiar with the general issues in a particular discipline should feel free to skip over the relevant section.

1.2.1 Philosophy

Philosophy has a long tradition of seeking to understand the concept of ‘well-being’, including its relationship with other important ethical concepts, such as ‘reason’ or ‘goodness’. A standard view is that ‘well-being’ refers to what is *non-instrumentally good for a subject S* (Crisp 2006; Woodard 2013). This notion is used to separate that which is intrinsically (i.e., non-instrumentally) good for a person—sometimes referred to as ‘prudential value’—from that which is merely good because of its instrumental role in leading to a greater level of well-being (e.g., income, employment, or social network). Therefore, a fully developed philosophical theory of well-being is concerned both with *enumerating* those things that are non-instrumentally good for someone (e.g., a mental state such as pleasure, or desire-satisfaction) and also *explaining* why the individual ought to pursue and promote the respective good (Lin 2017; Crisp 2006; Tiberius 2015). These two theoretical objectives can come apart, such that there can be agreement between two theories regarding the

enumerated goods for a particular theory (e.g., friendship) but disagreement concerning the reasons *why* these goods have prudential value (e.g., friendship satisfies an informed desire or fulfils an important part of our nature). Although it has been extended or challenged over the last couple of decades (Haybron 2008; Woodard 2013; Sumner 1996), a (simplified) typology for well-being theories, famously introduced by Derek Parfit (1984), can help organise the various philosophical theories of well-being into hedonistic theories, desire-fulfilment theories, and objective list theories. This typology is sufficient for our present purposes.

Hedonistic theories claim that all that matters for well-being is the amount of pleasure and pain experienced by an individual, either at some point in time or over the course of their life. Different theories may diverge on how these states should be measured (i.e. their hedonic level) but will agree that more pleasure is good and more pain is bad. According to hedonists, if activities or objects such as music, love, food, or expressions of gratitude are good for us (the enumerative component), it is *in virtue of* their bringing about mental states such as pleasure and avoiding mental states such as pain (the explanatory component).

Desire-fulfilment theories claim that it is good for us to get what we desire, and conversely, if our desires remain unfulfilled or frustrated this will lead to a decrease in our well-being. As with the other two theories, micro-debates exist within this class of theories that try to deal with a variety of possible objections. For example, desire-fulfilment theories are often objected to on the basis that the fulfilment of certain desires (e.g., the desire to stream one more television show rather than reading a book, or to eat processed meat rather than a healthier plant-based alternative) clearly leads to a diminished level of well-being. As such, desire-fulfilment theorists will seek to make the initial claim more precise and may argue that only those desires that are informed (i.e. held on the basis of rational deliberation and relevant evidence) should be considered.

Whereas desire-fulfilment and hedonistic theories make reference to subjective attitudes that an individual possesses, objective list theories claim that well-being is constituted by *some* list of goods that are prudentially valuable irrespective of the attitude that an individual may hold towards them. Aside from this feature of *attitude-independence*, as Fletcher (2016) labels it, the list of non-instrumental goods may have little in common. They could simply be a diverse list including goods such as achievement, friendship, pleasure, knowledge, and virtue, among others.

Each of the above classes of theories is home to a series of micro debates, e.g., whether the process of obtaining some good must be experienced by the subject to entail an improvement in their overall well-being. These debates are a worthwhile theoretical enterprise but need not concern us for our present purposes. Moreover, in recent years, philosophers have focused on how it may be possible to integrate the various disciplines that study well-being in order to show how they can collectively contribute to an increased understanding of well-being (Alexandrova 2017; Bishop 2015; Haybron 2008). For example, Bishop states that we should begin with the assumption that “both philosophers and scientists are roughly right about well-being, and then figure out what it is they’re *all* roughly right about” (2015, p. 2). Psychology, as we will see in the next sub-section, is one of these sciences.

Philosophy and Digital Well-Being

Guy Fletcher (University of Edinburgh)

I understand ‘digital well-being’ to mean the impact of digital technologies upon well-being as opposed to some specific dimension of well-being (for an introduction to philosophy and well-being generally, see Fletcher 2016).

There is a rich seam of work at the intersection of politics, philosophy, and journalism on the ways in which social media, big data, and the like function to undermine democratic institutions. I will leave this, very interesting, work to one side to focus on philosophical work that concerns the direct impact of digital technologies upon individual well-being.

Philosophers are interested in the myriad ways that digital technologies can promote or undermine well-being. One major focus of attention has been social media and the way in which social media impacts *friendship*, an important prudential good (whether instrumental or intrinsic). Social media creates new categories of purported friendship (‘Facebook friends’), makes it possible to make and sustain purely online relationships, and also has the capacity to affect our real-world friendships in ways that might be positive or negative for well-being (e.g. Elder 2014; Fröding and Peterson 2012; Jeske 2019; Sharp 2012; Vallor 2012). Philosophers are also interested in the way in which digital technologies such as social media impact upon the construction and expression of our personalities (e.g. Garde-Hansen 2009; Stokes 2012).

Digital technologies are also philosophically significant in their ability to affect our powers, capacities and virtues. Recent philosophical work has examined the weakening of our powers of *attention* in a world of endless, readily-available, digital distraction, and the interaction between technology and the virtues (e.g. Williams 2018; Vallor 2016). One live question is whether it is possible to use or amend the technology itself to reduce its attention-grabbing nature. Another more squarely philosophical question is whether we can equip ourselves with powers and capacities to mitigate the attention-hogging effects of digital technologies, by developing specific virtues of attention and the like (e.g. Vallor 2016).

Biography

Dr. Guy Fletcher is senior lecturer in philosophy at the University of Edinburgh. His work examines the nature of moral discourse, philosophical theories of well-being, and theories of prudential discourse. He edited the *Routledge Handbook of Philosophy of Well-Being* (2016) and co-edited *Having It Both Ways: Hybrid Theories in Meta-Normative Theory* (Oxford University Press 2014). He is author of *An Introduction to the Philosophy of Well-Being* (Routledge 2016) and has another book, *Dear Prudence*, forthcoming with Oxford University Press.

1.2.2 Psychology

Well-being has become an important indicator of progress for many governments around the world, thanks in part to empirical research that has shown it to be associated with a range of positive outcomes such as “effective learning, productivity and creativity, good relationships, pro-social behaviour, and good health and life expectancy” (Huppert and So 2013). Unlike philosophy, the behavioural and cognitive sciences—including psychology—are less concerned with whether these goods are non-instrumentally valuable, but rather with what causes them to fluctuate and how best to measure them.

To understand the current theoretical focus of psychological theories of well-being, it is worth mentioning the emergence of *positive psychology*. Positive psychology emerged as a distinct disciplinary enterprise at the turn of the century. Writing in 2000, Seligman and Csikszentmihalyi stated that, “[p]sychology has, since World War II, become a science largely about healing. It concentrates on repairing damage within a *disease model* of human functioning” (Seligman and Csikszentmihalyi 2000, p. 5, emphasis added). This disease model assumed that well-being arose from the removal of mental disorders such as depression, and thus required no separate study or distinct methodology of its own. Positive psychology rejected this model and instead sought to reorient psychological science towards a better understanding of valuable subjective experiences in their own right (e.g., happiness, contentment, or satisfaction). Its goal was to determine which environmental features are needed to achieve an *optimal level of human flourishing* for individuals and communities.

To achieve this goal, it was necessary to establish a distinct set of theoretical tools which could be used to measure and validate various psychological constructs that constitute well-being. Perhaps the most famous of these scales is *subjective well-being* (SWB), which comprises three components: frequent positive affect, infrequent negative affect, and an evaluation of the subject’s ‘satisfaction with life’ (Diener et al. 1985). The assessment of SWB typically relies on self-report (i.e. answers given by an individual in response to a question and on the basis of introspection), and because of this reliance the measurement of SWB can be affected by a range of cognitive or memory biases that impact an individual’s ability to accurately recall and report on the subjective experience being assessed (e.g. frequency of positive emotions). Methods such as experience sampling (Csikszentmihalyi 2008) have improved the reliability of SWB measures, by allowing researchers to deliver near real-time assessments of an individual’s experience through notifications that prompt users to reflect on their well-being at specific times of the day and during different activities, providing what is sometimes referred to as ‘ecologically-valid data’. More recently, suggestions to extend these methodologies by leveraging advances in ubiquitous computing have been proposed (Reeves et al. 2019).

SWB is widely assumed to be multidimensional, but there is disagreement over just how many dimensions (or factors) to include. Huppert and So (2013), for example, argue that ten factors are needed: competence, emotional stability, engagement,

meaning, optimism, positive emotion, positive relationships, resilience, self-esteem, and vitality. In contrast, Ryff (1989) claims that only six factors are needed: autonomy, environmental mastery, personal growth, positive relationships, purpose in life and self-acceptance. In spite of these disagreements, there is often significant overlap between different theories, and many often rely on the same psychometric scales for measuring subjective well-being (e.g. Satisfaction With Life, Positive and Negative Affect Scale). We can separate the various psychological theories into two groups: *hedonic* and *eudaimonic*.

