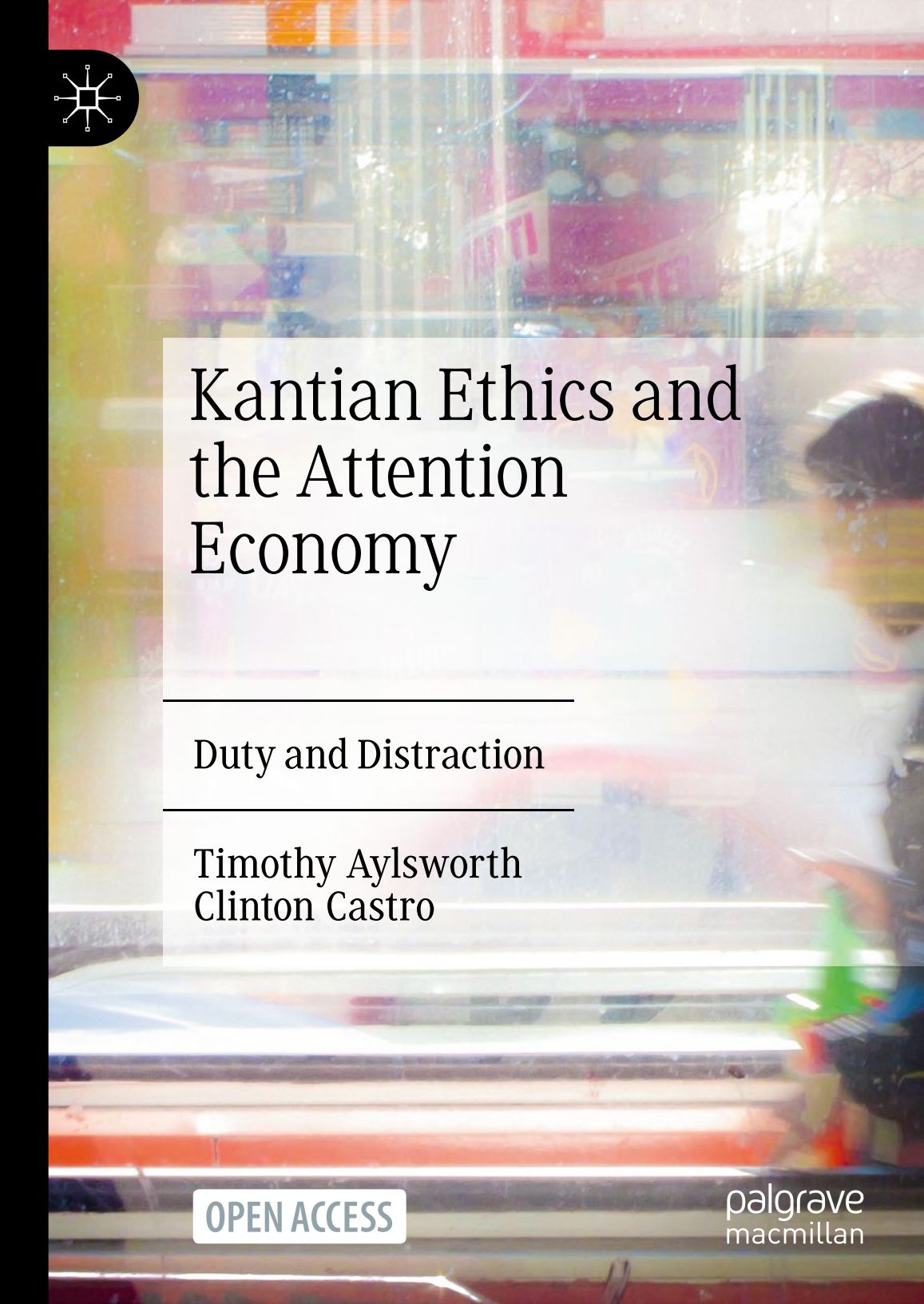

Kantian Ethics and the Attention Economy

Duty and Distraction

Timothy Aylsworth
Clinton Castro

OPEN ACCESS

palgrave
macmillan

Kantian Ethics and the Attention Economy

Timothy Aylsworth • Clinton Castro

Kantian Ethics and the Attention Economy

Duty and Distraction

palgrave
macmillan

Timothy Aylsworth
Florida International University
Miami, FL, USA

Clinton Castro
The Information School
University of Wisconsin-Madison
Madison, WI, USA

ISBN 978-3-031-45637-4 ISBN 978-3-031-45638-1 (eBook)
https://doi.org/10.1007/978-3-031-45638-1

Cover illustration: © Alex Linch shutterstock.com

This Palgrave Macmillan imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Paper in this product is recyclable.

To Iris, Ray, Rosa, and Hanlon, whose autonomy we are dedicated to fostering.

Preface

In the early days of the COVID-19 pandemic, the two of us began to complain to each other about how we were spending too much time on our phones. Like many people during that difficult time, we found it hard to tear ourselves away from the constant stream of bad news, alarmist videos, and anxiety-inducing social media posts. That kind of behavior became so widespread that it was eventually given a catchy name: "doomscrolling." Long before these conversations, however, we had both developed an appreciation for the work of Cal Newport, a professor of computer science who has written extensively about attention, productivity, and the threat posed by our unhealthy relationship with technology.

We were sympathetic to Newport's arguments, but neither of us believed that there was much of a philosophical dimension to these concerns. The majority of books written on this topic fall under the umbrella of self-help. They typically argue that by spending less time on our phones, we could be happier, less anxious, and more productive. All of those arguments seemed plausible to us. The empirical literature on smartphone use has become vast, and the evidence is damning.

But at some point, we started to wonder if there was an underappreciated moral dimension to these concerns. Both of us are committed to Kantian ethical principles about the importance of autonomy, so the initial thought was quite natural. If autonomy matters morally and smartphones are bad for autonomy, then we have a moral reason to be concerned about our phones.

At first glance, this seemed like a simple story to tell. All we had to do was explain how the value of autonomy justifies the existence of duties to

oneself (a project that has already been undertaken by both Kant scholars and contemporary ethicists) and then present enough empirical evidence to justify the concern about smartphones. We put that argument together fairly quickly and we published our first paper on this topic in the *Journal of Applied Philosophy*.

Before long, we realized that there was more to be said. It is not enough to think of this issue in terms of duties we owe to ourselves. No one builds their relationship with their phone in a vacuum. We began to consider the roles that tech companies, developers, employers, teachers, and parents play. In many cases, we have a profound effect on how people around us use technology, so we extended our initial argument by considering duties that we owe to each other. Finally, we expanded our focus to societal concerns about polarization, democracy, and the many ways that our relationship with social media has an impact on our ability to govern ourselves collectively.

After publishing our second paper on these issues in *Philosophy & Technology*, we realized that we had more than enough material to write a book on this topic. Our two papers had only scratched the surface. We wanted to develop a more detailed account of autonomy; we were hoping to take a deeper dive into the empirical issues; and we wanted to spend more time justifying the Kantian credentials of the duty. We were delighted that Palgrave Macmillan was interested in the project, and we are especially grateful to our editor, Philip Getz.

This project has benefited immensely from thoughtful feedback that we received from colleagues, friends, students, and audiences at many conferences. We are greatly indebted to Daniel Muñoz for giving us many rounds of insightful comments and for encouraging us to pursue this project in the first place. We would also like to thank audiences at Denison University, Texas A&M, Vanderbilt, UW-Madison, the Pacific APA (2023, 2021), the Central APA (2020), the Florida Philosophical Association (2021, 2022), the Southern Meeting of the North American Kant Society, and the MANCEPT Workshop at the University of Manchester.

There are many individuals we would like to thank in particular: Gus Skorburg, Dylan White, Perica Jovchevski, Bartek Chomanski, Giles Howdle, Lindsey Schwartz, William Chan, Laurie Shrage, Paul Warren, Céline Leboeuf, Diana Acosta-Navas, Hanna Gunn, Olga Lenczewska, Owen Ware, Patrick Frierson, Will Fleisher, Elizabeth Stewart, Stefano Lo Re, Stephen Daniel, Kristi Sweet, Ted George, Chris Menzel, Kenny Easwaran, and Peter Zuk. We are especially grateful to Elizabeth

Scarbrough and Kamila Forson for spending a supererogatory amount of time reading the manuscript and giving us helpful feedback. We would also like to thank some of our students at Florida International University for sharing their thoughts with us: Judith Chavarria, Archer Amon, Ilana Smith, George Lopez, Jorge Valdes, Carolina Mesa, Gabriel Parra Nogueiras, Omar Urbina, Anne Gazzera, Valeria Gonzales, and Milo Dupuis. Last (but certainly not least) we would like to thank Molly Castro and Katie Rainwater for their support and encouragement.

Although we have published two papers on this topic, almost all of the writing in this book is new. We wanted to explore the issues in much greater depth, and this required us to produce entirely new content. But we did borrow a handful of sentences from our other work, so we would like to gratefully acknowledge those journals for giving us permission to reprint parts of those papers. We would like to thank the *Journal of Applied Philosophy*, *Philosophy & Technology*; and the *Journal of Business Ethics* and we would like to note that in this work, as in all of our previous work, we have contributed equally as authors.

Finally, we are very grateful to the University of Wisconsin-Madison and Florida International University for providing funding to make this book available through Open Access. In particular we would like to thank Walter Van Hamme and FIU's School of Integrated Science and Humanity. This book is available through Open Access thanks to the University of Wisconsin Information School's Sarah M. Pritchard Faculty Support Fund and Mary Elizabeth Koch Fund, and the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin–Madison with funding from the Wisconsin Alumni Research Foundation.

Miami, FL, USA — Timothy Aylsworth
Madison, WI, USA — Clinton Castro

Contents

About the Authors

Timothy Aylsworth is an assistant professor of philosophy at Florida International University. He wrote his dissertation on Kant's concept of freedom, and he works on a variety of issues related to Kant's ethics. He is especially interested in topics related to autonomy, manipulation, and the problem of collective harm.

Clinton Castro is an assistant professor in The Information School at University of Wisconsin-Madison. Before that, he was an assistant professor in the Department of Philosophy at Florida International University; much of this book was conceived while he, Tim, and their children were running and biking around the parks and bike paths of South Miami. He specializes in information ethics, epistemology, and fair machine learning.

Abbreviations[1]

AF	Anthropology Friedländer in Kant (2013)
Anth.	*Anthropology from a Pragmatic Point of View* in Kant (2007a)
Collins	Collins Lectures on Ethics in Kant (1997)
G	*Groundwork of the Metaphysics of Morals* in Kant (2008)
Idea	Idea for a universal history with a cosmopolitan aim in Kant (2007a)
KU	*Critique of the Power of Judgment* in Kant (2007b)
KpV	*Critique of Practical Reason* in Kant (2008)
MS	*Metaphysics of Morals* in Kant (2008)
TP	"On the common saying: That may be correct in theory, but it is of no use in practice" in Kant (2008)
UP	Lectures on Pedagogy in Kant (2007a)
VE	Lectures on Ethics in Kant (1997)

[1] All translations of Kant's writings are taken from The Cambridge Edition of the Works of Immanuel Kant. We adopt the standard convention of citing the *Critique of Pure Reason* by referring to the 1781 (A) and 1787 (B) editions. For unpublished notes, we give the *Reflexionen* number. For all other works, we abbreviate the title and give the volume and page number from the *Akademie Ausgabe* edition of *Kants Gesammelte Schriften*, edited by Königlich Preussische Akademie der Wissenschaften (29 vols. Berlin: de Gruyter, 1900–)

List of Figures

CHAPTER 1

Introduction

Distractions must never be tolerated, least of all at school, for they eventually produce a certain tendency in that direction, a certain habit. Even the most beautiful talents perish in one who is subject to distractions.
—Kant, Lectures on Pedagogy (UP 4:473–74)

We are moving from a world where computing power was scarce to a place where it now is almost limitless, and where the true scarce commodity is increasingly human attention.
—Microsoft CEO Satya Nadella, memo to employees (See *Mosendz 2014*)

Ethical concerns about technology are nothing new. They go back at least as far as Plato, who worried that the invention of writing would weaken our capacity to remember things on our own. Some of these concerns may have been overblown, while others have proven to be real problems.[1] The rapid deployment of new technologies in the last few decades has led to a veritable explosion of such worries. Discussion of these issues now involves a dizzying array of topics—everything from the use of biased algorithms in

[1] For concerns about technology and literacy, see Carr (2010), Wolf (2018), and Loh and Kanai (2014). Plato's worry can be found at *Phaedrus*, 274b–277a (1997). We should add here that Plato—in a classic display of Platonic irony—communicated this thought in writing.

T. Aylsworth, C. Castro, *Kantian Ethics and the Attention Economy*,
https://doi.org/10.1007/978-3-031-45638-1_1

the criminal justice system to the rise of superintelligent AI and killer robots.[2]

Our focus in this book—the negative effects that smartphones have on our autonomy—may seem quite modest by comparison. The issues that we will address are not ones that are depicted in the media as existential threats to humanity, nor are they frequently discussed by policymakers. We are worried about something that most people would never dream of subjecting to moral scrutiny. But, in many cases, the dangers we fail to appreciate are more pernicious than those that make headlines. By the time the water is boiling, it is too late to jump out of the pot.

The technologies that concern us in this book are ones that we have already invited into our lives. They have become such an ingrained part of our daily activities that it is hard to imagine leaving the house without them. Indeed, nomophobia (no mobile phone phobia, i.e., fear of being without a mobile phone) is an increasingly common phenomenon, with researchers estimating that "approximately 100%" of university students have it (Tuco et al. 2023). Many people have a smartphone on their person *every minute of the day*.[3] We routinely gaze into our screens and find ourselves compulsively tethered to the endless drip of vibrations, dings, and notifications. According to one estimate, Americans collectively check their phones eight *billion* times *per day*.[4]

Yet many of us do not, in fact, *want* to look at screens as much as we do. Nearly 60% of adults say that they use their phones "too much," and that sentiment is even more prevalent among Millennials and Zoomers.[5] Mobile phone use has also been shown to get in the way of our closest relationships.[6] Over one third of parents find that they use their own

[2] For a helpful overview of the topic of algorithmic bias, see Fazelpour and Danks (2021); for killer robots, super intelligent AI (and more) see Müller (2020).

[3] A 2015 Pew poll found that 90% of cell phone owners say they "frequently" carry their phone, and a majority of them "rarely" (47%) or "never" (36%) turn them off. See Rainie and Zickuhr (2015). More recent poll data have shown that these numbers have gone up considerably. A 2022 Gallup poll found that over 90% of Americans keep their phones near them "almost all the time during waking hours" and well over 80% "Keep it near at night when sleeping" (Saad 2022).

[4] See Eadicicco (2015).

[5] These figures come from the 2022 Gallup poll. For smartphone users between the ages of 18–29, 81% of them reported using their phone too much, and for those between 30 and 49, it was 74%. See Saad (2022).

[6] See Dwyer et al. (2018), Kushlev and Dunn (2019), Misra et al. (2016), Chotpitayasunondh and Douglas (2018), Pancani et al. (2021), and Wang et al. (2020).

phones too much, and over 50% of teens report that their parents or caregivers are often or sometimes distracted by their phones when they are trying to have a conversation with them.[7]

In 2018, a group of economists offered social media users various amounts of money in order to determine how much it would take for them to quit Facebook for just four weeks. The average number ended up around $180, which was higher than most estimates.[8] The great irony of the study is that the majority of users reported feeling *happier* without it. They spent more time with family and friends, and they used Facebook much less after the experiment was over. Perhaps most shockingly, they found that those who quit for a month reported increased well-being equal to the jump one would get *from earning about $30,000 more in annual income* (Allcott et al. 2020, 654). Faced with data of this kind, it starts to look like people are no longer in charge of their own lives when it comes to their phones and their use of social media. People have, as Thoreau once said, "become the tools of their tools" (1991, 32).

It is this relationship with technology that worries us.

More specifically, we will argue that a variety of "smart" technologies have captured our attention in such a way that we have forfeited some of our autonomy to our devices. This leads us to believe that there are compelling moral reasons to restructure our relationship with technology. To better understand the nature of this concern, consider the following cases:

Esther. "I wish I could read. I really do. I try to read. I buy books. I open books. And then I black out and I'm on Instagram, and I don't know what happened" (Povitsky 2020).