Similar to philosophical hedonism, hedonic psychology claims that well-being consists of subjective experiences of pleasure or happiness, and can include “the preferences and pleasures of the mind as well as the body” (Ryan and Deci 2001, p. 144). Eudaimonic psychology, by contrast, claims that well-being consists of achieving one’s potential, as determined by human nature. According to eudaimonic psychology, human flourishing occurs when “people’s life activities are most congruent or meshing with deeply held values and are holistically or fully engaged” (Ryan and Deci 2001, p. 146).

These theoretical perspectives are often broadly characterised and can encompass a wide variety of different theories within their scope. For example, self-determination theory (SDT) is characterised as a eudaimonic theory (Ryan and Deci 2001). Briefly, SDT is a theory of human motivation and personality that is concerned with identifying the basic psychological needs of human individuals as well as the environmental conditions that are required to supply people with the nutrients to thrive and grow psychologically (Ryan and Deci 2017). SDT identifies three basic needs (competence, autonomy, and relatedness), which must be satisfied for an individual to experience an ongoing sense of psychological integrity and well-being. It is discussed further in Chap. 2 of this collection (Supporting human autonomy in AI systems: A framework for ethical enquiry), in a contribution from Rafael A. Calvo, Dorian Peters, Karina Vold, and Richard M. Ryan.

Psychology and the Study of Digital Technologies

Amy Orben (University of Cambridge)

Psychologists are becoming increasingly involved in the study of novel technologies like social media. With the field’s focus being mainly on the individual, much of the work has examined digital technology’s effect on people’s well-being, cognition or behaviour (e.g. Burke and Kraut 2016). This research has routinely taken a broad view: examining the use of digital technologies as a whole, and trying to quantify how this affects the whole population or certain broad sections of society. Yet the diversity of digital technology uses and users might be the crucial aspect missing in current psychological investigations.

(continued)

A lot of research has been done examining correlations or simple longitudinal relations between ‘screen time’ and general well-being outcomes, yet little concrete results have been found (Orben and Przybylski 2019; Jensen et al. 2019). We now know that increased time spent on digital technologies is routinely correlated with decreased well-being, but it is unclear whether this tiny correlation is causal or influential (Orben et al. 2019; Ferguson 2009). These issues are compounded by the low transparency of work done in the area, especially in the light of the recent replication crisis and open science movements (Munafò et al. 2017).

The next years will see more and more psychologists moving away from general ‘screen time’ to using more digital tracking and fine-grained digital usage data—if such data is provided by the companies that hold them (Ellis et al. 2019). They could then examine how specific uses of technologies might affect certain cognitions (e.g. self-comparison), which could in turn affect well-being (Verduyn et al. 2017). Furthermore, psychologists are increasingly integrating more robust and transparent research methods into their work, while also acknowledging that in-depth longitudinal studies will be needed to tease apart the cause-and-effect relationships that the public and policy are so interested in. Such work would ultimately allow researchers to come closer to understanding whether the increased use of digital technologies causally decreases population well-being by triangulating different types of evidence, diverse study designs and various measurement methodologies (Munafò and Smith 2018; Orben 2019).

Biography

Dr. Amy Orben is College Research Fellow at Emmanuel College and the MRC Cognition and Brain Sciences Unit. Her work using large-scale datasets to investigate social media use and teenage mental health has been published in leading scientific and psychology journals. The results have put into question many long-held assumptions about the potential risks and benefits of ‘screen time’. Alongside her research, Amy campaigns for the use of improved statistical methodology in the behavioural sciences and the adoption of more transparent and open scientific practices, having founded the global *ReproducibiliTea* initiative. Amy also regularly contributes to both media and policy debate, in the UK and internationally.

1.2.3 Economics

The development of new psychological measures of well-being has also brought about changes to the socio-economic study of well-being. Welfare economics, for example, is typically concerned with the measurement of aggregate levels of well-being. Its aim is to construct a *social welfare function*, which can be used to rank order a collection of social states (e.g., the differential allocation of public resources),

and in turn help decide which of a possible set of social policies would maximise social well-being. This normative approach assumes a *preference satisfaction* view of well-being, in which rational agents are assumed to choose what is best for them and to *reveal* their preferences through overt choice behaviour (Binmore 2008). Obtaining this data at scale, however, is challenging and so surrogate indicators for national (or aggregate) well-being are often used instead.

Until recently, one of the most popular indicators of national well-being was gross domestic product (GDP) per capita. As Diener and Seligman (2004) note, this is because economic indicators of this type are “rigorous, widely available, and updated frequently, whereas few national measures of well-being exist.” In addition, increased GDP per capita is assumed to lead to an increase in the freedom of choice available to individuals, which from the perspective of the preference satisfaction view means a greater ability to maximise well-being (or utility).

However, the use of such indicators as surrogates for national well-being has been widely criticised (e.g. Stiglitz et al. 2008), most notably from approaches within development economics, which often eschew the idea of a preference satisfaction view of well-being (Nussbaum 2011). One example is the capability approach (Robeyns 2005; Nussbaum and Sen 1993). In short, the capability approach draws attention to what people are “actually able to do and to be” in their environment, rather than simply assuming that their choice behaviour reveals a stable and ordered set preferences (Robeyns 2005)—an assumption that is also heavily challenged by research in behavioural economics that focuses on cognitive biases in judgement and decision-making (Kahneman 2011). A motivating idea here is that individuals need the freedom to pursue distinct capabilities, which may include health, education, arts and entertainment, political rights, social relationships, and so on. These diverse capabilities are poorly captured by a single indicator such as GDP per capita, and so a richer framework for measuring well-being is required.

The influence of the capabilities approach can be seen in the United Nations Human Development Index and related programmes such as the Sustainable Development Goals (United Nations 2019). It also influenced a report, commissioned by the then President of the French Republic, Nicholas Sarkozy, who stated that he was “unsatisfied with the present state of statistical information about the economy and the society” and that economic progress and social development required more relevant indicators than simply GDP (Stiglitz et al. 2008). As one of their key recommendations, the commission suggested that “[m]easures of both objective and subjective well-being provide key information about people’s quality of life” and that “statistical offices should incorporate questions to capture people’s life evaluations, hedonic experiences and priorities in their own survey” (Stiglitz et al. 2008, p. 12). Chapter 9 of this collection (Big Data and Wellbeing: An Economic Perspective), by Clement Bellet and Paul Frijters, offers a helpful overview of the recent developments that have followed this recommendation, leveraging insights derived from data-driven technologies, such as machine learning.

Public Policy, Well-Being and Digital Technology

Florian Ostmann (Alan Turing Institute)

There are two prominent strands of inquiry at the intersection of well-being and digital technology that are of interest to public policy researchers and increasingly relevant to policymaking agendas.

The first strand may be referred to as *digitally derived insights about well-being*—work that leverages technology-enabled methods and big data analytics to measure well-being and understand and manage its determinants. In the context of *measuring economic welfare*, this includes the use of novel analytical techniques and unconventional data sources (e.g. electronic payments, social media, or business news data) to predict GDP growth and related indicators in real-time (Anesti et al. 2018; Galbraith and Tkacz 2018) or with greater accuracy compared to traditional approaches. It also includes the use of massive online choice experiments for welfare measurement—for instance, to estimate the economic value of zero-priced goods, which fails to be captured by measures of GDP (Brynjolfsson et al. 2019).¹ In the context of work that is dedicated to *measuring subjective well-being*, digital methods have impactful applications as well, illustrated by the use of digital surveys or novel inferential methods (e.g. sentiment analysis applied to social media activity or digitized books) to arrive at estimates of present or historical levels of subjective well-being (Hills et al. 2019). Finally, technology and data analytics can enable pathbreaking insights about specific *factors that impact well-being*—such as urban air quality, for example—improving our understanding of and ability to manage these factors (Hamelijck et al. 2019; Warwick Machine Learning Group 2019).

The second strand concerns the *well-being effects of digital technologies* (i.e., the positive or negative consequences that the adoption of relevant technologies may have for individual and societal well-being). Consequences of interest from a public policy perspective may be intrinsically related to the technology in question or be characterised by a more indirect relationship, spanning a wide range of different policy domains (OECD 2019). Correspondingly, understanding the well-being effects of digital technologies and developing policy strategies that support the realisation of benefits while managing negative effects constitutes a wide-ranging area of research. This area includes the potential of technological innovation to enable well-being-enhancing improvements in the design and delivery of goods and services, especially in essential areas where accessibility and quality improvements may be particularly impactful for disadvantaged members of society (e.g. health, education, financial services and the judicial system). It also

(continued)

¹The most prominent zero-priced goods and services are often digital goods themselves, such as search engines or social media platforms.

comprises questions around digital exclusion, concerns about the risk of certain forms of innovation rendering consumers vulnerable to exploitative commercial practices, and a growing policy debate around ‘online harms’ (e.g. disinformation, cyberbullying, encouragement of self-harm, online grooming, and access to age-inappropriate material) (Vidgen et al. 2019; UK Government 2019). Finally, there are important questions around the relationship between digital innovation and more abstract welfare-related categories of analysis including economic growth, labour market dynamics, and competition and market power.

Biography

Dr. Florian Ostmann is the Policy Theme Lead within the Public Policy Programme at The Alan Turing Institute. His work focuses on the societal implications of data science and AI, concentrating on the use of data science and AI to address governmental and social challenges and on the ethical and regulatory questions raised by their application in different sectors. As part of this work, he leads projects across a range of thematic areas, including financial services, criminal justice, and combatting modern-day slavery. Florian also has strong interests in the role of value judgments in applied economics, health policy, and questions concerning the future of work and social welfare systems.

1.2.4 Health

While conceptually distinct, health and well-being are also intimately related, and therefore some brief remarks are helpful. The World Health Organisation defines health as “a state of complete physical, mental and social *well-being* and not merely the absence of disease or infirmity” (World Health Organisation 2019), and Crisp (2017) notes that “[p]opular use of the term ‘well-being’ usually relates to health”.

In the medical sciences, as Alexandrova (2017, p. 168) notes, “the stand-in for well-being is health-related *quality of life*”. Quality of life (QOL), and related variants such as quality-adjusted life years (QALYs) or disability-adjusted life years (DALYs) (Hausman 2015), is used in similar ways to economic constructs (i.e. as an input to calculations that help to determine the efficiency of policy decisions and to allocate healthcare resources). As with psychology, the medical sciences also rely on a range of more specific measures of well-being, which can be tailored to individual diseases or patients and sometimes extend to the well-being of caregivers. Although these measures will often rely on clinical diagnosis and observable indicators, subjective evaluation and self-report is also seen within healthcare in the

form of patient-reported outcomes (e.g. Alexandrova 2017; Haybron and Tiberius 2015). We offer more detailed comments on the links between digital health technologies and digital well-being in Sect. 1.3.2.

Digital Health: The Future Role of Policy

Indra Joshi and Jessica Morley (NHSX)

In July 2019 NHSX—a new joint unit bringing together staff from NHS England, the Department of Health, and NHS Improvement—came into being. It was created to ensure the NHS benefits from the best digital health thinking from government, the NHS, academia and industry.

For too long, there has been a consistent lack of investment in digitising the health service, slow adoption of technology (it takes approximately 17 years for a new innovation to spread through the NHS (Leigh and Ashall-Payne 2019), and fear stemming from past failures such as Care.Data (Sterckx et al. 2016) and the National Programme for IT (Justinia 2017). These setbacks have left NHS staff reliant on technology that was outdated in the 90s and forced patients to turn to digital services provided by unvetted third parties, in an attempt to manage or improve their health and well-being. This situation has introduced huge opportunity costs, economic costs (Ghafur et al. 2019), and risk into healthcare systems across the globe—not just the NHS (Mackey and Nayyar 2016). It is, therefore, clearly untenable.

There is a need for a step change in the way that healthcare system providers approach digital health and well-being. Policymakers need to adopt a principled, proportionate approach to its governance (Morley and Joshi 2019)—one that is open to the significant opportunities for improving outcomes, cutting costs, and ultimately saving lives—but mindful of the clinical and ethical risks (Morley and Floridi 2019b). This requires introducing policies that ensure digital health technologies are: designed for specific users (Themistocleous and Morabito 2012); developed in the open (Goldacre et al. 2019); interoperable; thoroughly and consistently evaluated (Ferretti et al. 2019); evidence-based (Greaves et al. 2018); economically viable; clinically safe and efficacious (Challen et al. 2019); and pro-ethically designed (Floridi 2016a).

According to the World Health Organisation (2019), “Health is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity.” Clearly harnessing digital technologies and the data that they generate in the right way, is going to be essential if healthcare systems, like the NHS, want to ensure their service users are able to achieve this state of wellbeing. Thus, while developing such policies will take time, we cannot afford to wait.

(continued)

Biographies

Dr. Indra Joshi is director of AI at NHSX, and an emergency clinician by training. She is responsible for creating a trusted environment that supports innovation of data-driven technologies while being: the safest in the world; appropriately responsive to progress in innovation; ethical, legal, transparent and accountable; evidence-based; competitive and collaborative; in alignment with the NHS Constitution.

Jessica Morley is currently pursuing the M.Sc. degree with Oxford Internet Institute, Oxford, U.K. She is also the policy lead with the DataLab, University of Oxford, Oxford, and the AI subject matter expert with NHSX, London, U.K. Her research focuses on the ethical and policy implications of digitizing healthcare.

1.3 Digital Well-Being: Three Themes for Further Discussion

In this section, we offer some thoughts on three themes that overlap with some of the subsequent chapters. These themes are inherently interdisciplinary in nature, and, therefore, are good examples of why the study of digital well-being requires a multidisciplinary approach. One thing they have in common is a strong emphasis on the importance of ethical design. As such, it is helpful to begin with a few clarificatory remarks.

The following two statements are widely accepted in the ethics of technology: (1) technological design and engineering is a value-laden process, and (2) the use and implementation of technology has the potential to fundamentally alter the way we understand ourselves, each other, and our environment, which may in turn create new ethical challenges (Floridi 2010). A single example can help illustrate and justify both statements: the design and use of wearable heart-rate monitors, such as smartwatches and fitness trackers.

Starting with the first statement, a common technique that modern wearables employ for measuring heart-rate is known as photoplethysmography (PPG). PPG uses a light-emitting sensor to estimate changes in arterial volume caused by pulsating blood pressure. However, the design choice of which colour light to use (e.g. red, green, or blue) can have different consequences, some of which raise ethical concerns. For instance, green light has higher levels of accuracy than red light when the device is in *constant motion*—a common occurrence for wearables that are used for fitness activities. Therefore, if one is optimising for accuracy as determined by the context of use (i.e. fitness) this would seem like the obvious choice. However, green light has lower accuracy for darker skin tones than lighter skin tones, leading to a potential bias against certain groups of people (Woolley and Collins 2019). Therefore, the choice of which colour of light to use in PPG can be treated as a

value-laden design choice, which may favour people with lighter skin and discriminate against people with darker skin.

Turning to the second statement, a number of studies have explored how data collected from heart-rate monitors can also be used to infer psychological information, including affective states (i.e., emotions) and psychopathological states (e.g., levels of anxiety) (see Burr and Cristianini 2019 for an overview). The process of recording and measuring biometric data and converting them into user-friendly types of information can be incredibly valuable for individuals who wish to track their health and well-being. In some cases, the bio-feedback that these devices provide can even allow individuals to acquire a degree of volitional control over the heart-rate, which could help alleviate symptoms of anxiety (Abukonna et al. 2013). However, as sociologists and philosophers have noted, the rapid increase in information to quantify and measure states of our bodies can also threaten our psychological integrity by negatively impacting our self-understanding and self-awareness (Lupton 2016; Floridi 2014). To illustrate, consider how the digital representation of our psychological states or process may be in competition with the internal representations that our brains have evolved to rely upon. Emotions, for example, are formed and refined on the basis of signals that originate from within our bodies. Smart devices aim to bypass this process—known as ‘interoception’—by inferring our psychological states through a variety of techniques, some of which rely on probabilistic machine learning algorithms.² In addition to ongoing questions regarding the accuracy and validity of these measurement procedures, there is a further concern about how this information is stored and presented to the user. For instance, unlike digital representations, our inner emotional states are not perpetually recorded in discrete forms *in silico*. Rather, our emotions are typically appraisals of our current context, and provide salient information about how to act in the current environment—they are *action-guiding* (Frijda et al. 1989). In this sense, emotions have an immediacy and embeddedness that connects us to our present surroundings in ways that permanently stored digital representations do not. Digital representations, by contrast, are detachable records of past states or processes. While this allows them to store historical information for reflection, in doing so they lose the immediacy that our mental representations provide. It is possible that an increased use of digital technologies to represent our mental states or processes could alter the level of trust that we have in our own interoceptive capabilities, and may result in destabilising effects for our psychological integrity and well-being due to the altered functional role they play in guiding our behaviour and self-understanding.

These brief examples of wearable heart-rate monitors help to emphasise the importance of ethical principles in the design and development of digital technologies (for further discussion, see Floridi 2010; Calvo and Peters 2014). The following three themes, influenced by the subsequent chapters, share this emphasis on the importance of ethical design.

² See (Calvo et al. 2015) for an introduction to current techniques in affective computing.

1.3.1 *Digital Gratitude*

The majority of readers will have an instinctive understanding of the concept ‘gratitude’, recognising it as either an emotion (e.g., feeling grateful towards an individual who has helped you), a behaviour (e.g., expressing gratitude to a friend or family member), or a virtuous trait (e.g., a praiseworthy disposition of an individual). However, the conjunction of ‘digital’ and ‘gratitude’ may not elicit the same instinctive understanding.

Our use of the term ‘digital gratitude’ is intended to emphasise the mediating role that digital technologies have on the feeling, expression, or trait of gratitude. As is well understood, digital technologies are not neutral. Their design is often motivated by commercial interests, as Charlie Harry Smith highlights in Chap. 3 of this collection (*Corporatised Identities ≠ Digital Identities: Algorithmic Filtering on Social Media and the Commercialisation of Presentations of Self*), and can also be used to manipulate user behaviour, as Michael Klenk discusses in Chap. 4 (*Digital Well-Being and Manipulation Online*). This is important, because as Lavinia Marin and Sabine Roeser note in Chap. 7 (*Emotions and Digital Well-being: The rationalistic bias of social media design in online deliberations*), digital technologies, such as social media platforms, “do not mediate the full range of human emotions and thus are an impediment for successful deliberations”, signifying a key risk of digital technologies. Similarly, we can ask what impact, both positive and negative, digital technologies may have on our conceptual understanding of ‘gratitude’, as well as the emotion itself.