Monica. "When I hear the Slack ping that announces a new message, I feel a Pavlovian pull to read it, right then, right away. There's a red circle noting the number of new messages that nudges me to drop whatever I'm doing and click. That's surely by design. Tristan Harris, a former Google employee turned industry critic, notes that red is a known trigger color. These psychological pulls are not great for my productivity and peace of mind" (Torres 2019).

And finally,

[7] See Rainie and Zickuhr (2015).
[8] See Allcott et al. (2020).

Damon. "Every time I log onto Facebook, I brace myself. My newsfeed—like everyone else's I know—is filled with friends, relatives and acquaintances arguing about COVID-19, masks and Trump. Facebook has become a battleground among partisan 'echo chambers'" (Centola 2020).

To many of us, these experiences are familiar. All too often we set out to complete a task but are interrupted and subsequently derailed by some social media service we have subscribed to for personal enjoyment or by some communication platform that our employer requires us to monitor. And few, if any of us, are lucky enough to be innocent of the horror of the contemporary newsfeed.

At a glance, it might appear that these cases lack a serious moral dimension or any connection to autonomy. Perhaps distraction is the price one pays to keep up with friends and family. Being interrupted by co-workers is part of having a job. Off-putting comments simply come with the territory of interacting with a large, diverse group of people. Some might say that these situations are little more than old wine in new bottles.

But there is more to these cases. Many of the most popular apps and platforms are meticulously designed behavior modification machines, leveraging behavioral psychology and well-known human vulnerabilities to attract us to them and habituate their use (see, e.g., Eyal 2014). Monica makes a sharp note of this, referencing both Pavlov's dog and trigger colors.

Thus, when we succumb to certain technological distractions, we are not only failing to do what is in our own self-interest (e.g., read more) but our behavior exemplifies what Kant calls "heteronomy"—we are allowing ourselves to be driven by external forces. Put another way, we are failing to self-govern; we are failing to act *autonomously*. In the course of this book, we will tap into the rich Kantian tradition of thinking about autonomy and its moral importance to argue that we have moral duties—individually and collectively—to shun technological heteronomy and protect autonomy and its attendant capacities, in ourselves and in each other.

If our arguments succeed, then these are not merely cases of weakness of will, annoying work conditions, or grating media content; rather, they involve genuine moral problems. What is more, we hope to show that they are instances of different, but related, moral failures.

Now, our goal here is not to pick on Esther, Monica's employer, or Damon's friends. Neither of us have transcended technological distraction

ourselves. Instead, our aim is to make salient the moral obligations that we need to be increasingly mindful of as mobile devices pervade our private lives, our workplaces, and the public sphere. We will make clear in these contexts—and others, such as parenting and education—that we as individuals have obligations to ourselves and to others and we will show how these obligations stem from the value of autonomy. In addition to these individual obligations, we hope to show that various collectives we belong to—such as the body politic—have similar obligations, grounded in our need to act collectively in the face of problems such as the COVID-19 pandemic or climate change.

Our book is organized in such a way that it allows us to track these differences. We develop a set of concepts and moral principles that will help us analyze the key feature of Esther's case. This concerns the duty we have to ourselves to be "digital minimalists." We define a digital minimalist as one whose interactions with digital technology are intentional, such that they do not conflict with the agent's ability to set and pursue her own ends. We ground this obligation in terms of a Kantian duty to oneself. Kant famously argues that we are required to respect rational agency even in our own person. If it is true that our relationship with technology threatens to undermine our capacities as rational agents, then this would mean that we have a moral duty to protect ourselves from this threat.

This helps to explain why our analysis of Monica's case will be very different, since it involves duties to others. While Monica's employer has a duty to respect her autonomy—an instance of what we call the duty to be an "attention ecologist" (i.e., one who promotes digital minimalism in others)—that duty is conditioned by Monica's sovereignty over herself. Finally, Damon's case will receive separate treatment, since it involves a different sort of autonomy altogether. In his case, he witnesses a breakdown of our collective ability to solve problems that require action at the group level.

We begin in Chap. 2 by developing an account of personal autonomy. The main argument of the book is that we have moral reasons to cultivate our autonomy and to protect it from the threats posed by our unhealthy relationship with mobile devices and the attention economy. So our first task is to explain what autonomy is and why it matters morally. Our view is generally Kantian. But when we use the word "autonomy," we are referring to what Kant called "humanity" (the rational capacity to set and pursue your own ends). As we will explain later, this means that we are dealing with personal autonomy rather than moral autonomy. We then break

autonomy down to two separate components: capacity and authenticity. To do this, we draw on several contemporary accounts of autonomy. We then present Kantian arguments about the moral weight of autonomy, and we explain why other ethical theories are committed to similar claims.

In Chap. 3, we turn to the empirical literature on mobile devices and their deleterious effects. Though the technology is fairly new, psychologists, neuroscientists, and social scientists have already written a great deal on this topic.[9] Mobile devices and the attention economy have been linked to negative effects on attention, working memory, executive function, sleep, depression, anxiety, and more.[10] We take the next step of our argument by connecting this empirical research to our discussion of autonomy in Chap. 2. These are the longest chapters of the book, as they provide the foundations for the moral arguments that follow.

Once the groundwork has been laid, we begin defending the existence of various moral duties. Each chapter deals with a different obligation, and the topics are broken down in ways that reflect Kant's taxonomy of duties in the *Metaphysics of Morals*. In Chap. 4, we begin, as Kant does, with duties to oneself. We argue that you have a moral duty to be a digital minimalist. We explain what we mean by "digital minimalism," and we show how this duty fits in within the broader framework of Kantian ethics.

In Chap. 5, we turn to duties to others, and we discuss various instances of these obligations. We argue that the duty to promote the autonomy of others is especially demanding for parents and teachers, who have special obligations to cultivate the autonomy of their children and students. We derive such imperfect duties (duties of love) from the Kantian requirement to respect rational agents as ends in themselves. We also discuss perfect duties (duties of respect) to refrain from using others as a mere means. We explore applications of this duty for employers and software developers in particular.

One thing that unifies all of the above obligations is that they are instances of what Kant calls "duties of virtue." They concern the moral duties of individuals. We shift this focus in Chap. 6 as we discuss the implications for policymakers, legally enforceable obligations, which Kant refers to as "duties of right." Although we refrain from making many specific policy recommendations, we outline the kind of Kantian reasoning that

[9] Many consider the release of the iPhone in 2007 to mark the advent of the smartphone era, even though devices like the Blackberry came out much earlier.

[10] See, e.g., Wilmer et al. (2017).

could be used to justify such regulations. Finally, we extend duties of digital minimalism to group agents in Chap. 7 as we defend the existence of collective moral obligations. If we were to restrict our focus to the ways that technology can harm us as individuals, we would overlook some morally significant harms to groups *qua* groups. Thus, in Chap. 7, we argue that addictive technology weakens our capacity to act autonomously *as a group*. We defend this claim by arguing that certain features of the attention economy (e.g., that it contributes to polarization) threaten to erode the legitimacy of political institutions.

We conclude in Chap. 8 by revisiting the three vignettes from this introduction and showing how the concepts and principles of the book make it possible for us to understand exactly what is going wrong in those cases.

References

Allcott, Hunt, Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow. 2020. The welfare effects of social media. *American Economic Review* 110: 629–676.

Carr, Nicholas. 2010. *The shallows: What the internet is doing to our brains*. New York: W. W. Norton & Company.

Centola, Damon. 2020. Why social media makes us more polarized and how to fix it. *Scientific American*. https://www.scientificamerican.com/article/why-social-media-makes-us-more-polarized-and-how-to-fix-it/. Accessed 1 Feb 2023.

Chotpitayasunondh, Varoth, and Karen M. Douglas. 2018. The effects of 'phubbing' on social interaction. *Journal of Applied Social Psychology* 48: 304–316.

Dwyer, Ryan J., Kostadin Kushlev, and Elizabeth W. Dunn. 2018. Smartphone use undermines enjoyment of face-to-face social interactions. *Journal of Experimental Social Psychology* 78: 233–239.

Eadicicco, Lee. 2015. Americans check their phones 8 billion times per day. *Time*. https://time.com/4147614/smartphone-usage-us-2015/. Accessed 2 May 2023.

Eyal, Nir. 2014. *Hooked: How to build habit-forming products*. New York: Penguin.

Fazelpour, S., & Danks, D. 2021. Algorithmic bias: Senses, sources, solutions. Philosophy Compass, 16(8):e12760.

Kushlev, Kostadin, and Elizabeth W. Dunn. 2019. Smartphones distract parents from cultivating feelings of connection when spending time with their children. *Journal of Social and Personal Relationships* 36 (6): 1619–1639.

Loh, Kep Kee, and Ryota Kanai. 2014. Higher media multi-tasking activity is associated with smaller gray-matter density in the anterior cingulate cortex. *PLoS One* 9: e106698.

Misra, Shalini, Lulu Cheng, Jamie Genevie, and Miao Yuan. 2016. The iPhone effect: The quality of in-person social interactions in the presence of mobile devices. *Environment and Behavior* 48: 275–298.

Mosendz, Polly. 2014. Microsoft's CEO sent a 3,187-word memo and we read it so you don't have to. *The Atlantic*. https://www.theatlantic.com/technology/archive/2014/07/microsofts-ceo-sent-a-3187-word-memo-and-we-read-it-so-you-dont-have-to/374230/. Accessed 17 July 2023.

Müller, Vincent C. 2020. Ethics of artificial intelligence and robotics. In *The Stanford encyclopedia of philosophy*, ed. Edward N. Zalta https://plato.stanford.edu/archives/sum2021/entries/ethics-ai/

Pancani, Luca, Tiziano Gerosa, Marco Gui, and Paolo Riva. 2021. 'Mom, dad, look at me': The development of the parental phubbing scale. *Journal of Social and Personal Relationships* 38: 435–458.

Plato. 1997. *Plato: Complete works*, ed. John M. Cooper and D. S. Hutchinson. Indianapolis: Hackett Publishing Co.

Povitsky, Esther. 2020. *Esther Povitsky: Hot for my name*. Aired July 17, 2020. Comedy Central.

Rainie, Lee, and Kathryn Zickuhr. 2015. How teens and parents navigate screen time and device distractions. http://www.pewinternet.org/2015/08/26/americans-views-on-mobile-etiquette/. Accessed 17 July 2023.

Saad, Lydia. 2022. Americans have close but wary bond with their smartphone. *Gallup*. https://news.gallup.com/poll/393785/americans-close-wary-bond-smartphone.aspx. Accessed 23 May 2023.

Thoreau, Henry David. 1991. *Walden*. New York: Vintage Books.

Torres, Monica. 2019. Slack is bad, actually. *HuffPost*. https://www.huffpost.com/entry/slack-is-bad-for-privacy_l_5d0bdc05e4b0aa375f49aa23. Accessed 1 Feb 2023.

Tuco, Kimberly G., Sharong D. Castro-Diaz, David R. Soriano-Moreno, and Vicente A. Benites-Zapata. 2023. Prevalence of nomophobia in university students: A systematic review and meta-analysis. *Healthcare Informatics Research* 29: 40–53.

Wang, Xingchao, Ling Gao, Jiping Yang, Fengqing Zhao, and Pengcheng Wang. 2020. Parental phubbing and adolescents' depressive symptoms: Self-esteem and perceived social support as moderators. *Journal of Youth and Adolescence* 49: 427–437.

Wilmer, Henry H., Lauren E. Sherman, and Jason M. Chein. 2017. Smartphones and cognition: A review of research exploring the links between mobile technology habits and cognitive functioning. *Frontiers in Psychology* 8: 1–16.

Wolf, Maryanne. 2018. *Reader, come home: The reading brain in a digital world*. New York: Harper.

CHAPTER 2

Respect for Humanity

Every human being has a legitimate claim to respect from his fellow human beings and is in turn *bound to respect every other. Humanity itself is a dignity; for a human being cannot be used merely as a means by any human being (either by others or even by himself) but must always be used at the same time as an end.*
—Kant, *The Metaphysics of Morals*, 6:462

Let's admit it, we in the consumer web industry are in the manipulation business. We build products meant to persuade people to do what we want them to do. We call these people "users" and even if we don't say it aloud, we secretly wish every one of them would become fiendishly addicted.
—*Nir Eyal* (See Eyal (*2012*). This quote comes from an article Eyal wrote for *Forbes*. He uses nearly identical language in chapter 6 of *Hooked*. *Cf.* Eyal (*2014*, 164–65).)

2.1 Introduction

Many of us share the intuition that autonomy matters. You ought to be the author of your own life story. And if you were to discover that your actions and desires are being manipulated by someone else, you would probably feel uncomfortable about that fact. In what follows, we will argue that this is precisely what is happening with our mobile devices.

But before moving on to our moral arguments about technology and digital minimalism, we must begin by establishing the ethical framework

T. Aylsworth, C. Castro, *Kantian Ethics and the Attention Economy*,
https://doi.org/10.1007/978-3-031-45638-1_2

that we will use throughout the book. We do not want to assume that everyone shares our Kantian commitments. Instead, we would like to defend them as independently plausible moral views. We begin with autonomy. Regardless of the moral theory that you happen to hold, you probably believe that autonomy is morally significant in some way. But how should we define autonomy? And why exactly does it matter morally? Our aim of this chapter is to answer those questions.

We understand autonomy as self-government: it is a rational agent's capacity to set and pursue her own ends. This may sound uninformative at first, but it will be fleshed out considerably in the next few sections. Once we have established a definition, we argue for a moral obligation to respect autonomy. This conclusion should start off with some intuitive appeal. It is fairly commonplace to acknowledge violations of autonomy as moral wrongs. That's why people often think that there is something morally objectionable about manipulation and coercion. Such acts undercut an agent's ability to set and pursue her own ends. Finally, we will show why certain technologies, like smartphones and social media, threaten to undermine our autonomy in Chap. 3.

2.2 Humanity: Autonomy as Rational Agency

When Kant talks about "autonomy," he is referring to *moral* autonomy, which he understands as an agent's capacity to be governed solely by a moral law that proceeds from her own rational will. He defines (moral) autonomy as "the property of the will by which it is a law to itself" (*G* 4:440). According to Kant, the moral law springs from one's own reason, so when we obey the moral law, we are submitting to a principle that we have given ourselves. For Kant, morality and autonomy are two sides of the same coin: morality consists in self-legislation through reason.

This idea of moral self-legislation harmonizes nicely with the Greek origin of the word: "autos" meaning "self" and "nomos" meaning "law." But Kant's lofty ideal sets the bar too high to match the usage in contemporary moral debates. Nowadays, when a moral philosopher refers to "autonomy," it typically means something quite different and much less demanding. More often than not, ethicists are talking about *personal* autonomy rather than *moral* autonomy. Personal autonomy is the ability to be in charge of our own lives, to act and reflect on the basis of our own beliefs and desires, and to pursue things in accordance with our own conception of what is good for us. This is radically different from Kant's

concept of autonomy according to which acting autonomously is synonymous with acting morally.[1] Throughout the book, whenever we talk about "autonomy," we are almost always talking about personal autonomy. We will say very little about Kant's idea of moral autonomy.

But it is not as if Kant makes no mention of personal autonomy.[2] On the contrary, he talks a great deal about the importance of rational agency. But given his special emphasis on moral autonomy, Kant uses a different word to refer to this capacity. The word he uses is "humanity."[3] Kant refers to "humanity" throughout his moral philosophy, and he tends to gloss it as "the capacity to set oneself an end" (*MS* 6:392).[4] Kant thinks that this capacity is what sets rational beings like us apart from the rest of the animal kingdom. Our inclinations do not necessitate our actions. Whatever desire we may have, we are capable of rejecting it for the sake of another end. What is more, we have the ability to reflect on the value of our chosen ends and revise those choices; we are free to pursue "any end whatsoever"

[1] As Joseph Raz puts it: "Personal autonomy, which is a particular ideal of individual well-being, should not be confused with the only very indirectly related notion of moral autonomy. The latter originates with the Kantian idea that morality consists of self-enacted principles" (1986, 370).