Gratitude is an important affective trait. It is recognised by psychologists, anthropologists, and evolutionary biologists, as an *other-directed* emotion (e.g. gratitude towards a friend, object, or state of the world), one that plays a prosocial role in communities by strengthening interpersonal relationships and generating positive behavioural norms within organisations and groups (Yost-Dubrow and Dunham 2018; Ma et al. 2017). Furthermore, gratitude is associated with higher levels of well-being—more grateful people are happier, express higher levels of life satisfaction, and also demonstrate greater levels of resilience to negative impacts on psychological well-being such as stress and burnout (Layous et al. 2017; Wood et al. 2008).

One reason that gratitude may have these benefits is because, as Allen (2018, p. 8, emphasis added) notes, “the experience of gratitude encourages us to *appreciate what is good in our lives* and compels us to *pay this goodness forward*.” Here, gratitude serves a dual role: it helps us identify and appreciate sources of prudential value (e.g. a mutually supportive online relationship with an anonymous stranger who helps an individual with difficult life challenges), and it encourages us to then increase the overall amount of prudential value by repeating the original behaviour and helping spread the feeling of gratitude to others. This latter role may also strengthen our original feelings of gratitude, generating a positive feedback loop.

Because of these benefits, digital technologies should (where relevant) be designed to promote feelings and expressions of gratitude, as well as additional

motivating psychological attitudes (see Chap. 2 of this collection). For instance, designers could introduce additional points of friction into the process of interacting with information online (e.g., sharing and reading content on social media platforms). This could allow users to reflect on how they are reacting to information, rather than just instinctively “liking” a post with little to no thought about the benefits they received from the original content. While this may reduce the amount of valuable data available to the companies (e.g., implicit feedback from user behaviour that updates recommender system algorithms), it could generate more meaningful engagement from users, further generating the perceived value of the social media platform (see Burke and Kraut 2016).

Beyond social media, designing digital technologies to promote feelings and expressions of gratitude could have additional benefits. For instance, it could help direct our attention to the intrinsic value of our digital environment and possibly generate more virtuous civic attitudes, rather than simply self-directed moral deliberation.³ By encouraging users to appreciate what is good in our lives, users may be encouraged to recognise the shared source of prudential value that is contained within the informational infrastructure that surrounds us—what we have previously referred to as the infosphere (Floridi 2014). For instance, AI offers myriad opportunities to improve and augment the capabilities of individuals and society, ranging from improved efficacy in healthcare decisions (Morley and Floridi 2019c) to identifying novel markers of social welfare in big datasets (see Chap. 9 of this collection). It is important that we a) continue to improve and augment our capabilities without reducing human control and b) continue to cultivate societal cohesion without eroding human self-determination (Floridi et al. 2018). A greater consideration of digital gratitude in the design of digital technologies could help us strike these balances, by motivating us to identify sources of prudential value, both individual and social.

However, as Andrew Gibson and Jill Willis demonstrate clearly in Chap. 8 (Ethical Challenges and Guiding Principles in Facilitating Personal Digital Reflection), the process of designing even simple gratitude enhancing technologies, such as digital self-reflective writing journals, can pose many complex and interrelated ethical challenges. Furthermore, as noted by Matthew Dennis in Chap. 6 (Cultivating Digital Well-being and the Rise of Self-Care Apps), the process of cultivating positive outcomes, such as well-being or gratitude, may sometimes generate a tension between the pursuit of the positive outcome on the one hand, and negative outcomes associated with too much screen time on the other hand. These topics are far from resolved, and we hope that this collection serves to motivate ongoing discussion and debate.

³ See (Vallor 2016; Howard 2018; Floridi 2010, Chapter 1) for a range of comments and approaches to moral virtues in the context of sociotechnical systems.

1.3.2 Automating Interventions

The final chapters in this collection discuss theoretical and conceptual issues related to the use of digital technologies for health care, starting with Chap. 10 by Amelia Fiske, Peter Henningsen and Alena Buyx (The implications of embodied artificial intelligence in mental healthcare for digital wellbeing); followed by Chap. 11 by Nick Byrd (Causal Network Accounts of Ill-being: Depression & Digital Well-being); and Chap. 12 by Michael Thornton (Malware as the Causal Basis of Disease). Many ethical challenges are intertwined with these developments, some of which are discussed in the aforementioned chapters (e.g., ensuring adequate data protection when dealing with big datasets; developing novel provisions for harm prevention). However, the broad, collective scope of these chapters also helps to draw our attention to another significant challenge: how to establish when a legitimate basis for an automated intervention has been secured. Or, to put it another way, how can we establish whether and when there is a right to intervene on the basis of an automated decision? A few clarificatory remarks are in order.

By now, it is well known that digital technologies, such as automated decision-making systems,⁴ have enabled clinical researchers and practitioners to augment their assessment, diagnostic, and treatment capabilities by leveraging algorithmically-derived insights from large-scale datasets (e.g. The Topol Review Board 2019; Watson et al. 2019; Dwyer et al. 2018; Morley et al. 2019). However, the use of these technologies outside of formal healthcare systems (e.g. in contexts such as education, employment, and financial services), and the corresponding ethical and public health challenges that arise from this deployment, is not as well appreciated (Burr et al. 2020a, b). For instance, school administrators are using predictive analytics and social media data to identify vulnerable students who may need additional support (Watson and Christensen 2017), and financial services firms have used artificial intelligence to proactively detect consumers who may experience additional financial difficulties caused by their mental health issues (Evans 2019). While these developments may lead to more proactive and personalised support, the transition away from clinical settings also raises several ethical challenges (see Palm 2013), including the question of whether there is a legitimate basis for intervening.

To understand why this is important it is helpful to contrast the clinical use of automated decision-making systems with non-clinical uses. In both cases, an automated decision can serve to establish a risk assessment or diagnosis of an individual and subsequently inform or select an intervention on the basis of these algorithmically-derived insights (e.g. nudging a user who experiences a dip in

⁴To clarify, our use of the term ‘automated decision-making systems’ is intended to be inclusive of systems that are fully automated (i.e. not requiring human oversight) and also decision support tools that keep a human-in-the-loop. Furthermore, we treat the act of classification as a decision (e.g. the classification of disease on the basis of a radiology image).

attention and engagement).⁵ This is problematic because, in clinical practice, interventions are typically decided upon following a process of *participatory decision-making* between the healthcare professional and their patient, due to the value-laden nature of health and well-being (Beauchamp and Childress 2013). Among other things, this process requires an assessment by the healthcare professional of the proportional risk associated with the intervention and the informed consent of the patient (Faden and Beauchamp 1986). It is currently unclear how automated decision-making systems should be incorporated into the process of participatory decision-making even *within* clinical settings (Morley and Floridi 2019b). Therefore, it is not possible to simply transpose existing bioethical guidance into the non-clinical settings, even as a starting point for further ethical analysis.

This is a vital and, in our opinion, unaddressed issue in the ongoing debate and discussion on informed consent in an age of big data and artificial intelligence. To clarify, informed consent in clinical decision-making is typically viewed as a *morally transformative* procedure that provides a normative justification for an act, such as a clinical intervention that carries an associated risk of harm (Kim 2019). The normative legitimacy of the informed consent process rests in part on the *professional accountability* established by formal healthcare systems and on the successful communication between healthcare professional and patient (Manson and O’Neill 2007), which in mental healthcare often requires ongoing *explanation* throughout treatment plans for chronic illnesses (e.g. depression). Accountability and explainability are, therefore, vital components of informed consent in mental healthcare but are currently poorly represented in digital health (Watson et al. 2019).

Guidelines and frameworks are currently being developed to help ensure the accountable design, development, and use of digital health tools (Henson et al. 2019; Torous et al. 2019; Jessica Morley and Floridi 2019a). However, these developments will not easily transpose into non-clinical settings where comparable mechanisms of accountability and behavioural norms are lacking (Mittelstadt 2019). While there is, in principle, no *a priori* reason to doubt that such mechanisms could be established in non-clinical settings (e.g. education, criminal justice), the contextual nature of mental health diagnosis and treatment—often emphasised by reference to the ‘biopsychosocial model’ (Burns 2014)—means that separate procedures will likely be required for each social domain where digital health tools are used to automate some part of a health intervention.

Additionally, there are also conceptual issues to address if we hope to have a robust account of what constitutes an intervention in the first place. For instance, as Michael Thornton notes in Chap. 12, novel digital technologies pose a challenge to existing conceptual accounts of ‘health’ and the boundaries of the body, as many devices can extend or augment human capabilities, thus placing pressure on our

⁵This example is based on the work of a research group at MIT’s Media Lab, who have developed a product (AttentivU) that seeks to improve attention through real-time monitoring of a user’s engagement, using a head-mounted device that use physiological sensors (i.e. electroencephalography) to measure engagement (Kosmyna et al. 2019).

existing theoretical concepts. It will be important, therefore, to develop robust accounts that are fit-for-purpose. This extends to concepts such as ‘psychological integrity’, which needs to be critically analysed if we are to make sense of the ethical significance of interventions that are more *informational* in nature (e.g. personalised recommendations for diet or lifestyle choices, or algorithmically-derived nudges).

Whereas the notion of bodily integrity is central to extant bioethical theories, comparatively less has been written about the normative status of interventions that impact an individual’s mental or psychological integrity,⁶ despite being established in Article 8 of the UK Human Rights Act 1998 and Article 3 of the EU Charter of Fundamental Rights. This is in part because of the close connection this usage has to existing theories of informed consent, personal autonomy and self-determination. However, these perspectives are insufficient when we reflect on the changing conceptual nature of concepts such as self-determination, autonomy, and informed consent in an age where the boundaries and interactions between human users and artificial agents is increasingly blurred (Floridi 2014). It will, therefore, be vital to address these conceptual, ethical, and legal challenges if we are to develop satisfactory guidelines and frameworks that can govern the use of automated interventions on individual and social health and well-being.

1.3.3 Sustainable Co-Well-Being

Derek Parfit (1984) famously offered a series of thought experiment concerning so-called “harmless torturers”, designed to query our intuitions about the possibility of imperceptible harms and benefits. We can reconstruct these thought experiments as a way to pump our intuitions about how the design of sociotechnical systems challenge ethical concepts such as ‘responsibility’, and whether a greater reflection on principles like *sustainability* can help overcome these difficulties.

First, and in line with Parfit’s original thought experiment, imagine you enter a room and see an individual strapped to a chair, connected to various pads and wires that are designed to deliver an electric current to the victim. In front of you there is a dial with numbers ranging from 1 to 1000 that controls the electric current. You turn the dial by a single increment, increasing the electrical current so slightly that the victim is unable to perceive any difference in intensity. While certainly not a morally praiseworthy action, you are unlikely to be reprimanded for causing any harm to the individual concerned. However, we now run the thought experiment for a second time, and in this alternate scenario you turn the dial by one increment at the same time as 999 other people turn similarly connected dials by one increment each. The net result of this collective action is an intensely painful electric shock

⁶A notable exception seems to be the literature on neuroethics (e.g. deep brain stimulation or direct brain interventions). However, it is more common to frame these discussions in terms of standard bioethical principles such as autonomy or informed consent (Pugh et al. 2017; Craig 2016).

that ends up killing the restrained victim. Your individual action has not changed between these two scenarios, but the relation in which your action stands to the actions of the other 999 individuals has altered drastically: you have now contributed to the death of another human being.

Next, consider the following scenario. You are waiting for a bus, tired from a long day at work. You are mindlessly scrolling through a list of possible videos that have been presented to you by a recommendation system that powers your video streaming app. You select a video of a fiery argument between two political pundits, in which one of them “destroys” their interlocutor. Ordinarily you would avoid selecting such a video, knowing that it is likely to be needlessly polarising and sensationalist. However, you’re tired and occasionally enjoy a spectacle as much as everyone else. Unfortunately, at a similar time, 999 other individuals, with similar viewing histories to yourself also click on the same recommended video. The effect is that the recommendation system learns that users similar to yourself and the 999 other individuals are likely to click on videos of this nature towards the end of the day. As such, in the future it will be more likely to recommend similarly low-quality, politically polarising videos to other users. While not as harmful, or morally reprehensible, as the death of an individual, this example nevertheless demonstrates that certain technologies—whether electric chairs or recommendation systems—have the potential to alter the moral status of our actions when they stand in a particular relation to the complementary actions of other individuals. However, what’s the particular lesson for digital well-being that we should draw from this example?

To begin, it is important to avoid the false charge that we are merely suggesting individual users must take greater responsibility for their actions online. The actions of the 1000 users impact subsequent recommendations in virtue of how the recommender system’s architecture is designed.⁷ Therefore, while the users do have a responsibility for their actions, it is a collective responsibility (similar in nature to Parfit’s harmless torturers) that emerges as a result of the interactions between the users and the system’s architecture. These interwoven interactions form complex sociotechnical systems, which connect human users and constrain their actions in important ways, leading to what we have elsewhere described as a form of ‘distributed moral responsibility’ (Floridi 2016b).

While it is immensely challenging to foresee the consequences and emergent effects of complex sociotechnical systems like recommender systems, ethical principles can serve as deliberative prompts for thinking through the ethical challenges. As such, they can offer designers a dual-advantage of identifying key opportunities to increase social value, while anticipating and avoiding costly risks (Floridi et al. 2018). One such principle we wish to propose is the need to orientate the design of digital technologies towards *sustainable* and *communal* well-being (hereafter, ‘sustainable co-well-being’). While this principle needs explaining and unpacking, the

⁷Technically, this is known as ‘collaborative filtering’, which is a method for using the collaborative actions of users (e.g., which videos they watch, how long they watch them for, and what rating they give them) as ‘implicit feedback’ to train a recommender system (see Burr et al. 2018; Milano et al. 2020 for further discussion).

focus on ‘well-being’ should be straightforward. As noted in Sect. 1.2, well-being is an intrinsic good at which much of human behaviour is directed. Therefore, although there is disagreement about what objects, activities, or states of the world bear prudential value, that well-being is a goal in itself and not merely as an instrumental means to other goods is relatively uncontroversial. To put it another way, well-being (or “the good life”), regardless of how it is understood at a subjective level, is an intrinsic good that system design should orientate towards.⁸ The ‘communal’ and ‘sustainable’ aspects of this principle requires a bit more explanation.

As we learn more about the relationship between well-being and digital technologies, the role played by *design* is emerging as crucial, not just in terms of individual fulfilment, but also, and perhaps even more significantly, in terms of communal well-being, or *co-well-being*. Precisely because it may be difficult to reach final conclusions about absolute thresholds or values of digital well-being, strategies to rectify and improve solutions already adopted will need to be considered as necessary. For part of any form of well-being consists in knowing that its erosion may not be irreversible. And the socialisation of well-being, increasingly stressed by its dependence also on digital technologies, will emphasise the socio-political aspects of co-well-being in ways probably unprecedented. In a world so connected, globalised, and mutually dependent, no discourse on digital well-being will be reasonable by focusing on individuals in isolation. Similar thoughts on these topics are explored by Loi in Chap. 5 (What contribution can philosophy provide to studies of digital well-being?) where aspects of co-well-being are present in the different concepts of digital well-being that are explored towards the end of his contribution in this collection.

Other frameworks and accounts have already emphasised the ethical significance of sustainability in design (see Floridi et al. 2018; Jobin et al. 2019). In addition, Calvo et al. in Chap. 2 of this collection note the following, regarding sustainable or circular design, “Just as we need to design in ways that preserve the natural environment for our survival, digital technologies, like YouTube, need to be designed in ways that minimise negative impact on individuals and societies to preserve a ‘sustainable’ *social* environment.” The sustainability part of our principle is similarly intended as a deliberative prompt to direct attention to the ethical significance of various design choices, while keeping a clear goal in mind (i.e. the sustainable promotion of co-well-being). This complementary focus could help steer design choices and help strike a balance between mitigating key risks while maximising opportunities. For instance, the risks of unsustainable design could include lock-in to a system that, while valuable on the basis of some outcome measures (e.g. entertainment, revenue), may nevertheless propagate bias, entrench social inequities, or create tensions between users and designers over questions such as responsibility and accountability. By considering the need for sustainability at the outset, designers may be able to consider more agile or fluid solutions, which can adapt to shifting

⁸This point is clearly demonstrated by the widespread adoption of well-being in recent frameworks or guidelines for ethical technology design, including AI (see Floridi et al. 2018).

values that change over an individual life course and across societal shifts. It is worth remarking, in particular, that digital contexts enable forms of reversibility unknown in analogue contexts but that are often underused. Relying on our previous example regarding the political video, one may notice the unwelcome training of the recommender system due to an unfortunate synchronisation of choices and could intervene to adjust or even reverse such training. More concretely, one may imagine clicking on a YouTube video, dislike it, and “take back” the click both for oneself and for others, thus avoiding both the wrong training and the “winner take all” effect of more clicks attracting even more clicks. At the time of writing, this option is not available, but it is trivially feasible, technically speaking.

As stated at the outset of this section, these themes are intended as a starting point for further discussion. Nevertheless, we hope that they will provide an informative starting point for further engagement with the subsequent chapters in this collection. If we have been able to generate some positivity or excitement in just a small handful of researchers, we will be content (and grateful) that we have contributed to a small increase in overall well-being.

References

- Abukonna, Ahmed, Xiaolin Yu, Chong Zhang, and Jianbao Zhang. 2013. Volitional Control of the Heart Rate. *International Journal of Psychophysiology* 90 (2): 143–148. <https://doi.org/10.1016/j.ijpsycho.2013.06.021>.
- Alexandrova, Anna. 2017. *A Philosophy for the Science of Well-Being*. New York: Oxford University Press.
- Allen, Sumner. 2018. *The Science of Gratitude*. UC Berkeley: Greater Good Science Center. https://ggsc.berkeley.edu/images/uploads/GGSC-JTF_White_Paper-Gratitude-FINAL.pdf.
- Anesti, Nikoleta, Ana Beatriz Galvão, and Silvia Miranda-Agrippino. 2018. ‘Uncertain Kingdom: Nowcasting GDP and Its Revisions.’ 764. Bank of England Staff Working Paper. <https://www.bankofengland.co.uk/working-paper/2018/uncertain-kingdom-nowcasting-gdp-and-its-revisions>.
- Beauchamp, Tom L., and James F. Childress. 2013. *Principles of Biomedical Ethics*. 7th ed. New York: Oxford University Press.
- Binmore, Ken. 2008. *Rational Decisions*. 1st ed. Princeton: Princeton University Press.
- Bishop, Michael. 2015. *The Good Life: Unifying the Philosophy and Psychology of Well-Being*. New York: Oxford University Press.
- Brynjolfsson, Erik, Avinash Collis, and Felix Eggers. 2019. Using Massive Online Choice Experiments to Measure Changes in Well-Being. *Proceedings of the National Academy of Sciences of the United States of America* 116 (15): 7250–7255.
- Burke, Moira, and Robert E. Kraut. 2016. The Relationship Between Facebook Use and Well-Being Depends on Communication Type and Tie Strength. *Journal of Computer-Mediated Communication* 21 (4): 265–281. <https://doi.org/10.1111/jcc4.12162>.
- Burns, Tom. 2014. *Our Necessary Shadow: The Nature and Meaning of Psychiatry*. London: Penguin.
- Burr, Christopher, and Nello Cristianini. 2019. ‘Can Machines Read Our Minds?’ *Minds and Machines*, March. <https://doi.org/10.1007/s11023-019-09497-4>.