[2] Paul Formosa argues against those who think that Kant lacks a morally significant conception of personal autonomy. He concludes that personal autonomy is the ability to "adopt merely permissible ends in accordance with your own personal conception of happiness" and he adds the "limiting condition that your will has normative authority for all rational agents" (2013, 209). This definition is fairly similar to our understanding of what Kant means by "humanity." As Allen Wood points out, Kant thinks that humanity "enables us not only to set ends but to compare the ends we set and organize them into a system (*KU* 5:426–427). Hence humanity also involves the capacity to form the idea of our happiness or well-being as a whole" (1999, 119). And Formosa's limiting condition would follow from Kant's formula of humanity, which we discuss in the next section.

[3] Careful readers of Kant are likely to notice that he attaches dignity (*Würde*) to moral autonomy rather than personal autonomy. To use Kant's jargon, dignity is grounded in "personality" rather than "humanity." This means that fellow Kant scholars might have reservations about our decision to focus on humanity (personal autonomy) rather than personality (moral autonomy). But there is no reason to get hung up on this point. After all, it is only because we are rational agents that we are capable of morality in the first place. As Kant puts it in the *Groundwork*: "Thus morality and **humanity, insofar as it is capable of morality**, is that which alone has dignity" (*G* 4:434–35; emphasis added). Furthermore, Kant identifies humanity as the categorical imperative's "end in itself." See Wood (1999) for a helpful discussion of why Kant chooses humanity rather than personality as the "end in itself" (118–21).

[4] This definition from the *Metaphysics of Morals* is very similar to the one given in the *Groundwork*: "Rational nature is distinguished from the rest of nature by this, that it sets itself an end" (*G* 4:437).

(*MS* 6:392). You might set out to become an investment banker, because you wanted to make a lot of money. But then you get to college and decide that you do not, upon reflection, care about getting rich, so you decide to pursue a career in academic philosophy. As a rational being, you get to set your own ends, whatever those might be.

That is the idea of autonomy that we will use throughout the book. Simply put, autonomy is the ability to set and pursue your own ends. Of course, we will need to say more about what it means for ends to be considered your "own," but this is a good starting place for the definition. Philosophers have employed a wide variety of metaphors to express the core idea of autonomy. To be autonomous is to be the author of your own life story (Raz 1986, 369). Autonomy requires sovereignty, the capacity to govern yourself. Autonomy means putting yourself in the driver's seat rather than being a mere passenger.

But metaphors can take us only so far. One way to progress beyond the metaphors is to examine the contrast class. We can develop a better sense of autonomy by considering examples of actions that *lack* autonomy. Sometimes we act for the sake of desires that are not truly our own. We may be coerced, manipulated, or in the throes of a powerful addiction. In such cases, we demonstrate an autonomy deficit by acting on "alien" desires. We act for reasons that are not our own, desires that were implanted in us (to employ yet another metaphor). In the next section, we will develop our account of autonomy further by considering different explanations of this phenomenon. What, then, are alien desires, and how are they inconsistent with autonomy?

2.2.1 Alien Desires

If someone pointed a gun at you and demanded that you turn over your wallet, it is fairly obvious that your decision to hand over the money was not autonomous. You were being coerced. Coercion can be understood quite simply as being forced to make a particular decision on the grounds that you have no acceptable alternative.[5] Other kinds of autonomy deficits are not so simple to grasp. It is trickier to explain what is going on when an unwilling drug addict acts against her better judgment or when a brainwashed cultist sells her belongings and moves to the compound.

[5] There are various conceptions of coercion in the literature. This view comes from Allen Wood. For further discussion, see Wood (2014).

What exactly is going on when someone acts on an alien desire? Many of the contemporary theories of personal autonomy were developed with an eye toward offering explanations of what is going on in those scenarios. According to one popular family of theories, the above actions should be understood as involving a lack of coherence between the agent's motivational states. For the unwilling addict, she may feel a strong desire for the next hit, but there is an important sense in which she does not *want* to do what she ends up doing. Harry Frankfurt (1971) offers a model that explains this behavior in terms of first-order and second-order desires. First-order desires refer to things that we want in the ordinary sense: you want to go for a walk, eat a piece of chocolate cake, or read the next chapter of this book. These are things that you want in the first-order sense. But at the second-order level, you could talk about whether you want to have those desires or not. You could also talk about whether or not you want to act on your first-order desire. Perhaps you are trying to avoid sugar, so you want to stop having a first-order desire for the cake. You do not want to act on your first-order desire to eat the cake. Those are second-order desires and volitions. For Frankfurt, the unwilling addict lacks freedom (or autonomy)[6] because there is a tension between her first-order and second-order desires. She wants to use heroin, but, at the second-order level, she wants to stop having that first-order desire.

There are several theories of autonomy that require something similar to the kind of coherence that Frankfurt put forward in his influential paper. The unifying feature of "coherentist" theories of autonomy is this. An agent's action is considered autonomous "if and only if she is motivated to act as she does because this motivation coheres with (is in harmony with) some mental state that represents her point of view on the action" (Buss and Westlund 2018). For Frankfurt, the coherence in question concerns the agent's first-order and second-order desires. Gerald Dworkin offers a similar model according to which a "person is autonomous if he identifies with his desires" when engaging in higher-order reflection (Dworkin 1981, 212). Dworkin says that autonomy consists in "a second-order capacity of persons to reflect critically upon their

[6] Frankfurt's classic paper (1971) frames the discussion in terms of "freedom" rather than autonomy. But his analysis can certainly be used to explain autonomy. Indeed, Dworkin's work explains autonomy by employing Frankfurt's language of lower- and higher-order desires. And Frankfurt himself returned to the subject in his later writings where he extended his work on freedom by applying it to the concept of autonomy.

first-order preferences, desires, wishes, and so forth and the capacity to accept or attempt to change these in light of higher-order preferences and values" (1988, 20).

Gary Watson suggests a slightly different approach. He argues that autonomy should be understood in terms of consistency between the agent's desire and her evaluative judgments about what is most worthwhile. Watson is skeptical about characterizing autonomy in terms of desires. He says that agents "do not (or need not usually) ask themselves which of their desires they want to be effective in action; they ask themselves which course of action is most worth pursuing" (1975, 219). When someone acts compulsively (and thus lacks autonomy), Watson would explain this by saying that "the desires and emotions in question are more or less radically independent of the evaluational systems of these agents" (1975, 220). Considering the case of a kleptomaniac who steals compulsively, Watson concludes that "it is because his desires express themselves independently of his evaluational judgments that we tend to think of his actions as unfree" (Ibid.).

One important thing to note about these accounts is that the harmony in question often appears to be synchronic. It could be evaluated at a particular moment in time, looking only at a single timeslice of an individual.[7] In the moment when the addict uses, we could ask, following Frankfurt, whether or not her first-order desire is consistent with her second-order volition. Or, following Watson, we could ask if her action coheres with her evaluative judgments. In contrast to this approach, Michael Bratman (2003)[8] argues for an account that involves the agent's long-term plans about her life. Agents form a variety of plans throughout their lives, and one of the central tasks of practical reason is to fit different plans together in a consistent and stable way. To act autonomously, we must act in a way that is consistent with the long-term plans that were the result of a deliberative process.

In addition to emphasizing the role of long-term plans, Bratman also mentions the importance of the historical processes that gave rise to the

[7] This is true of Frankfurt's position, which is ahistorical, but Dworkin's view could not be described in this way. Although many theories of autonomy set out to discern whether or not a particular action or desire is autonomous, Dworkin argues that judgments about autonomy primarily concern evaluations of a person's entire life. He says that such questions must "be assessed over extended portions of a person's life. It is a dimension of assessment that evaluates a whole way of living one's life" (1988, 16).

[8] Cf. Bratman (2007, 2018).

agent's plans. He suggests that a "full story about human autonomy will also need to appeal to some sort of historical condition that blocks certain extreme cases of manipulation, brainwashing, and the like" (2003, 175–76). To see why this is an issue of vital concern, let us return to the example of the brainwashed cultist. Imagine that she forms her desire to move to the compound only after undergoing some powerful forms of manipulation. Perhaps she was subjected to an extreme form of operant conditioning. If the manipulation was successful, it is perfectly conceivable that her first-order and second-order desires might be in harmony. Not only does she have first-order desires to sell her belongings and move to the compound, she *wants* to have those desires and she wants them to determine her will. It seems clear, however, that we should not regard her decision as autonomous. These desires were the result of manipulation, so it would be a mistake to consider them autonomous.[9]

It is for precisely this reason that Dworkin (1976, 1988) stresses the importance of procedural independence when the agent identifies with her desires. It is not enough to merely identify with her desires; this identification must not be the result of manipulation. To address concerns like these, we might want a view of autonomy that requires us to evaluate the process through which the desire was developed. John Christman defends just such a view. In order to regard a desire as autonomous, Christman suggests that we must ask whether or not the agent resisted (or would have resisted) the development of the desire if she were to attend the process of its development (Christman 1991, 11). Furthermore, he adds that the lack of resistance must not take place "under the influence of factors

[9] In business ethics, a similar argument has been made about advertising and autonomy. Lippke (1989) argues that we should not be satisfied with the consistency between first- and second-order desires, given that advertising does more than simply create first-order desires. He suggests that advertising tries to induce an "uncritical acceptance of the consumer lifestyle as a whole," and if it succeeds in doing this, then we should not be surprised to find that our second-order reflection endorses the first-order desire (Lippke 1989, 39). After all, advertising sold us on the idea that satisfying such first-order desires will make us happy, so the mere fact that we do not repudiate the desire upon reflection is not enough to ensure that the desire is autonomous. For an overview of this argument, see Aylsworth (2020).

As we explain in the next chapter, the same point has been made about online manipulation: "[S]oftware agents can indirectly influence behaviour through second-order effects. Second-order effects include changes to the user's utility function through prolonged exposure to certain kinds of content (e.g., behavioural addiction) and changes to their doxastic attitudes (e.g., political polarisation)" (Keeling and Burr 2022, 257). See also Burr et al. (2018).

that inhibit self-reflection" and that the self-reflection must be (minimally) rational and involve no self-deception (Ibid.). We can safely assume that the cultist would repudiate the development of her desires if she were to understand the manipulative processes that shaped them.

But there are those who believe that Christman's historical approach does not go far enough when it comes to rooting out influences that undermine autonomy. One thing that all of the above approaches have in common is that they are "content neutral" and "value neutral." They are not committed to any particular theory of the good, and they make no judgments about the content of the decision. Each of those views would allow us to judge that a person is autonomous with respect to a desire even if we think that the desire in question is bad for the agent. They include no substantive requirements about attitudes or capacities (like self-trust or self-respect) that must be present when the agent identifies with her desires. They would also make it possible to view a desire as autonomous even though it was shaped by conditions of profound injustice. For instance, consider the situation of a woman in a very conservative society in the Global South, where oppressive gender norms have made it common for women to deprive themselves of food (even to the point of starvation) in order to make sure that their husbands and sons can eat far more than they need.[10] Similarly, we are likely to have reservations about the abused wife's preference to stay with her violent husband.[11]

Cases of this kind might pose a problem for content-neutral, or "procedural," models of autonomy. The woman who chooses to starve herself in order to overfeed her husband and son might get a pass on all of the procedural views discussed above. Using the Frankfurt/Dworkin model, she might approve of her first-order desire when she engages in higher-order deliberation. With Watson, she might argue that her action is perfectly consistent with her evaluative judgments. She might genuinely believe that it is more important to nourish her husband and son. On Bratman's view, we might find that her action is consistent with her long-term plans. Even Christman, whose view is explicitly concerned with the

[10] Natalie Stoljar (2014) uses this example to demonstrate how conditions of oppression can undermine autonomy (237), but the example was discussed earlier in Serene Khader's (2011) work on adaptive preferences. Khader cites reports from social scientists who have documented these gender norms in South Asia. See Papanek (1990), Sen (1990), and Ramachandran (2006). David Enoch (2020) also uses the example to show why he thinks that value-neutral models, like those defended by Christman and Dworkin, are inadequate.

[11] See Terlazzo (2016).

process of desire formation, might lack the resources to explain why her decision is not fully autonomous. When the woman looks back on the process that developed her desire, she might find moments in her upbringing when her mother taught her these values or when societal pressures reinforced them, but, absent any outright manipulation this might not be enough for her to repudiate it as an alien desire.[12]

But it stretches credulity to suggest that her desire to starve herself is autonomous. We should not think of her as the author of this desire. For precisely this reason, a number of philosophers—particularly feminists—have been critical of content-neutral, procedural views of autonomy. According to the procedural view, a desire is autonomous as long as the procedure that one uses to endorse the desire (e.g., by higher-order reflection) is independent of things like manipulation and coercion. Feminist views of autonomy go further, as they frequently stress the claim that the autonomous person must have "substantive" independence as well (see Stoljar 2000; Benson 1987; Oshana 2006; Mackenzie and Stoljar 2000). Substantive independence may take many forms. For instance, autonomous choices should not take place in a social context in which your set of options is highly constrained by things like oppressive gender norms. You must have a sufficiently wide range of options (Brison 2000). What is more, we cannot simply look at the procedure the agent uses to endorse her desire. We might ask questions about the content of the desire itself. Or substantive theories may require that the agent's conclusion not demonstrate a lack of self-respect or self-worth (Benson 1994). It has also been argued that full autonomy requires that a preference not be causally influenced by injustice (Enoch 2020).

Feminist critics have also argued that we inherited a flawed view of autonomy from the Enlightenment (especially from thinkers like Kant). They criticize the "atomistic" view on the grounds that it asks us to conceptualize the autonomous person as a fully self-sufficient individual who is bereft of any dependence on social relations. Natalie Stoljar suggests that this view "is abstracted from the social relations in which actual agents are embedded. Such a conception of the self is associated with the claim that autonomous agents are, and ought to be, self-sufficient, which in turn is associated with the character ideal of the 'self-made man'" (Stoljar

[12] Enoch (2020) faults Christman for precisely this point. Enoch says that Christman comes close to admitting outright that the historical view may yield the surprising conclusion that "oppression is consistent with autonomy" (Enoch 2020, 181).

2018). And since gender norms tend to involve expectations that women ought to value certain social relations (such as the family), this view "denies women, in particular, the social and political advantages associated with the label 'autonomous'" (Ibid.).

As a result, some feminist philosophers have defended "relational" views of autonomy. According to these views, a person's autonomy is not compromised by the fact that she is not fully self-sufficient. No one is or ought to be an island; we can be autonomous even if we recognize the moral significance of interdependence and social relationships. There is no denying that we are, to some extent, products of socialization. And given our dependence on one another, we ought to promote social arrangements that facilitate the development of autonomy. We should avoid restricting agents' choices in such a way that their seemingly voluntary choices demonstrate autonomy deficits, like the woman who "voluntarily" chooses to starve herself. Thus, in the political context, feminists who champion autonomy see it "as a valuable kind of individual freedom that political arrangements ought to promote" (Stoljar and Voigt 2022).

But this is precisely what leads some political and moral philosophers to hesitate from endorsing a substantive view of autonomy. According to certain views of liberalism, public reason requires us to refrain from offering justifications that are wedded to substantive value commitments that may not be shared by other members of our liberal democracy (Rawls 2005). The motivation for this reluctance is that we typically want to avoid forcing someone to live by another person's conception of the good. We ought to offer justifications that any "reasonable"[13] person could accept, regardless of the comprehensive doctrine to which she is committed.

On the flip side, there are those who are skeptical about this conception of public reason. There are some issues with the idealization that is involved in asking whether or not a "reasonable" person would consent to a particular claim. The basic idea behind public reason is that we want to offer justifications that citizens will consent to because they view them as reasonable. The authoritative force of policies under liberalism is grounded in the consent of the citizens; this is the source of their legitimacy. But, at

[13] It is important not to read too much into Rawls's use of the word "reasonable" here. According to his view in *Political Liberalism*, a reasonable person is one who is "ready to propose principles and standards as fair terms of cooperation and to abide by them willingly, given the assurance that others will likewise do so" (Rawls 2005, 49). For further elaboration, see Alm and Brown (2021).

times, this tends to abstract away from actual, non-ideal humans, because they may be unreasonable in a variety of ways (racist, sexist, etc.). So it requires us to ask questions about the consent of idealized, "reasonable" persons. But then it becomes unclear whether or not public reason can do the very thing it set out to do. It appears to obviate the role of consent. Joseph Raz (1990) puts this in the form of a dilemma. Either political consent comes from actual human beings, in which case it might be so weak that it cannot yield the very basic legal principles that we want from it (such as those barring racial discrimination). Or else, we must abstract from real citizens and derive consent from idealized reasoners, but then it becomes unclear what role consent is really playing. After all, the motivation of liberalism's commitment to public reason is to ground the government's legitimacy in the consent of the governed.[14]

For now, there is no need to settle these difficult questions. But it is important to highlight the tradeoff between procedural and substantive views of autonomy. This may not matter much in moral contexts where political issues are not at stake. Two people could agree that autonomy matters *morally* even if one of them holds a procedural view and the other prefers a substantive view. So this issue is of less concern for the present chapter. But the value or content neutrality of autonomy becomes much more important when considering the coercive power of the state, an issue that we will address in Chap. 6.