- Burr, Christopher, Nello Cristianini, and James Ladyman. 2018. An Analysis of the Interaction Between Intelligent Software Agents and Human Users. *Minds and Machines* 28 (4): 735–774. <https://doi.org/10.1007/s11023-018-9479-0>.
- Burr, C., J. Morley, M. Taddeo, and L. Floridi. 2020a. Digital Psychiatry: Risks and Opportunities for Public Health and Wellbeing. *IEEE Transactions on Technology and Society* 1 (1): 21–33. <https://doi.org/10.1109/TTS.2020.2977059>.
- Burr, Christopher, Mariarosaria Taddeo, and Luciano Floridi. 2020b. The Ethics of Digital Well-Being: A Thematic Review. *Science and Engineering Ethics*, January. <https://doi.org/10.1007/s11948-020-00175-8>.
- Calvo, Rafael A., and Dorian Peters. 2014. *Positive Computing: Technology for Well-Being and Human Potential*. Cambridge, MA: The MIT Press.
- Calvo, Rafael A., Sidney D’Mello, Jonathan Gratch, and Arvid Kappas. 2015. *The Oxford Handbook of Affective Computing*. Oxford: Oxford Library of Psychology.
- Challen, Robert, Joshua Denny, Martin Pitt, Luke Gompels, Tom Edwards, and Krasimira Tsaneva-Atanasova. 2019. Artificial Intelligence, Bias and Clinical Safety. *BMJ Quality and Safety* 28 (3): 231–237. <https://doi.org/10.1136/bmjqs-2018-008370>.
- Craig, Jared N. 2016. Incarceration, Direct Brain Intervention, and the Right to Mental Integrity – A Reply to Thomas Douglas. *Neuroethics* 9 (2): 107–118. <https://doi.org/10.1007/s12152-016-9255-x>.
- Crisp, Roger. 2006. Hedonism Reconsidered. *Philosophy and Phenomenological Research* 73 (3): 619–645. <https://doi.org/10.2307/40041013>.
- . 2017. Well-Being. In *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta and Fall 2017. Stanford: Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/fall2017/entries/well-being/>.
- Csikszentmihalyi, Mihaly. 2008. *Flow: The Psychology of Optimal Experience*. 1st Harper Perennial Modern Classics. New York: Harper Perennial.
- Diener, Ed, and Martin E.P. Seligman. 2004. Beyond Money: Toward an Economy of Well-Being. *Psychological Science in the Public Interest* 5 (1): 1–31.
- Diener, E.D., Robert A. Emmons, Randy J. Larsen, and Sharon Griffin. 1985. The Satisfaction with Life Scale. *Journal of Personality Assessment* 49 (1): 71–75.
- Dwyer, Dominic B., Peter Falkai, and Nikolaos Koutsouleris. 2018. Machine Learning Approaches for Clinical Psychology and Psychiatry. *Annual Review of Clinical Psychology* 14 (1): 91–118. <https://doi.org/10.1146/annurev-clinpsy-032816-045037>.
- Elder, Alexis. 2014. Excellent Online Friendships: An Aristotelian Defense of Social Media. *Ethics and Information Technology* 16 (4): 287–297.
- Ellis, David A., Brittany I. Davidson, Heather Shaw, and Kristoffer Geyer. 2019. Do Smartphone Usage Scales Predict Behavior? *International Journal of Human-Computer Studies* 130 (October): 86–92. <https://doi.org/10.1016/j.ijhcs.2019.05.004>.
- Evans, Katie. 2019. ‘Financial Transactions Data, AI and Mental Health: The Challenge’. Money and Mental Health Policy Institute. January 2019. <https://www.moneyandmentalhealth.org/wp-content/uploads/2019/01/FCA-financial-transactions-data-discussion-note-footnotes-at-end.pdf>.
- Faden, Ruth R., and Tom L. Beauchamp. 1986. *A History and Theory of Informed Consent*. New York: Oxford University Press.
- Ferguson, Christopher J. 2009. An Effect Size Primer: A Guide for Clinicians and Researchers. *Professional Psychology: Research and Practice* 40 (5): 532–538. <https://doi.org/10.1037/a0015808>.
- Ferretti, Agata, Elettra Ronchi, and Effy Vayena. 2019. From Principles to Practice: Benchmarking Government Guidance on Health Apps. *The Lancet Digital Health* 1 (2): e55–e57. [https://doi.org/10.1016/S2589-7500\(19\)30027-5](https://doi.org/10.1016/S2589-7500(19)30027-5).
- Fletcher, Guy. 2016. *The Philosophy of Well-Being: An Introduction*. London: Routledge.
- Floridi, Luciano, ed. 2010. *The Cambridge Handbook of Information and Computer Ethics*. Cambridge: Cambridge University Press.

- , ed. 2014. *The Fourth Revolution: How the Infosphere Is Reshaping Human Reality*. Oxford: Oxford University Press.
- . 2016a. Tolerant Paternalism: Pro-Ethical Design as a Resolution of the Dilemma of Toleration. *Science and Engineering Ethics* 22 (6): 1669–1688. <https://doi.org/10.1007/s11948-015-9733-2>.
- . 2016b. Faultless Responsibility: On the Nature and Allocation of Moral Responsibility for Distributed Moral actions. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374 (2083): 20160112. <https://doi.org/10.1098/rsta.2016.0112>.
- Floridi, Luciano, Josh Cows, Monica Beltrametti, Raja Chatila, Patrice Chazerand, Virginia Dignum, Christoph Luetge, et al. 2018. AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines* 28 (4): 689–707. <https://doi.org/10.1007/s11023-018-9482-5>.
- Frijda, Nico H., Peter Kuipers, and Elisabeth Ter Schure. 1989. Relations Among Emotion, Appraisal, and Emotional Action Readiness. *Journal of Personality and Social Psychology* 57 (2): 212.
- Fröding, Barbro, and Martin Peterson. 2012. Why Virtual Friendship Is No Genuine Friendship. *Ethics and Information Technology* 14 (3): 201–207.
- Galbraith, John W., and Greg Tkacz. 2018. Nowcasting with Payments System Data. *International Journal of Forecasting* 34 (2): 366–376.
- Garde-Hansen, Joanne. 2009. MyMemories?: Personal Digital Archive Fever and Facebook. In *Save As... Digital Memories*, 135–150. London: Springer.
- Ghafur, S., S. Kristensen, K. Honeyford, G. Martin, A. Darzi, and P. Aylin. 2019. A Retrospective Impact Analysis of the Wanna Cry Cyberattack on the NHS. *Npj Digital Medicine* 2 (1): 98. <https://doi.org/10.1038/s41746-019-0161-6>.
- Gluckman, Sir Peter, and Kristiann Allen. 2018. ‘Understanding Wellbeing in the Context of Rapid Digital and Associated Transformations’. The International Network for Government Science Advice. <https://www.ingsa.org/wp-content/uploads/2018/10/INGSAs-Digital-Wellbeing-Sept18.pdf>.
- Goldacre, Ben, Caroline E. Morton, and Nicholas J. DeVito. 2019. Why Researchers Should Share Their Analytic Code. *BMJ* 367 (November): 16365. <https://doi.org/10.1136/bmj.16365>.
- Greaves, Felix, Indra Joshi, Mark Campbell, Samantha Roberts, Neelam Patel, and John Powell. 2018. What Is an Appropriate Level of Evidence for a Digital Health Intervention? *The Lancet* 392 (10165): 2665–2667. [https://doi.org/10.1016/S0140-6736\(18\)33129-5](https://doi.org/10.1016/S0140-6736(18)33129-5).
- Hamelijnck, Oliver, Theodoros Damoulas, Kangrui Wang, and Mark Girolami. 2019. ‘Multi-Resolution Multi-Task Gaussian Processes’. In: Thirty-Third Conference on Neural Information Processing Systems, Canada, 8–14 Dec 2019
- Hausman, Daniel. 2015. *Valuing Health: Well-Being, Freedom, and Suffering*. New York: Oxford University Press.
- Haybron, Daniel M. 2008. *The Pursuit of Unhappiness: The Elusive Psychology of Well-Being*. Oxford: Oxford University Press.
- Haybron, Daniel M., and Valerie Tiberius. 2015. Well-Being Policy: What Standard of Well-Being? *Journal of the American Philosophical Association* 1 (4): 712–733. <https://doi.org/10.1017/apa.2015.23>.
- Henson, Philip, Gary David, Karen Albright, and John Torous. 2019. Deriving a Practical Framework for the Evaluation of Health Apps. *The Lancet Digital Health* 1 (2): e52–e54. [https://doi.org/10.1016/S2589-7500\(19\)30013-5](https://doi.org/10.1016/S2589-7500(19)30013-5).
- Hills, Thomas, Eugenio Proto, Daniel Sgroi, and Chanuki Illushka Seresinhe. 2019. Historical Analysis of National Subjective Wellbeing Using Millions of Digitized Books. *Nature Human Behaviour* 3: 1271–1275.
- Howard, Don. 2018. Technomoral Civic Virtues: A Critical Appreciation of Shannon Vallor’s Technology and the Virtues. *Philosophy & Technology* 31 (2): 293–304. <https://doi.org/10.1007/s13347-017-0283-1>.

- Huppert, Felicia A., and Timothy T.C. So. 2013. Flourishing Across Europe: Application of a New Conceptual Framework for Defining Well-Being. *Social Indicators Research* 110 (3): 837–861. <https://doi.org/10.1007/s11205-011-9966-7>.
- Jensen, Michael, Madeleine J. George, Michael R. Russell, and Candice L. Odgers. 2019. Young Adolescents' Digital Technology Use and Mental Health Symptoms: Little Evidence of Longitudinal or Daily Linkages. *Clinical Psychological Science*, August, 216770261985933. <https://doi.org/10.1177/2167702619859336>.
- Jeske, Diane. 2019. *Friendship and Social Media: A Philosophical Exploration*. Abingdon/Oxford/ New York: Routledge.
- Jobin, Anna, Marcello Ienca, and Effy Vayena. 2019. The Global Landscape of AI Ethics Guidelines. *Nature Machine Intelligence* 1 (9): 389–399. <https://doi.org/10.1038/s42256-019-0088-2>.
- Justinia, T. 2017. The UK's National Programme for IT: Why Was It Dismantled? *Health Services Management Research* 30 (1): 2–9. <https://doi.org/10.1177/0951484816662492>.
- Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. London: Penguin.
- Kim, Nancy S. 2019. *Consentability: Consent and Its Limits*. Cambridge: Cambridge University Press.
- Kosmyna, Nataliya, Caitlin Morris, Utkarsh Sarawgi, Thanh Nguyen, and Pattie Maes. 2019. AttentiVU: A Wearable Pair of EEG and EOG Glasses for Real-Time Physiological Processing. In *2019 IEEE 16th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, 1–4. Chicago: IEEE. <https://doi.org/10.1109/BSN.2019.8771080>.
- Layous, Kristin, S. Katherine Nelson, Jaime L. Kurtz, and Sonja Lyubomirsky. 2017. What Triggers Prosocial Effort? A Positive Feedback Loop between Positive Activities, Kindness, and Well-Being. *The Journal of Positive Psychology* 12 (4): 385–398. <https://doi.org/10.1080/017439760.2016.1198924>.
- Leigh, Simon, and Liz Ashall-Payne. 2019. The Role of Health-Care Providers in MHealth Adoption. *The Lancet Digital Health* 1 (2): e58–e59.
- Lin, Eden. 2017. Enumeration and Explanation in Theories of Welfare. *Analysis* 77 (1): 65–73. <https://doi.org/10.1093/analysis/anx035>.
- Lupton, Deborah. 2016. *The Quantified Self*. Hoboken: Wiley.
- Ma, Lawrence K., Richard J. Tunney, and Eamonn Ferguson. 2017. Does Gratitude Enhance Prosociality?: A Meta-Analytic Review. *Psychological Bulletin* 143 (6): 601–635. <https://doi.org/10.1037/bul0000103>.
- Mackey, Tim K., and Gaurvika Nayyar. 2016. Digital Danger: A Review of the Global Public Health, Patient Safety and Cybersecurity Threats Posed by Illicit Online Pharmacies. *British Medical Bulletin* 118 (1): 110–126. <https://doi.org/10.1093/bmb/ldw016>.
- Manson, Neil C., and Onora O'Neill. 2007. *Rethinking Informed Consent in Bioethics*. Cambridge/ New York: Cambridge University Press.
- Milano, Silvia, Mariarosaria Taddeo, and Luciano Floridi. 2020. Recommender Systems and Their Ethical Challenges. *AI and Society*, February. <https://doi.org/10.1007/s00146-020-00950-y>.
- Mittelstadt, Brent. 2019. Principles Alone Cannot Guarantee Ethical AI. *Nature Machine Intelligence* 1 (11): 501–507. <https://doi.org/10.1038/s42256-019-0114-4>.
- Morley, Jessica, and Luciano Floridi. 2019a. How to Design a Governable Digital Health Ecosystem. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3424376>.
- . 2019b. NHS AI Lab: Why We Need to Be Ethically Mindful About AI for Healthcare. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3445421>.
- . 2019c. 'The Limits of Empowerment: How to Reframe the Role of MHealth Tools in the Healthcare Ecosystem'. *Science and Engineering Ethics*, June. <https://doi.org/10.1007/s11948-019-00115-1>.
- Morley, J., and I. Joshi. 2019. Developing Effective Policy to Support Artificial Intelligence in Health and Care. *Eurohealth* 25 (2): 11–14. <https://apps.who.int/iris/bitstream/handle/10665/326127/Eurohealth-V25-N2-2019-eng.pdf?sequence=1&isAllowed=y>.
- Morley, Jessica, Caio Machado, Christopher Burr, Josh Cows, Mariarosaria Taddeo, and Luciano Floridi. 2019. 'The Debate on the Ethics of AI in Health Care: A Reconstruction and Critical

- Review'. SSRN Scholarly Paper ID 3486518. Rochester: Social Science Research Network. <https://papers.ssrn.com/abstract=3486518>.
- Munafò, Marcus R., and George Davey Smith. 2018. Robust Research Needs Many Lines of Evidence. *Nature* 553 (7689): 399–401. <https://doi.org/10.1038/d41586-018-01023-3>.
- Munafò, Marcus R., Brian A. Nosek, Dorothy V.M. Bishop, Katherine S. Button, Christopher D. Chambers, Nathalie Percie du Sert, Uri Simonsohn, Eric-Jan Wagenmakers, Jennifer J. Ware, and John P.A. Ioannidis. 2017. A Manifesto for Reproducible Science. *Nature Human Behaviour* 1 (1): 0021. <https://doi.org/10.1038/s41562-016-0021>.
- Nussbaum, Martha C. 2011. *Creating Capabilities*. Cambridge, MA/London: Harvard University Press.
- Nussbaum, Martha, and Amartya Sen. 1993. *The Quality of Life*. New York: Oxford University Press.
- OECD. 2019. How's Life in the Digital Age? Opportunities and Risks of the Digital Transformation for People's Well-Being. Organisation for Economic Co-operation and Development. https://www.oecd-ilibrary.org/science-and-technology/how-s-life-in-the-digital-age_9789264311800-en.
- Ofcom. 2018. *Adults' Media Use and Attitudes Report*. The Office of Communications. https://www.ofcom.org.uk/__data/assets/pdf_file/0011/113222/Adults-Media-Use-and-Attitudes-Report-2018.pdf.
- Orben, Amy. 2019. 'Teens, Screens and Well-Being: An Improved Approach'. PhD Thesis (University of Oxford). https://ora.ox.ac.uk/objects/uuid:198781ae-35b8-4898-b482-8df7201b59e1/download_file?file_format=pdf&safe_filename=_main.pdf&type_of_work=Thesis.
- Orben, Amy, and Andrew K. Przybylski. 2019. The Association between Adolescent Well-Being and Digital Technology Use. *Nature Human Behaviour* 3 (2): 173–182. <https://doi.org/10.1038/s41562-018-0506-1>.
- Orben, Amy, Tobias Dienlin, and Andrew K. Przybylski. 2019. Social Media's Enduring Effect on Adolescent Life Satisfaction. *Proceedings of the National Academy of Sciences of the United States of America* 116 (21): 10226–10228. <https://doi.org/10.1073/pnas.1902058116>.
- Palm, Elin. 2013. Who Cares? Moral Obligations in Formal and Informal Care Provision in the Light of ICT-Based Home Care. *Health Care Analysis* 21 (2): 171–188.
- Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Oxford University Press.
- Pugh, Jonathan, Hannah Maslen, and Julian Savulescu. 2017. Deep Brain Stimulation, Authenticity and Value. *Cambridge Quarterly of Healthcare Ethics* 26 (4): 640–657. <https://doi.org/10.1017/S0963180117000147>.
- Reeves, Byron, Nilam Ram, Thomas N. Robinson, James J. Cummings, C. Lee Giles, Jennifer Pan, Agnese Chiatti, et al. 2019. Screenomics: A Framework to Capture and Analyze Personal Life Experiences and the Ways That Technology Shapes Them. *Human-Computer Interaction*, March, 1–52. <https://doi.org/10.1080/07370024.2019.1578652>.
- Robeyns, Ingrid. 2005. The Capability Approach: A Theoretical Survey. *Journal of Human Development* 6 (1): 93–117.
- Ryan, R.M., and E.L. Deci. 2001. On Happiness and Human Potentials: A Review of Research on Hedonic and Eudaimonic Well-Being. *Annual Review of Psychology* 52: 141.
- Ryan, Richard M., and Edward L. Deci. 2017. *Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness*. New York: Guilford Press.
- Ryff, Carol D. 1989. Happiness Is Everything, or Is It? Explorations on the Meaning of Psychological Well-Being. *Journal of Personality and Social Psychology* 57 (6): 13.
- Seligman, Martin E.P., and Mihaly Csikszentmihalyi. 2000. Positive Psychology: An Introduction. *American Psychologist* 55 (1): 5–14.
- Sharp, Robert. 2012. The Obstacles Against Reaching the Highest Level of Aristotelian Friendship Online. *Ethics and Information Technology* 14 (3): 231–239.
- Sterckx, S., V. Rakic, J. Cockbain, and P. Borry. 2016. "You Hoped We Would Sleep Walk into Accepting the Collection of Our Data": Controversies Surrounding the UK Care. Data Scheme and Their Wider Relevance for Biomedical Research. *Medicine, Health Care and Philosophy* 19 (2): 177–190. <https://doi.org/10.1007/s11019-015-9661-6>.