At this point, it would be helpful to take stock of these views in a way that will allow us to draw on them later. First, we can separate procedural, content-neutral views from substantive ones that involve commitments to particular values or that condemn certain social arrangements. Within the family of procedural views, there are several that involve questions about the coherence of the agent's motivational states. Such views might require us to ask about the agent's desires at a particular time or they may require us to take a deeper look at the person over time.

These are the procedural views that we have discussed so far:

Frankfurt/Dworkin: An agent is autonomous with respect to a desire just in case she approves of her desire when engaging in higher-order reflection. What is more, the higher-order deliberation must be free of external manipulation and coercion.

[14] See also Enoch (2013, 2015).

Watson: An action is autonomous if and only if it is consistent with the agent's evaluative judgments. It accords with her deeply held commitments about what is most worthwhile.

Bratman: An action is autonomous just in case it harmonizes with her long-term plans.

Christman: Autonomous desires are ones that the agent would not reject if she were to rationally reflect on the development of the desire.

Finally, it is important to note a subtle distinction that is present in these summaries. We can and should distinguish between different domains of autonomy. At times, it may be important to ask whether or not a particular *desire* or *action* is autonomous, and in other cases we may inquire about autonomy on a larger scale and ask what it means for a *person* to be autonomous. Procedural views could be fruitfully applied in each domain.[15]

It would be difficult to provide a similar overview of substantive theories of autonomy because there are too many of them, and the details vary widely. At their core, substantive theories reject the content neutrality of procedural approaches by proposing certain constraints. There is a distinction between theories that constrain the "content of autonomous preferences, values, or commitments (strong substantivism)" and those that merely put constraints on "the self-affective attitudes required for self-governing agency (weak substantivism)" (Mackenzie 2022, 36). A strong substantive theory might rule out certain options entirely (for instance, the choice to be a subservient housewife), whereas weak substantive theories may restrict themselves to asking whether or not the agent's decision was fully autonomous because her deliberation was undermined by a lack of self-trust or self-respect (see, e.g., Govier 2003; McLeod 2002).

Our aim in this section is not to defend or attack any of these views of autonomy. On the contrary, we are open to all of these accounts, and our definition of "humanity"—the ability to set and pursue one's own ends—is meant to be broad enough that it is consistent with a variety of approaches. So although our view of autonomy is not wedded to any particular account, it will be useful to draw on various features of these accounts in order to explain why a particular action or desire is at odds with the agent's ability to set and pursue her own ends. In some instances,

[15] Some views, like Christman's, are more explicitly concerned with autonomous and alien desires, whereas Dworkin's view places greater emphasis on long-term considerations about persons.

it may be useful to invoke Frankfurt's language of second-order desires. In other places, we may deploy Christman's historical model, or we might use Watson's approach and ask questions about the agent's evaluative judgments. We did not provide explanations of these views in order to settle the question of which one does the best job of defining autonomy. Instead, these accounts will provide a conceptual toolkit that we will draw on throughout the book in order to better understand how smartphones and the attention economy undermine our autonomy. In the next section, we revisit our definition of "humanity" in order to provide a more concrete understanding by means of the views we have explored in this section.

2.2.2 Humanity Revisited: The Ability to Set and Pursue One's Own Ends

Recall once more, that our definition of autonomy throughout the book will correspond to what Kant has in mind when he uses the word "humanity." This is the rational agent's capacity to set and pursue her own ends. We may begin by noticing that this definition comprises two importantly distinct elements: capacity and authenticity. The capacity condition refers to the agent's *ability* to set and pursue ends, and the authenticity condition requires that the ends be, in some sense, *her own*. It is common for accounts of autonomy to note this distinction and to regard these two conditions as separately necessary and jointly sufficient for autonomy.[16] If an agent lacks either one of them, then she is not fully autonomous. If she has both, then she is.

2.2.2.1 Agential Capacity

We will begin with capacity. For starters, to qualify as an autonomous agent, someone must fulfill the baseline criteria for rational self-government. The sense of rationality at issue here is fairly minimal. It does not require a person to be perfectly logical and consistent, but it does require basic rational competence: the capacity to make fitting inferences from available information, to evaluate and reconsider one's commitments, to be self-reflective, and to respond appropriately to

[16] For instance, Christman (2011) separates the conditions of "competence" and "authenticity."

inconsistencies.[17] In addition to the baseline criteria for *setting* ends, it is equally important that the agent satisfy the criteria for *pursuing* ends. She must be capable of forming intentions, adhering to the plans she sets (which requires the ability to merge plans), adapting to changes in her environment, making use of available resources, etc.[18]

It is in virtue of these baseline capacities that we do not think of infants or young children as fully autonomous. They may have the potential to become autonomous as they grow older, but, at early stages of development, children lack the capacities necessary to be autonomous. It is precisely because of this lack of capacity that paternalistic constraints are appropriate when applied to children but often inappropriate when applied to adults. A small child might want to play with sharp knives or eat nothing but ice cream, but we are under no obligation to respect such choices.

But this is only the starting point. Once the baseline criteria have been satisfied, there are additional questions about the capacities an agent must have in order to successfully exercise her rational agency. Paternalistic interventions, such as the parent's restriction of the child's dinner options, are only one of many ways that someone might undermine another person's sovereignty. We can easily imagine adults, who possess all the rational capacities that infants lack, but who are unable to successfully exercise their capacities because of external constraints.[19] For instance, someone may be held hostage, coerced by threats of violence, or misled through deception. There are many ways to prevent people from successfully

[17] Formal or procedural views of autonomy (like Christman's and Dworkin's) typically construe baseline epistemic capacities in a fairly minimal way. There are those like Jonathan Knutzen who see this as a target for criticism. Knutzen (2020) argues that autonomy requires a more substantive capacity for reasons-responsiveness.

[18] When Kant discusses moral duties regarding our own humanity, he stresses the importance of this point. Respecting humanity in ourselves requires us to cultivate our capacities so that we are capable of pursuing a wide variety of ends (*MS* 6:392).

[19] Feinberg (1989) distinguishes between the capacity for self-government, the successful exercise of this capacity, the personal ideal of self-government, and the moral and political claims that one may assert in order to demand that others respect her sovereignty over herself. See also, Rubel et al. (2021, 24).

exercising their capacities, and some of these are fairly obvious.[20] Other constraints may be more subtle, however.

Once again, feminist perspectives on this issue are rich with examples of autonomy deficits that are difficult to diagnose. Trudy Govier (2003) and Carolyn McLeod (2002) defend the claim that "self-trust" is necessary for autonomy.[21] Someone might satisfy the baseline criteria for rational agency but nevertheless be unable to successfully exercise those capacities because those around her (or various traumatic experiences) have made it difficult for her to trust her own judgment. So although she has rational capacity in principle, she routinely defers to the judgment of others in practice because she lacks self-trust. Benson (1994), drawing on the film *Gaslight*, provides an example of a woman whose husband—a physician whom she trusts—convinces her that she is "unstable" and "hysterical." As a result, she begins to feel helpless and disoriented, further eroding her self-confidence. Benson concludes that the woman's ability to make her own decisions has been severely compromised, even though her rational faculties are fully intact.

This gives us yet another condition to consider when evaluating an agent's rational capacities. So far, we have established that rational agency requires certain baseline capacities (reasoning with evidence, forming coherent plans, etc.), freedom from external constraints (being held hostage, coerced by threats of violence, etc.), and freedom from influences that would undermine the agent's ability to successfully exercise her rational capacities (gaslighting, deceit, etc.). But this still might not exhaust the set of conditions required for autonomy.

Consider the example of a boy in West Virginia in the late 1800s. He is bright, talented, and successful in school. But almost everyone he knows is a coal miner, and he grows up believing that coal mining is the only way for him to make a living. When he comes of age, he is free to choose between one of three different coal mines. He lacks the financial means to move from the area (and he has no clue what he would do for a living if he

[20] Of course, not all constraints are inconsistent with autonomy. Laws constrain one's options in many ways, but we would certainly not want to suggest that the existence of the state is inconsistent with autonomy. On the contrary, we side with Kant in believing that the state is a necessary condition of personal autonomy. Kant argues that individuals cannot coexist with one another in a state of freedom without the state.

[21] For a helpful overview of feminist perspectives on autonomy, see Stoljar (2018).

left). So he weighs his options, evaluating the different pay and hours, and he chooses to work at one of the mines.[22]

Someone could say that he was free, in some abstract sense, to set and pursue any ends that he wanted. There were no immediate or apparent constraints on his decision. No one held him hostage or put a gun to his head. But if his freedom amounts to nothing more than choosing between one of three coal mines, we are likely to think that this falls short of full autonomy. We may agree with Susan Brison (2000) in thinking that in order to make autonomous choices, "we need to have a range of significant options to choose from, and so we need to live in a society that makes these possible" (2000, 283).[23] She points out that a wide variety of things might promote or inhibit our access to a range of options. This would include "personal, familial, social, political, and economic relations with others" (Ibid.). Furthermore, she rightly recognizes that we must be able "to perceive the availability of these options and to recognize their achievability *by us*" and that this requires us "to live in a culture in which the norms and expectations do not preclude such recognition" (Ibid.). The boy in West Virginia may have felt his options were restricted for any number of reasons. It may have been the result of pressure from his family and friends. Maybe he recognized that becoming a physician is an option for some, but he felt that such a thing was not a possibility *for him*.

This gives us a fourth condition of rational capacity. It is not enough to have a choice; one must choose between a sufficiently wide range of options.[24] For this reason, Rubel et al. (2021) distinguish between local and global autonomy. To have local autonomy, there must be a domain in

[22] Joseph Raz (1986) provides examples of even more restrictive situations. A young man falls into a pit, and his autonomy is limited to free decisions about when to nap or where to lay his head. He may have free rein in this domain but he clearly lacks autonomy because his set of options is so constrained (1986, 373). Similarly, Nussbaum (2011) and Sen (1999) show how severe poverty limits autonomy by constraining options and capabilities.

[23] Of course, Brison is not the only person to recognize the range of options as a condition of autonomy. Joseph Raz identifies three conditions: "appropriate mental abilities, an adequate range of options, and independence" (1986, 372).

[24] It is worth noting that one's autonomy is not *always* enhanced by increasing the number of available options. Although having too few options constrains the agent's ability to set and pursue her own ends, psychologists have observed that people often fail to make good choices when they are overloaded with options. In *The Paradox of Choice*, psychologist Barry Schwartz argues that the overabundance of options can lead to decision paralysis, and people often fail to make choices that harmonize with their evaluative judgments. People frequently end up being less happy with their decision when they have too many options. This means that having too *few* options is a constraint on your autonomy, but having too *many* options makes it less likely that you will successfully exercise your autonomy. See Schwartz (2009).

which the agent is free to "make decisions about actions with immediate effect and may be able to ensure that those decisions comport with their values" (2021, 23). Rubel et al. consider the employee Ali, whose employer routinely assigns her to work on projects that she does not enjoy, and she has very little say over what she does at work. But, in general, she is a very talented person who could easily find another job and who is perfectly capable of leaving. With respect to her job, Ali lacks autonomy, but she is autonomous in the more general sense. She has global autonomy but lacks local autonomy (Ibid.). They contrast Ali with Bari whose global options are severely restricted. Bari is all but forced to care for her siblings and older relatives. She cooks and cleans for them. But she has incredibly wide latitude when it comes to fulfilling those obligations. So Bari has local autonomy but lacks global autonomy (Ibid.).

The fourth condition (having a sufficiently wide range of options) underscores an important sense in which autonomy must be seen as relational. No one's capacities are developed in a vacuum. Culture, upbringing, and education are inextricably linked to the development of our capacities. They shape our sense of what is possible and what we can achieve. Indeed, even Kant, who is often maligned for defending an excessively atomistic view of autonomy, was sensitive to this point. As we will explain in Chap. 5, Kant argues that the cultivation of "humanity" through education is a moral imperative. Although it is true that Kant thinks each individual is morally obligated to cultivate her own capacities, he recognizes that this is something that we must undertake collectively, a point he repeats throughout his writings on religion, history, and anthropology.[25] What is more, he sees the collective perfection of the entire human species as the final end of history.[26] In his lectures on pedagogy, he says, "The human being can only become human through education. He is nothing

[25] Wyrębska-Đermanović puts this point nicely: "Nevertheless, it would be a mistake to consider Kant's account of moral perfection outside of the context of living within a society, among fellow human beings and in constant interaction with them. Even individual moral progress is not independent of the character of these relations, as Kant repeatedly stresses in his writings on religion, history and the methodology of practical reason … Moreover, besides the moral perfection of the individual, Kant conceptualises the idea of perfecting the entire human species and represents this ideal as the final end of history … Moral development has therefore two dimensions—individual and collective and in each Kant gives us hints about the possible ways of facilitating progress" (2021, 1046).

[26] This idea features prominently in *Idea for a Universal History with a Cosmopolitan Aim*. See Kant (2007a, 8:27ff).

except what education makes out of him. It must be noted that the human being is educated only by human beings, human beings who likewise have been educated" (*UP* 9:443).[27] This is one of the many ways that our autonomy is promoted (rather than inhibited) by social interdependence.

With those four conditions in place, we should have a much better understanding of the capacities required for autonomy. It requires

(1_a) baseline competence (the ability to form intentions, make coherent judgments, etc.),
(2_a) a lack of external constraints,
(3_a) the absence of cognitive inhibitions (such as self-doubt), and
(4_a) a sufficiently wide range of options (with the ability to pursue them effectively).

Of course, all of these things come in degrees. External constraints can be more or less severe. Questions about self-doubt, self-trust, and baseline competence will rarely be answered by a simple yes or no. You may have days when you are racked with self-doubt and imposter syndrome, but you may be teeming with confidence at other times. Or you may experience both of those feelings at the same time but with respect to different domains. Some students are totally confident in classes that require them to write essays, but their "math phobia" causes them to experience self-doubt in others. So the presence or absence of these conditions of autonomy is a matter of degree.

2.2.2.2 Authenticity

For now, this should be enough to understand what kinds of capacities matter when it comes to autonomy. Authenticity is a separate issue, but, as we will see, they are closely related. For instance, interfering with someone's capacities (e.g., by gaslighting them) could lead them to adopt

[27] In our own writing, we try to be very careful about the use of gendered language. Some of our quotations of Kant use gendered language, however. This passage from Kant's lectures on pedagogy (über Pädagogik) uses gendered language (the pronouns are 'er' and 'ihm', but he uses 'der Mensch' here which could be read as 'humanity' more broadly). We have chosen not to alter the quotes or change the translations. We have chosen to bring attention to Kant's sexism rather than erase it by altering the quotations. Pauline Kleingeld (2019) argues persuasively against the temptation of replacing Kant's words with gender-inclusive language. She points out how doing this runs the risk of construing Kant as saying things that he did not say, and it would be better to preserve his language so that we can observe the tension between his egalitarian principles and his inegalitarian applications of them.

inauthentic or alien desires. To understand authenticity, it would be helpful to recall the lessons from the previous section. Your authentic desires are ones that did not result from manipulation or coercion. They are generally consistent with your long-term plans (Bratman), your higher-order reflection (Dworkin), and your evaluative judgments about what is most worthwhile (Watson). Finally, you should not feel alienated from your desire when you reflect on the process through which the desire was developed (Christman).

Once again, these things come in degrees. All of our desires are subject to *some* amount of external influence. We should not make the mistake of thinking that desires are one's own only if they were developed in the complete absence of outside influence. That would set the bar too high. So where should the line be drawn? Perhaps we should begin with the clearest cases and work our way toward those that are more subtle and difficult. Arguably the clearest and most compelling case would be coercion. As noted earlier, when you decide to hand your money to the armed robber, it is very clear that this desire is not authentically yours. This desire is being forced upon you because the alternative is unacceptable (Wood 2014).

So when it comes to the first condition of authenticity (being free from manipulation and coercion), it is not hard to see why desires that result from coercion are inauthentic. There is a very clear sense in which you are being compelled to adopt someone else's desire. Manipulation is more difficult to grasp. For starters, there are fundamental debates about what manipulation even is. Two people might agree that manipulation is at odds with autonomy even though they have radically different ideas about how to define the concept. Yet again, it would be helpful to begin with some clear cases of manipulation.

Consider Iago's manipulation of Othello. Iago feels that Othello unfairly passed him over for promotion, so Iago formulates a plan to manipulate Othello into demoting his lieutenant, Cassio. Iago accomplishes his aim by deceiving Othello in a variety of ways. First, Iago acquires the handkerchief that Othello gave to his wife, Desdemona. Iago lies to Othello, telling him that he saw Cassio with the handkerchief (implying that Cassio is having an affair with Desdemona). This sparks jealous rage in Othello, and Iago convinces him to eavesdrop on Cassio's conversation. Othello overhears Cassio talking about his affair with his mistress Bianca. Iago deceives Othello once more by leading him to believe that Cassio was

talking about Desdemona. Othello becomes so angry that he orders Iago to kill Cassio.

It seems safe to say that Iago manipulated Othello. We can also see *how* Iago accomplished his aim: via deception. He lied to Othello about the handkerchief and about the meaning of Cassio's conversation. So the first lesson that we can draw from this case is that you can manipulate someone by deceiving them. And although Iago deceived Othello by telling many outright lies, this is certainly not the only form of deception. Allen Wood (2014) claims that telling flat-out lies is manipulative because it "feeds the person false information, on the basis of which he makes choices the person presumably might not have made if he had known the truth" (35). In addition to outright lies, Wood goes on to note that "Other, subtler forms of deception—misleading, encouraging false assumptions, fostering self-deception—do this in more devious ways" (Ibid.).

In addition to deception, someone might manipulate you by pressuring you in various ways: guilt, shame, browbeating, social pressure, etc. Now, if the pressure is so extreme that they leave you with no acceptable alternative, then this would rise to the level of coercion. But manipulation often falls short of coercion. Consider the following case. Sofia is a college student who wants to major in philosophy. After thorough deliberation, she decides that her only criterion for choosing a major is studying what she loves, and she is certain that her passion lies in philosophy. When she breaks the news to her parents, they try to talk her out of it. After rational persuasion fails, they turn to emotional appeals, telling her, with tears in their eyes, that they would be heartbroken if she does not follow in her mother's footsteps to practice medicine. They even threaten to stop paying her tuition. Assuming that Sofia has an acceptable alternative (such as going to a cheaper college or taking out student loans) this would not amount to coercion. But it certainly seems manipulative.

Wood encourages us to notice what is distinctive about manipulative "inducements." He says that they "offer the person 'the wrong sort of reason'—a sort of reason that the person would not endorse on reflection, if behaving rationally and operating with normal, healthy motivations intact" (Ibid.). The young woman had already decided that the only thing that mattered to her was following her passion. Pleasing her parents and avoiding student loans may be reasons to study medicine, but, for Sofia,

they are not the right sort of reasons.[28] Seeing the response from her parents provokes so much emotional distress that Sofia's judgment is clouded and she ends up studying medicine. She later regrets the decision, and she feels that she acted on the basis of reasons that she could not rationally endorse.

These two examples are meant to show two possible means of manipulation. Iago manipulates Othello by deception, and the parents manipulate their daughter by using emotional appeals to get her to act on the basis of reasons that she does not reflectively endorse. It would be helpful to consider one last class of examples before offering a general account of manipulation. Critics of persuasive advertising frequently argue that marketing is morally problematic insofar as it undermines our autonomy by manipulating our desires.[29] It is outside the scope of our project here to weigh in on that debate, but it would be helpful to consider one of the most popular examples. Commentators frequently cite the (apocryphal)[30] story of a New Jersey movie theater that drastically increased popcorn and soda sales by subjecting moviegoers to subliminal advertisements. The moviegoers did not know that they saw an advertisement at all, but it caused them to get up during intermission to purchase refreshments. Roger Crisp argues that the moviegoers are not acting autonomously when they do this; he quips that they are acting "**automaton**ously" (1987, 413; emphasis added). That is, their behavior resembles that of an automaton rather than that of an autonomous agent. This is because the subliminal advertisement induced their actions without engaging their rational capacities in any way. They were not able to reflect on their reasons because the influence was affecting them at a subconscious level.

The debate concerning persuasive advertising is ongoing, and there is no consensus in the literature about whether or not advertising is

[28] To better understand what it would mean to have the "right" and "wrong" kinds of reasons, consider David Enoch's example of being offered a lot of money to believe that the number of stars in the universe is even. If he makes you this offer, you certainly have a reason to believe this, but it is not the right kind of reason (2020, 177). This is an important point for Enoch, since he argues that adaptive preferences are ones that are formed for reasons of the wrong kind. Enoch's point is that adaptive preferences that are based on reasons of the wrong kind are pro-tanto irrational.

[29] See Aylsworth (2020) for an overview of these arguments.

[30] Although this story is frequently discussed, many have argued that the initial report was fabricated by James Vicary. Karremans et al. write: "Nobody has ever replicated Vicary's findings and his study was a hoax" (2006). See also Rogers (2001).

manipulative. The mere fact that advertising has an influence on our desires is insufficient for establishing the claim that it manipulates us. Something further must be established. Anne Cunningham (2003) argues that some critics of advertising have created an unreasonably demanding standard for autonomy. She offers an alternative approach for addressing concerns about autonomy and advertising. She suggests that we should consider a desire autonomous (or alien) by evaluating it according to the model defended by Noggle (1995). In Noggle's view, alien desires are the product of "discordant quasi-beliefs" (1995, 65). Quasi-beliefs are representational states that can function like beliefs insofar as they promote desires or actions, but they lack several important properties belonging to "straightforward beliefs." Straightforward beliefs are integrated into our broader web of beliefs; "are formed by cognitive epistemic means and are subject to cognitive epistemic control" (60); are available to us upon introspection; and are ones that we are, other things being equal, willing to assert to ourselves (Noggle 1995, 59–60). Quasi-beliefs lack all of these features, and a quasi-belief is discordant just in case it is inconsistent with some straightforward belief.

To see how this would work, imagine a case where advertising is capable of achieving what the critics suggest that it is guilty of doing. For example, let's say Rosa's desire to drink Coca-Cola was brought about because the ads full of young, beautiful, happy people successfully led her to associate Coca-Cola with her unconscious desire for youthful exuberance and joie de vivre. If the desire driving her purchase is truly unconscious, then she would not be able to rationally deliberate over her choice or endorse it through higher-order reflection. The desire for Coca-Cola was brought about by manipulating her unconscious desires. Using Noggle's model, we could say that her action was prompted by a discordant quasi-belief. She surely does not believe that drinking Coca-Cola will bring back her youth or make her life as fun as the advertisement, but that is precisely the content of her quasi-belief. Of course, those who defend advertising, like Cunningham (2003) and Arrington (1982), argue that advertising rarely does what the critics claim. But if it *did* accomplish what critics like Crisp (1987) and Wood (2014) suggest, then we can see why it would be inconsistent with autonomy.

Examples of manipulation extend well beyond deception, emotional appeals, and subliminal/subconscious advertising. In some cases, all that is required is the presence of an influence that leads to a decision that falls short of the agent's ideal form of rational deliberation. For instance,

consider the realtor who bakes cookies before the open house.[31] Of course, prospective buyers are likely to feel welcomed by such a pleasant smell, but if Hanlon were to make an offer on the house on the grounds that the smell brought back fond childhood memories of feeling cozy in his abuela's house, then it is easy to see how he is falling short of his ideal for deliberation. The fact that the house smelled like cookies was not a good reason for buying it.[32]

In each case described above, manipulation involves influencing people's choices in ways that interfere with their capacity for rational decision making. Manipulation makes it more difficult for the agent to make choices that are consistent with her evaluative judgments. Deception does this by tricking people into acting in ways that they would not want to act if they had accurate information. Emotional manipulation works by giving people inducements they would not endorse on reflection.[33] If advertising is indeed manipulative in some cases, it accomplishes this by changing desires through means not open to introspection or rational consideration. We could conclude, as Allen Wood does, that manipulation circumvents or undermines rational agency in some way (Wood 2014).

It is important to avoid overgeneralizing about manipulation. We would not want to suggest that all forms of non-rational influence are manipulative. That would set the bar much too high for autonomy. In many cases, we may appeal to someone's emotions without manipulating them. You might want to influence the behavior of a close friend, so you decide to explain how their actions have hurt you in the past. Assuming that your friend cares about you, their emotional response to your suffering should count as an appropriate reason to change their behavior. You

[31] Barnhill (2014) uses this example, and she cites evidence from realtors who explicitly encourage this practice (58).

[32] Discerning between good and bad reasons does not require a commitment to an objective account about which kinds of reasons are good and bad. It can simply be the case that such reasons are not adequate for the agent in question. In this case, the prospective homebuyer has other values (the affordability of the house, its location, size, etc.). For such a buyer (indeed almost any buyer), the fact that the house smells like cookies is not a good reason to buy the house.

[33] It is important to note that not all emotional appeals are manipulative. Sometimes our reasons that are based on emotions are the right kinds of reasons. It would also be a mistake to draw a sharp distinction between reason and emotions. Rationalist thinkers in the early modern period surely overemphasized the separateness of the two. The key question when it comes to manipulation is whether or not the appeal to emotion undermines the agent's capacity for rational decision making.

have influenced their behavior by means of an emotional appeal, but you have not manipulated them. It should also be pointed out that manipulation can be covert or overt. Othello covertly manipulates Iago, but Sofia's parents make an overt effort to manipulate her. Covertness is not a necessary condition of manipulation.[34]

With these views in hand, it is not difficult to see how manipulation and coercion are at odds with autonomy. Given that autonomy requires authenticity, our chosen ends must be, in some sense, our own. When we are coerced or manipulated, we act for reasons that are not our own. As we mentioned above, acts of coercion and manipulation are often seen as morally wrong, but it is worth noting that they are not wrong by definition.[35] As Roger Crisp points out, a skillful actor deftly manipulates our emotions, but this is not morally objectionable (1987, 414). Allen Wood says the same is true of coercion. A police officer might point a gun at the bank robber, coercing him to stop what he is doing, but coercion of this kind does not seem to be unjust or immoral (2014, 27).

In some cases, we can employ the local/global distinction to understand why a manipulative or coercive action is not immoral. For instance, parents may rightly manipulate their child into eating a healthy, balanced diet. They manipulate her into trying a wide variety of foods. A kindergarten teacher might manipulate (or even coerce) a child into sitting quietly and listening. In both cases, the parents and teacher are doing something that is at odds with the child's *local* autonomy.[36] With respect to that particular choice, the child's autonomy is constrained. But both actions can be seen as promoting the child's *global* autonomy. By learning to enjoy a variety of foods, the child will have more dietary options in the future. Her

[34] Klenk (2022) makes a powerful case for this point, against the claim from Susser et al. (2019a, b) that the hallmark of manipulation is hidden influence.

[35] This sets our definition apart from explicitly moralized concepts of manipulation. Marcia Baron's 2003 APA address comes very close to putting forward a moralized version of manipulation. Her condemnation of manipulation has both Kantian and Aristotelian elements.

[36] It is also important to note that constraining the local autonomy of the child does not fundamentally disrespect their humanity. Some moral theories, like consequentialism, could condone this action by claiming that the greater good (promoting the child's global autonomy) outweighs the local harm (undermining her local autonomy). But as Patrick Frierson (forthcoming) notes, this move is not available to Kantians who are committed to the claim that we must *always* respect humanity as an end. Thus, Kant's theory of education is predicated on the idea that the child's capacities are not fully developed, so she does not possess humanity in the relevant sense. Frierson's comments echo Schapiro's (1999) Kantian, nonconsequentialist justification of treating children paternalistically. Cf. Schapiro (2003).

choices will be less constrained than if she were to become a picky eater. Similarly, by making the student sit down and listen, the teacher is helping the student develop her capacities and thus promoting her autonomy.

This sums up the first condition of authenticity. Authentic preferences are ones that are free of manipulation and coercion. Just as we argued above, when discussing the conditions of agential capacities, most of these conditions should be understood as existing on a spectrum rather than as binaries. The more that a preference is the result of manipulation, the less autonomous it is. If questions of this kind are seen as a binary, the conclusions run the risk of being implausibly strong. Once more, we must avoid overgeneralizing about manipulation and autonomy. The account would be implausible if it condemned every attempt to exert non-rational influence on someone.

The second condition of authenticity is one that was discussed extensively above in the context of "alien" desires. Your authentic desires and preferences are ones that are consistent with your deep commitments. They demonstrate a certain kind of internal coherence that shows that you are the author of your own life story. We may side with Frankfurt and Dworkin by understanding this coherence in terms of first-order and second-order desires. With Watson, we may use the model of evaluative judgments, or we could follow Bratman by understanding this in terms of the agent's long-term plans. Desires and preferences are authentic to the extent that you identify with them.

In addition to identifying with our desires, authenticity also seems to require the absence of alienation. At first glance, it may seem that non-alienation is already a feature of internal coherence accounts. But critics like Christman insist that they are distinct. One important feature of Christman's account is that he regards mere "identification with" a desire as too low of a bar for personal autonomy.[37] This is why he argues that the agent must evaluate the historical process through which the desire was developed. In some cases, an agent might identify with a desire (e.g., she regards it as her own when engaging in higher-order reflection), but she

[37] More precisely, he argues that the identification approach is either too strong or too weak. If the account involves merely identifying with the desire, then Christman thinks it is too weak. Identification of this kind is not enough to guarantee autonomy. On the other hand, if identification involves robust evaluative endorsement, then he thinks it is too strong.

would feel alienated from it if she were to reflect on the historical development of that desire.[38] Christman writes:

> Consequently, I suggest that the proper test for the acceptability of the characteristic in question is one where the person does not feel deeply *alienated* from it upon critical reflection. Alienation is not simply lack of identification, in that I can fail to identify with a trait but not be alienated from it. I can simply be indifferent to it or undecided about it. Alienation is a stranger reaction; it involves feeling constrained by the trait and wanting decidedly to *repudiate* it. (2011, 143)

He goes on to note how alienation goes beyond identification models in that it involves both "judgment and affective reaction" (144). He contrasts this model with identification accounts that focus too narrowly on cognitive judgments. This is the third condition of authenticity. Your desire is authentic only if you are not alienated from it. Furthermore, you would not feel alienated from the desire if you were to reflect on the process that gave rise to it. The brainwashed cultist is a helpful example. She might identify with her desire to move to the compound, but she would likely repudiate the desire if she were to reflect on the manipulative processes that shaped her preferences.

Finally, as we explained above, there are those who criticize Christman for not going far enough when it comes to issues involving oppression and adaptive preferences.[39] To see why, we can return to the case of the woman who chooses to starve herself in order to overfeed her husband and son. This woman might identify with her desire, and she might not repudiate the desire when reflecting on the processes that developed it. In the autonomy literature, this is often referred to as an "adaptive preference." Her preference is "adaptive" in the sense that it was shaped by restrictive conditions that limited her set of options.

Indeed, the broadest accounts of adaptive preferences describe them in just such a way. According to those views, adaptive preferences are ones that were shaped, at least in part, by constraints placed on someone in virtue of their circumstances. But when using such a broad definition, it is difficult to see why anyone would regard them as problematic. As Rosa Terlazzo points out, her preference for a wheat-based diet is likely the

[38] Christman makes this point several times. He tells us that he is trying to shift away from "identification with" a desire in favor of a model of non-alienation (2011, 143, 214).

[39] See, for example, Enoch (2020).

result of growing up in the US, and she might very well prefer a rice-based diet if she had grown up in China (2016, 206). But that preference does not present any moral concerns. Many of our preferences are adaptive in that sense. Terlazzo notes that we might feel quite differently about a different set of cases:

> [T]he case of the woman who thinks that her husband is justified in beating her, the poor person who has no desire for an education, the woman who wants to undergo breast augmentation to make herself more attractive to men, or the man who refuses treatment for depression because he takes stoicism to be central to masculinity (Ibid.).

Now let us ask: what is it, exactly, that distinguishes the problematic cases of adaptive preferences from ones that are not so troubling?