- Sterelny, Kim. 2003. *Thought in a Hostile World: The Evolution of Human Cognition*. Malden: Blackwell.
- Stiglitz, Joseph, Amartya Sen, and Jean-Paul Fitoussi. 2008. *Report by the Commission on the Measurement of Economic Performance and Social Progress*. <https://ec.europa.eu/eurostat/documents/118025/118123/Fitoussi+Commission+report>.
- Stokes, Patrick. 2012. Ghosts in the Machine: Do the Dead Live on in Facebook? *Philosophy & Technology* 25 (3): 363–379.
- Sumner, Leonard W. 1996. *Welfare, Happiness, and Ethics*. Oxford: Clarendon Press.
- The Topol Review Board. 2019. *The Topol Review: Preparing the Healthcare Workforce to Deliver the Digital Future*. Health Education England. topol.hee.nhs.uk.
- Themistocleous, M., and V. Morabito. 2012. How Can User-Centred Design Affect the Acceptance and Adoption of Service Oriented Healthcare Information Systems? *International Journal of Healthcare Technology and Management* 13 (5–6): 321–344. <https://doi.org/10.1504/IJHTM.2012.052550>.
- Tiberius, Valerie. 2015. Prudential Value. In *The Oxford Handbook of Value Theory*, ed. Iwao Hirose and Jonas Olson, 158–174. New York: Oxford University Press.
- Torous, John, Gerhard Andersson, Andrew Bertagnoli, Helen Christensen, Pim Cuijpers, Joseph Firth, Adam Haim, et al. 2019. Towards a Consensus Around Standards for Smartphone Apps and Digital Mental Health: Towards a Consensus Around Standards for Smartphone Apps and Digital Mental Health. *World Psychiatry* 18 (1): 97–98. <https://doi.org/10.1002/wps.20592>.
- UK Government. 2019. *Online Harms White Paper*. UK Government. <https://www.gov.uk/government/consultations/online-harms-white-paper>.
- United Nations. 2019. The Sustainable Development Agenda. *United Nations Sustainable Development Goals* (blog). 2019. <https://www.un.org/sustainabledevelopment/development-agenda/>.
- Vallor, Shannon. 2012. Flourishing on Facebook: Virtue Friendship & New Social Media. *Ethics and Information Technology* 14 (3): 185–199. <https://doi.org/10.1007/s10676-010-9262-2>.
- . 2016. *Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting*. New York: Oxford University Press.
- Verduyn, Philippe, Oscar Ybarra, Maxime Résibois, John Jonides, and Ethan Kross. 2017. Do Social Network Sites Enhance or Undermine Subjective Well-Being? A Critical Review. *Social Issues and Policy Review* 11 (1): 274–302. <https://doi.org/10.1111/sipr.12033>.
- Vidgen, Bertie, Helen Margetts, and Alex Harris. 2019. *How Much Online Abuse Is There? A Systematic Review of Evidence for the UK*. The Alan Turing Institute. <https://www.turing.ac.uk/research/research-programmes/public-policy/programme-articles/how-much-online-abuse-there>.
- Warwick Machine Learning Group. 2019. London air quality. *The Alan Turing Institute*. 2019. <https://www.turing.ac.uk/research/research-projects/london-air-quality>.
- Watson, Ryan J., and John L. Christensen. 2017. Big Data and Student Engagement among Vulnerable Youth: A Review. *Current Opinion in Behavioral Sciences* 18 (December): 23–27. <https://doi.org/10.1016/j.cobeha.2017.07.004>.
- Watson, David S, Jenny Krutzinna, Ian N Bruce, Christopher EM Griffiths, Iain B McInnes, Michael R Barnes, and Luciano Floridi. 2019. Clinical Applications of Machine Learning Algorithms: Beyond the Black Box. *BMJ*, March, 1886. <https://doi.org/10.1136/bmj.1886>.
- Williams, James. 2018. *Stand Out of Our Light: Freedom and Resistance in the Attention Economy*. Oxford: Oxford University Press.
- Wood, Alex M., John Maltby, Gillett Raphael, P. Alex Linley, and Stephen Joseph. 2008. The Role of Gratitude in the Development of Social Support, Stress, and Depression: Two Longitudinal Studies. *Journal of Research in Personality* 42 (4): 854–871. <https://doi.org/10.1016/j.jrp.2007.11.003>.
- Woodward, Christopher. 2013. Classifying Theories of Welfare. *Philosophical Studies* 165 (3): 17.
- Woolley, Sandra, and Tim Collins. 2019. Some Heart-Rate Monitors Give Less Reliable Readings for People of Colour. *The Conversation*. 8 January 2019. <http://theconversation.com/some-heart-rate-monitors-give-less-reliable-readings-for-people-of-colour-121007>.

World Health Organisation. 2019. Constitution. 2019. <https://www.who.int/about/who-we-are/constitution>.

Yost-Dubrow, Rachel, and Yarrow Dunham. 2018. Evidence for a Relationship between Trait Gratitude and Prosocial Behaviour. *Cognition and Emotion* 32 (2): 397–403. <https://doi.org/10.1080/02699931.2017.1289153>.

Christopher Burr is a Philosopher of Cognitive Science and Artificial Intelligence. He is a Senior Research Associate at the Alan Turing Institute and a Research Associate at the Digital Ethics Lab/Oxford Internet Institute, University of Oxford.

His current research explores philosophical and ethical issues related to data-driven technologies, including the opportunities and risks that such technologies have for mental health and well-being. A primary goal of this research is to develop frameworks and guidance to support the governance, responsible innovation and sustainable use of data-driven technology within a digital society.

To support this goal, he has worked with a number of public sector bodies and organisations, including NHSx; the UK Government’s Department for Health and Social Care; the Department for Digital, Culture, Media and Sport; the Centre for Data Ethics and Innovation; and the Ministry of Justice.

He has held previous posts at the University of Bristol, where he explored the ethical and epistemological impact of big data and artificial intelligence as a postdoctoral research and also completed his PhD in 2017.

Research Interests: Philosophy of Cognitive Science and Artificial Intelligence, Digital Ethics, Decision Theory, Public Policy and Human-Computer Interaction.

cburr@turing.ac.uk

Luciano Floridi is the OII’s Professor of Philosophy and Ethics of Information at the University of Oxford, where he is also the Director of the Digital Ethics Lab of the Oxford Internet Institute. Still in Oxford, he is Distinguished Research Fellow of the Uehiro Centre for Practical Ethics of the Faculty of Philosophy and Research Associate and Fellow in Information Policy of the Department of Computer Science. Outside Oxford, he is Turing Fellow of the Alan Turing Institute (the national institute for data science and AI) and Chair of its Data Ethics Group and Adjunct Professor (‘Distinguished Scholar in Residence’) of the Department of Economics, American University, Washington D.C.

He is deeply engaged with emerging policy initiatives on the socio-ethical value and implications of digital technologies and their applications. He has worked closely on digital ethics (including the ethics of algorithms and AI) with the European Commission, the German Ethics Council and, in the UK, with the House of Lords, the House of Commons, the Cabinet Office and the Information Commissioner’s Office, as well as with multinational corporations (e.g. Cisco, Google, IBM, Microsoft and Tencent). Among his current commitments, he is Chair of the Ethics Committee of the Machine Intelligence Garage project, Digital Catapult, UK innovation programme; Member of the Board of the UK’s Centre for Data Ethics and Innovation (CDEI); the Advisory Board of The Institute for Ethical AI in Education; the EU Commission’s High-Level Group on Artificial Intelligence; EY AI Advisory Board; and the Advisory Board of the Vodafone Institute for Society and Communications.

Research Interests: Digital Ethics (including the ethics of AI, and Information and Computer Ethics), Philosophy of Information and Philosophy of Technology.

Among his recent books, all published by Oxford University Press (OUP): *The Logic of Information* (2019); *The Fourth Revolution - How the infosphere is reshaping human reality* (2014), winner of the J. Ong Award; *The Ethics of Information* (2013); *The Philosophy of Information* (2011).

luciano.floridi@oii.ox.ac.uk