According to one popular view, adaptive preferences are worrisome only in cases where the preference is *bad for* the agent in some way.[40] It is bad for the abused wife to believe that she deserves to be hit; it is bad for the woman to starve herself; and so on. When Aesop's fox tells himself that the grapes are sour simply because they are out of reach, his preference does not seem to be adaptive in the problematic sense.[41] After all, he might very well be better off believing that the inaccessible grapes would taste bad. Otherwise he may agonize over what he is missing. Similarly, Nussbaum provides an example of a child who comes to prefer a career in philosophy after learning that she was not cut out for being an opera singer (2001a). Again, this does not seem problematic. So what exactly is distinctive about the adaptive preferences that are morally worrisome? One way of explaining why some adaptive preferences are bad for agents would be to suggest that adaptive preferences are instances of autonomy deficits. But this account runs the risk of vicious circularity. Which adaptive preferences are the bad ones? The ones that undermine autonomy. Which ones undermine autonomy? The bad ones.

In order to escape this circle, we must provide a procedural account of adaptive preferences that would explain why some adaptive preferences undermine autonomy while others do not. Terlazzo (2016) defends a view of autonomy that is mostly procedural but which includes a couple of

[40] Nussbaum understands adaptive preferences in this way. See Nussbaum (1993, 2001a, b, 2011).

[41] Elster (1983).

substantive constraints. Recall that substantive theories of autonomy are ones that do not remain content neutral when making judgments about autonomy. For instance, if a feminist view rules out the possibility of autonomously choosing to be a subservient housewife, then it would not be content neutral; it would be substantive.[42] Interestingly, Terlazzo's view of autonomy does not rule out any preferences with respect to the value of their content, but it does require the agent to choose from a set of valuable, live options. In addition to requiring basic capacities and continence, she argues that a preference is autonomous if and only if:

(1_b) the agent reflectively endorsed the preference "at some point in its formation,"
(2_b) her reflection took place "in the presence of recognized alternatives,"
(3_b) "at least some of these alternatives were valuable ones," and
(4_b) "some of these valuable alternative options were live ones (that is, they were ones that X could reasonably see herself exercising, given her current values and ambitions)" (Terlazzo 2016, 215).

[42] Again, the potential problem with such accounts is that they might undermine the legitimacy of political interventions. Consider Mill's discussion of the Mormon woman who wants to be in a polygamous marriage. If we rule this out as non-autonomous, we might be overlooking the possibility that her desire (adaptive though it may be) is consistent with her deeply held religious values. She may very well be autonomous with respect to this desire. Mill might be right to suggest that it would be unjust to deploy the coercive power of the state to prevent her from entering into such an arrangement. This is one reason why it is important to distinguish between the normative *ideal* of autonomy and the concept of autonomy that grounds the right of non-intervention. Some agents might fall short of autonomy (if, for instance, their preferences are adaptive), but we will nevertheless be required to respect their choices, if they are autonomous enough to make decisions for themselves.

Interestingly, Kant rejects polygamy in his discussion of marriage in the Doctrine of Right. He argues that marriage ought to involve equal possession between partners and that polygamy wrongs the woman by giving her an inferior status: "For the same reasons, the relation of the partners in a marriage is a relation of *equality* of possession, equality both in their possession of each other as persons (hence only in *monogamy*, since in polygamy the person who surrenders herself gains only a part of the man who gets her completely, and therefore makes herself into a mere thing), and also equality in their possession of material goods" (*MS* 6:278).

She uses this view of autonomy in order to argue that adaptive preferences should be understood as core preferences that lack autonomy precisely in virtue of failing to meet one of these conditions.[43]

According to this view of adaptive preferences, we do not need to worry ourselves with questions about whether or not the preference is bad for the agent. We can also avoid making substantive claims that abandon content neutrality. Terlazzo does not require autonomous preferences to be ones that promote value for the agent; she only requires that the agent make her choice among a set that includes a few valuable, live options. So when a kidnapping victim develops Stockholm syndrome, we can safely say that her preference to stay with the kidnapper is adaptive in the problematic sense. She made this choice only after judging that she had no hope of escape. According to the well-being view of adaptive preferences, it might be hard to make such a judgment. It is difficult to judge whether or not the preference is bad for the victim. Perhaps Stockholm syndrome was a coping mechanism that allowed her to survive the traumatic situation. In that sense, the preference might have been good for her (at least in a local sense).[44]

This view allows us to render an intuitively plausible verdict about the examples we mentioned. On the one hand, we can see why the preferences of the abused wife or the kidnapping victim are not autonomous. On the other hand, we can explain why there is nothing wrong with the fox's preference to pursue food other than the grapes. The fox chooses to abandon his preference for the grapes only after realizing that they are unattainable. Not only is this perfectly rational, it is consistent with autonomy.[45] As long as the fox had other live options (e.g., berries that are closer to the ground), his choice was not constrained in a problematic way. Similarly,

[43] The preference either fails to meet conditions (2_b)-(4_b) or the agent fails to do (1_b) in virtue of failing to do (2_b)-(4_b).

[44] Stockholm syndrome might be bad for her in other respects (it may contribute to confusion or PTSD afterwards, for instance). Determining whether it is "all things considered" bad would require further details. But it is not inconceivable to imagine a case where Stockholm syndrome is good for the victim. Perhaps her feeling of affection for the kidnapper is the only thing that keeps her from committing suicide. If she ends up being rescued a few days after refraining from committing suicide, it seems plausible to suggest that having Stockholm syndrome was good for her. It kept her alive long enough for her to be rescued. Indeed, it is not uncommon for psychologists to treat it in this way. Alexander et al. regard Stockholm syndrome as "a means of coping and surviving since it certainly enables, on many occasions, hostages to deal with extreme and life-threatening circumstances" (2009, 18).

[45] See Enoch (2020).

we do not regard it as an autonomy deficit when an American consumer chooses between many varieties of apples but cannot easily acquire durian or rambutans. Adaptive preferences are worrisome only in cases where we judge that the agent came to have a preference precisely because she did not have any other valuable, live options.

In sum, this gives us four conditions for authentic ends. Authentic preferences are ones that are:

(1_c) Free of manipulation and coercion,
(2_c) consistent with the agent's motivational states,
(3_c) free from alienation, and
(4_c) not adaptive.

We are not suggesting that these categories are perfectly discrete. There is considerable overlap and interplay between them. For instance, a preference that results from manipulation or coercion (condition 1_c) is likely to be inconsistent with some of the agent's motivational states (condition 2_c) and it is also likely to spark feelings of alienation (condition 3_c). What is more, there are several connections between the conditions of capacity and the conditions of authenticity. One of the requirements of autonomous capacity is that the agent's reasoning must be free of influences that undermine her ability to deliberate (3_a). It would be helpful to recall Paul Benson's example of the husband who is gaslighting his wife, convincing her that her faculties are unreliable so that she defers to his judgment. In the previous section, this was characterized as a threat to her agential capacities. In the absence of self-trust, the wife is unable to successfully deliberate and choose for herself. But the wife's decision to defer to her husband also lacks authenticity. It was the product of manipulation. The husband's action has a twofold effect on her autonomy; it undermines both her capacity and her authenticity.

So we do not mean to suggest that there is no overlap between these eight conditions (four for capacity, four for authenticity). What is more, we do not want to claim that each of these picks out a necessary condition. This may be true of *some* of the conditions. When it comes to global autonomy, it is fairly clear that the baseline capacities outlined in the previous section are necessary conditions. That's why infants and small children lack autonomy. This means that some of these conditions are necessary, but this may not hold for all of them. It would be fair, however, to regard them as jointly sufficient conditions of autonomy.

If someone satisfies all of the conditions of authenticity and capacity, it is safe to say that she is autonomous. But what should we conclude when an agent or an agent's preference fails to satisfy one or more of these conditions? Perhaps the safest claim here would be that every violation of a condition counts as prima facie evidence that the agent's autonomy has been undermined. This means that the failure to satisfy one of these conditions is *evidence* of an autonomy violation but it is not necessarily *constitutive* of a violation.[46]

To sum up, throughout the book we will understand autonomy along the lines of what Kant meant by humanity. It is a rational agent's capacity to set and pursue her own ends. It involves a variety of conditions for both capacity and authenticity. It requires that the agent have the *ability* to set and pursue ends, and it requires that the ends be, in some sense, *her own*.

2.3 Why Humanity Matters Morally

Kant is probably the most well-known and ardent proponent of the idea that we are morally obligated to respect autonomy. Indeed there are few thinkers in the history of philosophy who put such a high premium on rational agency. Kant famously characterizes the moral law as an obligation to respect this capacity in others and in ourselves. This has come to be known as the categorical imperative's "formula of humanity." In the *Groundwork* he describes the moral imperative in this way:

Formula of Humanity: So act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means (*G* 4:429).

Kant argues that there is something distinctive about the value of rational agency that sets it apart from everything else in the world. Because human

[46] Rubel et al. make a similar move when it comes to authenticity. They write: "[A]uthenticity is important for two reasons. One is that authenticity is evidence of personal autonomy. The ability to reflect on one's values and preferences and to recognize them as compatible with one's sense of self and practical identity over time is an important test of the degree to which one's values and preferences are one's own, and (hence) that one self-governs. Likewise, lack of authenticity—that is, where a person would be alienated from their preferences and values upon reflection—is defeasible evidence that one's personal or social autonomy is compromised" (2021, 38).

beings are rational agents, we have the capacity to act morally, and it is in virtue of this capacity that we have what Kant calls "dignity"—a value that is elevated above any price.

Given that this book is a work of Kantian ethics, we could simply assume that Kant is right about the value of humanity and move on to the moral duties that follow from this principle. But that would be too quick. We can (and should) ask whether or not Kant is right. Why should we think that rational agency (humanity) is so important? What is distinctive about humanity that merits this elevation above everything else?

Our aim in this section is to present a compelling case for the value of humanity and to explain how Kant derives moral duties from this idea. We begin in the next section by stepping away from Kant to explore the common intuition that autonomy is morally important and that we have a duty to respect it. We then return to Kant and explain how he arrives at the formula of humanity. We conclude by showing how other moral theories (such as consequentialism and virtue ethics) should arrive at similar conclusions about the moral significance of autonomy. Although Kant believes that autonomy has intrinsic value, other theories might see it as instrumentally valuable. For classical utilitarians, such as Bentham, pleasure is the only thing that is intrinsically good, and autonomy is instrumentally good only insofar as it is a means to pleasure. We will revisit this idea in later chapters when we defend the existence of moral obligations to be autonomous with respect to technology and the attention economy. We believe that Kantian ethics does the best job of capturing what is morally worrisome about the attention economy: if autonomy is non-instrumentally morally valuable and the attention economy undermines autonomy, then the moral concern is quite clear. But, as we will argue in later chapters, other moral theories can reach the same endpoint even if they take a different route.

2.3.1 The Commonsense Intuition

Most readers are likely to share the intuition that there is something valuable about autonomy. Other things being equal, an autonomous life is better than one that lacks autonomy. This probably rings true in your own experience. You want to be the author of your own life story. When it comes to major life decisions, you would rather be in the driver's seat than ride along as a mere passenger. You would feel wronged if someone made those decisions without your consent or if they undermined your capacity

to make those choices for yourself. To see why this is compelling, we can contrast autonomous lives with similar ones that *lack* this element of self-direction.[47] It would be better for Sofia to choose her own career rather than to have her parents decide for her. Similarly, people should be free to choose their own life partners rather than be forced into arranged marriages.[48] This is also true in mundane situations. Imagine relinquishing all control over your diet and letting random strangers decide what you will eat for every meal. The value of such freedoms might be so obvious that they hardly need any defense.

Indeed, there are many situations in which the moral significance of autonomy is already taken as a given. There might be no domain in which the value of autonomy is more obvious than the field of medical ethics. Although "respect for autonomy" is only one of the four widely accepted principles of bioethics (the other three being beneficence, nonmaleficence, and justice), autonomy is often given pride of place.[49] Even those who regard these principles as "equal" sometimes say that autonomy should be seen as "first among equals" (Gillon 2003).[50] It would be hard to overstate the importance of respect for autonomy in medical contexts. When a patient (who has *decision-making capacity*, i.e., the ability to understand relevant information, appreciate one's situation, reason about options, and communicate choices)[51] refuses lifesaving treatment, physicians are required to respect the decision. This is what leads bioethicists to conclude that "autonomy is assumed to be more valuable than human life" (Brudney and Lantos 2011).

The importance of autonomy is also recognized in a variety of political contexts. For instance, within the tradition of liberalism, there is widespread opposition to paternalism. In spite of their many differences on other issues, Kant and Mill converge on this point. In *On Liberty*, Mill

[47] The following examples are similar to those provided by Enoch (2020). In arguing why consent matters, Enoch claims that the value of autonomy is something of a bedrock. To illustrate the intuitive plausibility of the value of autonomy, he gives a handful of examples to show why, ceteris paribus, autonomous lives are preferable to lives that lack autonomy.

[48] The ceteris paribus clause applies in each case, even if we avoid the redundancy of mentioning it in every instance. For example, it seems better to be in a happy, arranged marriage than it is to be in a freely chosen, abusive relationship.

[49] The four principles of bioethics were put forward by Beauchamp and Childress (2012).

[50] Gillon (2003) was the first to use this phrase, but the priority of autonomy is a popular idea in bioethics. See also Smith, "The Pre-Eminence of Autonomy in Bioethics." Cf. Pugh (2020).

[51] See Appelbaum and Grisso (1998).

articulates the famous harm principle, as he argues that the state should interfere with an individual's liberty only in order to prevent harm to others (1988, 9). But he thinks that we should not use the state's coercive power to prevent individuals from harming themselves. Similarly, Kant thinks we cannot rightly force other people to live according to our vision of what would make them happy. Interestingly, Mill's opposition to paternalism bears a striking resemblance to Kant's. As we will see in Chap. 6, Kant sees individual freedom as the only suitable justification of the state. The government exists to promote and safeguard the freedom of individuals.

In his "Theory and Practice" essay, Kant goes so far as to say that paternalistic government would amount to "the greatest despotism thinkable." He writes:

> No one can coerce me to be happy in his way (as he thinks of the welfare of other human beings); instead, each may seek his happiness in the way that seems good to him, provided he does not infringe upon that freedom of others to strive for a like end which can coexist with the freedom of everyone in accordance with a possible universal law (i.e., does not infringe upon this right of another). A government established on the principle of benevolence toward the people like that of a *father* toward his children—that is, a *paternalistic government* (*imperium paternale*), in which the subjects, like minor children who cannot distinguish between what is truly useful or harmful to them, are constrained to behave only passively, so as to wait only upon the judgment of the head of state as to how they *should be* happy and, as for his also willing their happiness, only upon his kindness—is the greatest *despotism* thinkable (a constitution that abrogates all the freedom of the subjects, who in that case have no rights at all). (TP 8:290–91)

In general, we tend to think that paternalistic interventions are justified only when someone lacks decision-making capacity. That is precisely why we do not allow small children to refuse medical treatment even though we must respect the decisions of capacitated adults who decline treatment. Similarly, we allow students to drop out of college, but education is compulsory for young children. Autonomy is the decisive factor in each of these situations.

And even when we put aside the state's coercive power, we tend to find something objectionable about the idea of being subject to the will of another person. This concept lies at the heart of contemporary republican views of freedom. Proponents of this view, such as Philip Pettit, have

argued that freedom consists in the absence of domination. To be free is to act in such a way that your decisions are not subject to the arbitrary will of another person (2014, 30).[52] Drawing on an intellectual tradition that goes all the way back to the Roman Republic, this view of freedom paints a stark contrast with the kind of domination that would have characterized the lives of slaves. Once again, Kant made valuable contributions to this way of thinking. In the margins of his copy of *Observations on the Feeling of the Beautiful and Sublime*, Kant says that there is "no misfortune more terrible to him who would be accustomed to freedom than to see himself delivered to a creature of his own kind, who could force him to do what he wants" (2005, 11).[53] When faced with a choice between slavery and risk of death, Kant thinks there would be no reservation about "preferring the latter" (Ibid.).

All of this goes to show that there is fairly widespread support for the idea that autonomy is morally significant. People generally share the intuition that we should respect autonomy (at least in the vast majority of situations). This is true in medicine, interpersonal contexts, and in the realm of politics. But this falls short of a philosophical justification. Even if this idea enjoys a great deal of intuitive support, we ought to ask what justifies it. What grounds the moral significance of autonomy? In the next section, we give a brief overview of Kant's reasons for thinking that autonomy matters morally. We explain why he picks humanity as the one and only "end in itself."

2.3.2 Kant on Humanity as an End in Itself

In order to understand why Kant identifies humanity as an end in itself, we must begin by explaining the philosophy of action that underpins Kant's moral philosophy. In the *Groundwork*, Kant puts forward the view that every intentional action should be analyzed in terms of a maxim, which he defines as "the subjective principle of volition" (*G* 4:401). Whenever you perform an intentional action, you are acting on the basis of a maxim. Whether you explicitly articulate it or not, your maxim can be

[52] Pettit specifies three specific conditions that must be met in order to realize the ideal of freedom as non-domination. You must have (1) "the room and resources to enact the option you prefer, (2) whatever your own preference over those options, and (3) whatever the preference of any other as to how you should choose" (2014, 30).

[53] Kant's note can be found in the *Notes and Fragments* collection (2005, 11).

characterized in terms of an act that you perform in some set of circumstances in order to achieve some end. For example, after seeing rain in the forecast, you pack your umbrella in order to avoid getting wet. Of course, you do some things without maxims (e.g., sneezing), but those are not intentional actions.

With this view of maxims in place, Kant gives his first formulation of the moral law, as he tells us that we should

Formula of Universal Law: act only in accordance with that maxim through which you can at the same time will that it become a universal law (*G* 4:421).

He then shows how to derive duties from this formula by explaining why we cannot will certain maxims to be universal laws (e.g., making a false promise to secure a loan, and refusing to ever help those in need). Immediately after this discussion of the formula of universal law, Kant starts to ask what this would look like for human beings, who always act for the sake of ends. Since he sees the moral law as valid for all rational beings, he acknowledges that it would also necessitate the actions of being with a "holy will" or "divine will" (*G* 4:414). But when it comes to human beings, the moral law takes the form of an "ought" precisely because it does not necessitate our actions. Our compliance with the moral law is not guaranteed.

He then distinguishes between two different forms of "oughts." First, there are hypothetical imperatives, which tell us that we ought to do something only because we desire something else. You ought to study for the LSAT if you want to go to law school. You ought to crack the eggs if you want to make an omelet. Kant wants us to notice the crucial limitation of hypothetical imperatives. They command us to do something only insofar as we desire the end in question. If you do not want to go to law school, you do not have to study for the LSAT. If you do not want an omelet, you do not have to crack the eggs. This is why he refers to the moral law as the categorical imperative. Unlike hypothetical imperatives, which depend on you willing some end, the moral law commands obedience from *all* rational agents. It does not matter what you want or who you are: the demands of the moral law are universal.

This is why Kant thinks that the moral law must be grounded in reason rather than anything empirical (e.g., human nature, or desire). Kant thinks we would never discover universal moral laws if we tried to ground them

in something like human nature. If we based morality on what people happen to want, we would end up with contingent laws that hold only insofar as someone has that particular desire. But it's important to recall that human beings always act for the sake of ends. This is why Kant follows his discussion of the universal law with questions about the possibility of an objectively valuable end. He wants to know if there is any end that has absolute worth for all rational beings. If there is, then it would be a suitable foundation for a universal law.

Kant answers his question in the affirmative by suggesting that humanity (rational nature) is the only thing that could be seen as an end in itself. He writes, "Now I say that the human being and in general every rational being *exists* as an end in itself, *not merely as a means* to be used by this or that will at its discretion; instead he must in all his actions, whether directed to himself or also to other rational beings, always be regarded *at the same time as an end*" (*G* 4:428). In order to understand why Kant thinks this, we must ask what is distinctive about rational agency. How does the value of humanity differ from the value of other things and why does Kant think this is the case?

The value of humanity is distinctive in three ways.[54] It is objective, unconditional, and non-fungible. An *objective value* is one that holds for all rational agents.[55] This is contrasted with relative value, which requires the presence of a contingent inclination. For instance, the value of a chocolate cake is merely relative, because it has value only insofar as agents desire it. Not everyone has this inclination (even if most of us do), so it cannot make a claim on the wills of all rational agents. This is what Kant means when he says that humanity is an "end in itself." It is something whose value does not depend on anything else (such as a desire). The value of humanity is also *unconditional*; it is not something that you could forfeit under any conditions. No matter what you do, you deserve respect as a human being. Finally, it is *non-fungible*, which is to say that human beings have the kind of value that does not admit of exchanges for things of equal value. Kant says in the *Groundwork* that "everything has either a price or a dignity. What has a price can be replaced by something else as its

[54] Our discussion of the formula of humanity is greatly indebted to Allen Wood's careful textual analysis of the *Groundwork* in *Kant's Ethical Thought* (1999, 114–38).

[55] Note, then, that our use of objective differs from that of contemporary philosophers, who commonly use "objective" to imply mind independent.

equivalent; what on the other hand is raised above all price and therefore admits of no equivalent has a dignity" (*G* 4:434–35).

It is also important to distinguish between ends that we are trying to bring about from ends that already exist. The omelet is an end that you are trying to bring about. It does not yet exist, but you would like to make one in order to satisfy your desire. It is an end that must be effected. The same goes for the law degree. Humanity, by contrast, is an "existent end" (*G* 4:437). The fact that humanity is valuable does not mean we are obligated to produce more of it. Thus, the value of humanity for Kant is not analogous to the value of pleasure for Mill. Kant is not arguing for something like Parfit's repugnant conclusion.[56] Instead, Kant is saying that we must recognize the value of human beings who already exist. As Allen Wood explains, "When I seek to relieve someone's suffering, my end to be effected is a certain state of affairs (the person's comfort); but I also act for the sake of a human being, whom I value in some way and is thus an existent end of my action" (1999, 116). Wood also shows how people sometimes express respect for an object, such as a flag, by taking off their hats. Those who wish to recognize the value of the flag do so by expressing respect for it, not by making more flags.

Those are the distinctive features of humanity's value. Rational agency is distinct from everything else because its value is objective, unconditional, and non-fungible. It is also to be seen as an existent end, rather than an end to be effected. This answers the first of the two questions. We now have a better understanding of what Kant thinks is special about humanity, but we have not answered the justificatory question. Why does Kant think humanity merits this elevation above everything else?

We can approach this conclusion, as Kant does, by process of elimination. He rules out other candidates for objective value in a widely discussed section of the *Groundwork* (G 4:428).

Kant begins by considering things that we value—the "objects of inclination." Some of us value chocolate cake, omelets, law degrees, or books about Kantian ethics. But we can immediately see why these are not

[56] The repugnant conclusion is a consequence of several approaches to population ethics (the study of ethical puzzles having to do with who is born and the size of future generations). It states, "For any possible population of at least ten billion people, all with a very high quality of life, there must be some much larger imaginable population whose existence, if other things are equal, would be better, even though its members have lives that are barely worth living" (Parfit 1984, 388). Kant's commitment to the value of humanity (as an existent end) does not tell us that we should produce as many humans as possible.

suitable as grounds for categorical value. First, not everyone values them. Second, these ends are valuable only insofar as someone desires them. Their value is instrumental rather than intrinsic. They are valued for the sake of something else (namely, the satisfaction of a desire).

This leads us to the next candidate. Since the objects of inclination do not have intrinsic value, perhaps it is the inclinations themselves. It is not the objects of desire that are good, it is desire-satisfaction that is good. As we saw earlier, however, we often repudiate our desires for one reason or another. The unwilling addict does not regard her desire for the addictive drug as a good thing at all.[57] On the contrary, she would rather be free of that inclination entirely. When your doctor tells you to avoid sugar in order to keep your insulin down, you no longer regard your desire for cake as something that is good.

For Kant, it is only by means of your rational agency that your inclinations become valuable. Because you are someone with the capacity to set and pursue your own ends, you evaluate your inclinations, and you decide which ones you should act on. It is for precisely this reason that you must regard your rational agency as the source of value. Reason confers value on your inclinations by making judgments about which ends matter to you. Korsgaard sees Kant's reasoning in this passage as a "regress upon the conditions" (1996, 120). We start by asking why the objects of inclination are valuable. They have value because of inclinations. Why do those inclinations have value? The answer is that rational agency has endowed them with value. When we ask why rational agency has value, we must realize

[57] This point will prove useful later on when we discuss smartphones and the attention economy. There might be skeptics who suggest that the compulsive use of smartphones merely shows that we value phones as we find ourselves eagerly scrolling for new content. The fact that something satisfies a desire (inclination) is not enough to show that it is good. As Korsgaard puts it, "Now even without fully endorsing what Kant says here, we can easily agree that there are some inclinations of which we want to be free: namely those whose existence is disruptive to our happiness. Take the case of a bad habit associated with a habitual craving—it would not be right to say that the object craved was good simply because of the existence of the craving when the craving itself is one that you would rather be rid of" (1996, 121).

that we have reached the end of the story. This is the "unconditioned condition." It is the bedrock on which all other value rests.[58]

Kant argues that we cannot possibly fail to recognize the value of our own rational agency. If we did, we would collapse into incoherence. Whenever we perform any action whatsoever, we demonstrate a commitment to the value of our rational agency. We cannot escape the fact that we act for the sake of reasons, and this very fact about our rational agency demonstrates the value of humanity as an end in itself. In this way, Kant's argument foreshadows Mill's point about how "questions of ultimate value do not admit of proof" and this is something that is "common to all first principles" (1957, 44).[59]

Indeed, in Kant's argument about the value of rational agency one can also hear echoes of Aristotle's defense of the principle of noncontradiction. It would be impossible to give a proof of the principle on noncontradiction, since any proof would presume the principle and thus beg the question. So Aristotle challenges his opponent to say anything at all. If the utterance picks out any object whatsoever, then it demonstrates a commitment to the principle. But what if our interlocutor refuses to speak? If so, then Aristotle thinks that his *actions* would betray a commitment to the principle. Even if he does not say anything, he must *do* something. And when he does, we could ask why he walks to the city of Megara instead of walking "into a well or over a precipice" (2001, *Metaphysics* 1008b16). By doing anything whatsoever, the skeptic shows that he does not *really* reject the principle.

The same can be said for Kant's defense of humanity as an end in itself. Rational agency is the only tool we have for deciding which things have value and which do not. What is more, we cannot possibly opt out of rational agency. We inevitably act for the sake of reasons. When we do so, we demonstrate our commitment to the value of humanity as an end in

[58] It is important here to stave off a common confusion. The claim here is *not* that humanity is valuable because it causes value to exist. That is, it is not like the claim that the espresso machine is *itself* valuable because it makes good espresso. Rather, as Wood (1999, 130) notes, a better analogy is authority: we can take an authority's advice as good in virtue of their giving only if we respect her authority. Her authority doesn't merely produce recommendations that are sensible (in the way that an espresso maker merely produces espresso, which is good—if it is good—in virtue of its own properties); instead, it (partially) *grounds* the fact that they are sensible.

[59] Allen Wood makes this point about Mill, but he does not draw the comparison with Aristotle (1999, 125).

itself. As we explained above, this means that its value is objective and unconditional. Everyone is compelled to recognize the value of rational agency. Unlike omelets and law degrees, we all acknowledge the value of rational agency. We do so—whether explicitly or implicitly—every time that we act.

But another skeptical challenge remains. What could be said to the egoist who argues that each individual recognizes *herself* as a source of value but fails to recognize the value of other rational beings? The Kantian response to the egoist is to point out the incoherence of the claim. You cannot possibly regard yourself as having a claim to objective value without recognizing the value of other beings whose rational agency is identical to your own. This is why the Golden Rule is effective even when it is used on small children. The egoistic child takes a toy from her friend who is smaller and weaker. So you ask her, "How would you like it if a bigger child took your toys?" Of course, the child should recognize that she would not like that at all. But she offers an egoistic rejoinder: "But that's not *me*. They should not take my toys, but it's fine for me to take their toys." The inconsistency is obvious. How can the child think that her desire or consent places limitations on others without recognizing that others can make equally legitimate claims on her? If Kant is right to suggest that the value of humanity is *objective,* then we must see that it holds for all agents, both as makers of claims and as recipients. We are both rulers and subjects in the kingdom of ends.

To summarize, Kant's argument goes like this. We begin by asking how anything could have value. Rational agency is the only answer. We acknowledge this fact (even if implicitly) every time that we act. That is, we regard ourselves as having objective value insofar as we are rational agents. On pain of inconsistency, we must recognize this in others as well: "But every other rational being also represents his existence in this way consequent on just the same rational ground that also holds for me" (*G* 4:429). This leads us to the conclusion that the value of humanity is objective and unconditional. It is the only end that could possibly make a valid claim for all rational wills, and this means that it can serve as the ground of the categorical imperative.

Once Kant reaches this conclusion, he immediately presents the formula of humanity: "So act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means" (*G* 4:430). But even if we grant Kant's point

about the value of humanity, we might wonder how that is supposed to lead to the formula of humanity.

Perhaps the best way to understand the formula of humanity is to think about how it differs from consequentialist or "teleological" moral theories. Kant's ethics is often characterized as "duty-based" or "deontological," and this is typically done to illustrate the contrast between Kantian ethics and consequentialist ethics. This contrast is commonly used to depict Kantian ethics in an uncharitable light. For example, consequentialists often argue that deontological ethics requires us to adopt a "fetishistic attitude toward rules against lying" while utterly ignoring the moral relevance of consequences, even when those consequences are disastrous (Wood 2007, 260). But this is an unfair caricature of Kant's view.[60] It would be a mistake to think that Kant dismisses the moral relevance of consequences, and it is important to underscore the importance of teleological thinking in Kant's ethics.[61] Indeed, this discussion of humanity should have made it clear that Kant is deeply sensitive to the idea that

[60] We will not try to resolve the infamous "killer at the door" objection here. A variety of Kant scholars have already done an excellent job of explaining why this thought experiment does not yield the conclusion that you should never say anything that is false, no matter the circumstances. See, for example, Wood (2007, 244–58), Korsgaard (1986), and Cholbi (2009). Kant's essay "On the Supposed Right to Lie from Philanthropy" addresses a specific case of lying as a legal declaration, and the point cannot be generalized to other instances of false statements. In Kant's lectures on ethics, he makes it clear that you can indeed lie to the robber who puts a knife to your neck and asks you where your money is. "I can also commit a *falsiloquium* when my intent is to hide my intentions from the other, and he can also presume that I shall do so, since his own purpose is to make a wrongful use of the truth. If an enemy, for example, takes me by the throat and demands to know where my money is kept, I can hide the information here, since he means to misuse the truth. That is still no *mendacium*" (VE 27:447). Lying is a technical and moralized term. Lying is analytically wrong in the same way that murder (defined as wrongful killing) is moralized. So Kant distinguishes between lies (which are automatically wrong since they violate a duty of right) and merely false utterances (which are not contrary to a duty of right).

[61] As Kristi Sweet puts it, "Kant also places an emphasis on ends in descriptions of moral goodness and, therefore, there is an undeniably 'teleological' aspect to his practical thought" (2013, 5). Allen Wood is even more emphatic on this point. He writes, "The Doctrine of Virtue is … overwhelmingly *teleological*. In fact, *within* the *system of ethical duties* (this is a crucial qualification), Kant is not (in the now commonly accepted sense of these terms) a 'deontologist' but a 'consequentialist.' For he asserts the priority of the 'good' (of the end to be effected) over the 'right'" (Wood 1999, 327). In a footnote, Wood goes on to explain the important qualification. In the Doctrine of Virtue, Kant's explication of morality involves "ends that are also duties," so the duties of virtue have a consequentialist flavor. They ultimately concern ends that we are morally obligated to promote (viz. our own perfection and the happiness of others). Wood points out how this style of reasoning is consequentialist even though the fundamental principle underlying them is not (1999, 414).

there is something of fundamental value (humanity, defined as the rational capacity to set ends), and that this existence of this value entails moral obligations for us.

So although Kant does not ignore the moral relevance of consequences and ends, his ethical system is importantly distinct from those that ground moral obligation in fundamentally teleological principles. For Kant, ethics cannot ultimately reduce to doing whatever brings about the best result. It was noted above that humanity's value must be recognized as an "existent end." This means that our moral duties do *not* involve bringing about such an end; they involve demonstrating the appropriate kind of respect for the humanity that already exists. As Korsgaard puts it, the escape from consequentialism ultimately involves abandoning the idea that "the business of morality is to *bring something about* … The primal scene of morality, I will argue, is not one in which I do something to you or you do something to me, but one in which we do something together. The subject matter of morality is not what we should bring about, but how we should relate to one another" (1996, 275).

When it comes to thinking about how we should relate to one another, Kant insists that we must treat each other with respect. If we recognize the value of humanity as something that is fundamentally greater than anything else, then we must express the appropriate attitude toward it when we act. It is the only thing whose value is objective, unconditional, and non-fungible; so, our actions should demonstrate respect for humanity.[62] Imagine that someone gives you their mother's ashes in a fragile vase. They tell you that this object has great importance to them, and they need you to hold onto it while they travel. If they were to return home and find you tossing it back and forth with a friend, they could rightly suggest that you failed to treat the object with the kind of respect it deserves. If you were to reply by pointing out that the vase did not break (so there were no harmful consequences), this would not be enough to show that your action was morally permissible.

[62] This approach to the formula of humanity has been called the "Respect-Expression" approach by Samuel Kerstein (2013). Kerstein argues against it on the grounds that it is inconsistent with common intuitions about actions that lead to loss of life such as euthanasia (2013, 29). Some advocates for the respect-expression approach, like David Velleman, have argued that respect for humanity should indeed give us pause in certain end-of-life situations (1999). Allen Wood (1999) also defends the claim that the formula of humanity is derived from the idea that we ought to express respect for the value of humanity. Oliver Sensen (2009) has defended the view on textual grounds.

Of course, in a case of that kind, we might think that the vase is not the thing that fundamentally matters. What matters in such a case is your friend and the fact that you betrayed their trust. Showing respect for the vase is the right thing to do only because you should respect your friend. To respect the humanity of another person is to recognize that their capacity to set and pursue ends is every bit as important as your own. You should not interfere with their rational capacities and you must refrain from doing things that would undermine their ability to set and pursue their own ends. This is why Kant thinks it is wrong to make a false promise to the banker. When you make a false promise to the banker about repaying a loan, your action expresses disrespect for the banker. If you recognize that rational agency is a faculty that deserves respect, then you cannot, on pain of inconsistency, disrespect the ends of other agents as if they did not matter.

The false promise to the banker involves treating someone as a mere means. This is a powerful way of explaining what is compelling about the formula of humanity. To treat someone as a mere means is to ignore or override the moral weight of their consent.[63] And such actions are almost always morally wrong; it is objectionable to *use* someone in such a way. Of course, this does not rule out every action that treats people as a means. We do that all the time in morally permissible ways. You treat your Uber driver as a means of transportation, the person behind the counter is a means of procuring certain goods, and so on. But in those cases, you do not treat people as *mere* means because you also respect their ends. The Uber driver and the store employee are being paid for their labor, so we cannot say that you have done something that is inconsistent with their consent.

[63] In *On What Matters*, Derek Parfit characterizes Kant's mere means principle in terms of consent. Parfit argues that there is something morally wrong with treating a person "in any way to which this person could not rationally consent" (2011, 181). Understanding the principle in terms of consent can be very helpful, but it is not the only approach. And this could be thought of in terms of either actual or possible consent. That distinction matters because we might want to accommodate intuitions where you perform CPR on someone who is unconscious and thus incapable of giving actual consent (but they would consent to it if they were awake). Treating a person as a mere means can also be understood as acting in such a way that the other person cannot possibly share your ends. This is useful for actions like false promises. The false promise would not possibly work if the other person were to know what you are doing.

But treating people as mere means is not the only way to fail to show respect for them. Kant thinks that you fail to respect someone as an end in themselves when you express contempt for them (*MS* 6:463). He thinks that such an action would be wrong even if it does not *use* the other person in any way. This means that there are multiple ways of failing to respect humanity. In some cases, the failure to respect humanity involves treating someone as a mere means; in other cases, it simply involves failing to show adequate respect for humanity as an end in itself. If we recognize the objective and unconditional value of humanity, then we are morally required to demonstrate respect for humanity when we act. If *this* is the ground of our moral obligations, as Kant thinks it is, then we can see how his ethical system differs from consequentialism. Our moral duties do not stem from requirements to bring about the best outcomes; they come from an obligation to express the appropriate kind of respect with our actions.

To bring this back to the issue at hand, we are now in a much better position to understand what the formula of humanity demands of us when it comes to respecting the autonomy of rational agents (including ourselves). If Kant is right about humanity, then there is something very special about our capacity to set and pursue ends. This capacity is something that we should cultivate and foster in ourselves. We ought to safeguard it against things that threaten to weaken our capacity to be in charge of our own lives. We must also respect humanity in others and refrain from doing things that would undermine their ability to set and pursue their own ends. To do so is to fail to show the kind of respect that all rational beings deserve in virtue of their rational agency. In the next chapter, we will argue that our relationship with technology threatens to undermine our autonomy. But before moving on, we would like to explore a handful of other moral theories in order to show why they are also likely to commit to the moral significance of autonomy.

2.3.3 The Centrality of Autonomy

Of course, Kant's ethical theory isn't the only one that values autonomy. Moral theories of all stripes have reasons to care about autonomy. But their reasons will differ. For starters, why should utilitarians value autonomy? Classical utilitarians are committed to *hedonism*—the claim that pleasure is the only thing that is intrinsically good, and pain is the only thing that is intrinsically bad. So it is not obvious why a utilitarian should

care about autonomy. Nevertheless, classical utilitarians like Mill and Bentham were indeed committed to the moral importance of autonomy. For Bentham, the value of autonomy is merely instrumental. When Bentham talks about liberty (or political sanctions) in his writings, he makes it very clear that pain and pleasure are the only real values at stake (1907, ch. 3). He thinks that it is often the case that someone feels pain of one kind or another when they suffer a loss of liberty.

Given what Mill says in *Utilitarianism*, it would be reasonable to expect him to concur with Bentham on this point. He makes it clear that happiness should be understood as nothing more than "pleasure and the absence of pain" (1957, 10). He then goes on to explain why things like fame, money, and power are good only insofar as they promote happiness. Given his commitment to hedonism in *Utilitarianism*, it is somewhat surprising to see what he says about autonomy in *On Liberty*. First, he clarifies what he has in mind by autonomy (though he does not use that word):

> A person whose desires and impulses are his own—are the expression of his own nature, as it has been developed and modified by his own culture—is said to have a character. One whose desires and impulses are not his own has no character, no more than a steam engine has a character. (1988, 57)

When Mill reflects on why the capacity to make our own choices is morally significant, we might expect him to cash it out in terms of pleasure or happiness. But he seems to suggest otherwise when he writes, "If a person possesses any tolerable amount of common sense and experience, his own mode of laying out his existence is the best, not because it is the best in itself, but *because it is his own*" (1988, 64, emphasis added). Here, Mill sounds like he is claiming that autonomy has intrinsic value. Being in charge of your own life matters not merely because you will make the best choices (the ones that maximize happiness) but because they are your own.[64]

[64] James Griffin takes this even further by suggesting the following thought experiment. What would you say if a benevolent despot offered to take control of your life, and they convinced you that they would do a better job of promoting your own happiness? Griffin says that he would decline the offer as he prefers to be his own master (1986, 9). Jonathan Pugh (2020) discusses the personal despot argument at length as he defends the claim that autonomy has intrinsic or final value.

Parfit is credited with giving the name to "objective list" theories of well-being. And Parfit includes "rational activity" and "the development of one's abilities" on his list of objective goods while putting "being manipulated" on his list of things that are objectively bad (1984, 499).

This means that consequentialists have two ways of committing to the moral importance of autonomy. It could be that autonomy matters instrumentally insofar as it promotes pleasure, or autonomy could have final value. Of course, consequentialists who commit to the final value of autonomy are not required to hold the Kantian view that autonomy is the *only* thing that has this kind of value. Some consequentialists, such as G.E. Moore, have pluralistic theories of the good that include things like beauty, knowledge, and friendship. So consequentialists could simply include autonomy as one of the many goods that we are morally obligated to bring about through our actions.

As for eudaimonist virtue ethics, perfectionist theories, and ethical egoism, the accounts are going to be fairly similar. Like utilitarianism, these theories are, at bottom, teleological. They begin with an account of the good, and ethical prescriptions are then made on the basis of promoting that good. Aristotelian virtue ethics, for instance, starts off with an account of human flourishing (eudaimonia). For Aristotle, happiness does not merely consist in pleasure and the absence of pain; instead, he believes that happiness requires the possession of certain objective goods. For most contemporary Aristotelians, autonomy is likely to be on their list of objective goods.[65] The same is true for Aristotle himself who regarded the exercise of one's own practical reason as partially constitutive of happiness. Perfectionist theories are quite similar to virtue ethics. Perfectionists ground moral and political prescriptions in an account of what is required for the proper development of human nature. Once again, both historical and contemporary perfectionists tend to value autonomy on the grounds that the exercise of autonomy is a fundamental component of human nature and happiness.[66] The same goes for ethical egoists, who typically count autonomy among the goods that rationally self-interested individuals ought to secure for themselves.

Finally, Scanlon's contractualism also places a great deal of weight on rational agency. Indeed, the core principle of contractualism is grounded

[65] For instance, Martha Nussbaum places "practical reason" on her list of ten basic capabilities. She defines practical reason as "Being able to form a conception of the good and to engage in critical reflection about the planning of one's life" (2006, 77).

[66] Spinoza is a good example of a perfectionist (and ethical egoist) who values autonomous agency. For Spinoza, our most fundamental task as human beings is to free ourselves from the bondage of the passions and to exercise our freedom through rational agency. Contemporary perfectionists like Thomas Hurka defend the value of autonomy as well (Hurka 1987).

in the equal moral status of persons in virtue of their capacity for rational agency. Scanlon writes,

> [R]especting the value of human (rational) life requires us to treat rational creatures only in ways that would be allowed by principles that they could not reasonably reject insofar as they, too, were seeking principles of mutual governance which other rational creatures could not reasonably reject. This responds to the problem of selecting among reasons in a way that recognizes our distinctive capacities as reason-assessing, self-governing creatures. (1998, 106)

Respect for rational agency is a core commitment for contractualists. It is in virtue of this respect that contractualism obligates us to act according to principles that no one could reasonably reject. As Ashford and Mulgan explain, "According to contractualism, morality consists in what would result if we were to make binding agreements from a point of view that respects our equal moral importance as rational autonomous agents" (Ashford and Mulgan 2018).

Of course, this list of moral theories is not exhaustive, but it should be sufficient to demonstrate that there are many paths to the same conclusion. When it comes to recognizing the moral importance of autonomy, we might be, as Parfit suggests, "climbing the same mountain on different sides" (2011, 385). A wide variety of ethical theories share this commitment to the value of autonomy. The capacity to be in charge of one's own life is something that matters morally. There are very good reasons for believing, just as Kant did, that we ought to respect the capacity of rational agents to set and pursue their own ends.

2.4 Conclusion

Our aim in this chapter was to answer two questions. What is autonomy and why does it matter morally? Throughout the book, we take autonomy to be synonymous with what Kant referred to as "humanity"—the capacity to set and pursue one's own ends. As we explained in Sect. 2.2, humanity comprises both a set of capacities and a set of conditions for authenticity. Exercising your humanity requires you to have certain capacities to set and pursue ends, and it requires that the ends be, in some important sense, your own. In each case, we identified four conditions. In order to have the capacities required for autonomy, you must have the following:

(1_a) baseline competence (the ability to form intentions, make coherent judgments, etc.),
(2_a) a lack of external constraints,
(3_a) the absence of cognitive inhibitions (such as self-doubt), and
(4_a) a sufficiently wide range of options (with the ability to pursue them effectively).

And, in order for your ends to be considered your own, they must be:

(1_c) free of manipulation and coercion,
(2_c) consistent with your motivational states,
(3_c) free from alienation, and
(4_c) not adaptive.

When it comes to understanding why autonomy matters morally, we offered a wide range of reasons derived from a variety of moral theories. Of course, our leanings are Kantian; we believe that autonomy has intrinsic or final value. Kant argues that we are morally required to respect this capacity both in ourselves and in others. This is expressed very clearly in the humanity formula of the categorical imperative. We then showed why other moral theories, such as consequentialism, virtue ethics, and contractualism are also committed to the claim that autonomy is morally significant. So it is with this conclusion in mind that we turn to the topic of the next chapter. If autonomy matters morally, as we believe it does, then we ought to be concerned about our relationship with technology, especially with devices like smartphones.

References

Alexander, David A., and Susan Klein. 2009. Kidnapping and hostage-taking: A review of effects, coping and resilience. *Journal of the Royal Society of Medicine* 102: 16–21.

Alm, Kristian, and Mark Brown. 2021. John Rawls' concept of the reasonable: A study of stakeholder action and reaction between British petroleum and the victims of the oil spill in the Gulf of Mexico. *Journal of Business Ethics* 172: 621–637.

Appelbaum, Paul, and Thomas Grisso. 1998. Assessing patients' capacities to consent to treatment. *New England Journal of Medicine* 319: 1635–1638.

Aristotle. 2001. *The basic works of Aristotle*. New York: Random House.

Arrington, Robert. 1982. Advertising and behavior control. *Journal of Business Ethics* 1: 3–12.

Ashford, Elizabeth, and Tim Mulgan. 2018. Contractualism. In *The Stanford encyclopedia of philosophy*, ed. Edward N. Zalta. https://plato.stanford.edu/entries/contractualism/. Accessed 9 May 2023.

Aylsworth, Timothy. 2020. Autonomy and manipulation: Refining the argument against persuasive advertising. *Journal of Business Ethics* 175: 689–699.

Barnhill, Anne. 2014. What is manipulation? In *Manipulation: Theory and practice*, ed. Christian Coons and Michael Weber, 51–72. New York: Oxford University Press.

Baron, Marcia. 2003. Manipulativeness. *Proceedings and Addresses of the American Philosophical Association* 77: 37–54.

Beauchamp, Tom L., and James F. Childress. 2012. *Principles of biomedical ethics*. 7th ed. New York: Oxford University Press.

Benson, Paul. 1987. Freedom and value. *Journal of Philosophy* 84: 465–486.

———. 1994. Free agency and self-worth. *Journal of Philosophy* 91 (12): 650–668.

Bentham, Jeremy. 1907. *An introduction to the principles of morals and legislation*. New York: Clarendon Press.

Bratman, Michael. 2003. Autonomy and hierarchy. *Social Philosophy and Policy* 20: 156–176.

———. 2007. *Structures of agency*. New York: Oxford University Press.

———. 2018. *Planning, time, and self-governance*. New York: Oxford University Press.

Brison, S.J. 2000. Relational autonomy and freedom of expression. In *Relational autonomy: Feminist perspectives on autonomy, agency and the social self*, ed. Catriona Mackenzie and Natalie Stoljar, 280–300. New York: Oxford University Press.

Brudney, Daniel, and John Lantos. 2011. Agency and authenticity: Which value grounds patient choice? *Theoretical Medicine and Bioethics* 32: 217–227.

Burr, Christopher, Nello Cristianini, and James Ladyman. 2018. An analysis of the interaction between intelligent software agents and human users. *Minds and Machines* 28: 735–774.

Buss, Sarah, and Andrea Westlund. 2018. Personal autonomy. In *The Stanford encyclopedia of philosophy*, ed. Edward N. Zalta. https://plato.stanford.edu/archives/spr2018/entries/personal-autonomy/. Accessed 3 May 2023.

Cholbi, Michael. 2009. The murderer at the door: What Kant should have said. *Philosophy and Phenomenological Research* 79: 17–46.

Christman, John. 1991. Autonomy and personal history. *Canadian Journal of Philosophy* 21: 1–24.

———. 2011. *The politics of persons: Individual autonomy and socio-historical selves*. Reissue ed. Cambridge: Cambridge University Press.

Crisp, Roger. 1987. Persuasive advertising, autonomy, and the creation of desire. *Journal of Business Ethics* 6: 413–418.
Cunningham, Anne. 2003. Autonomous consumption: Buying into the ideology of capitalism. *Journal of Business Ethics* 48: 229–236.
Dworkin, Gerald. 1976. Autonomy and behavior control. *The Hastings Center Report* 6: 23.
———. 1981. The concept of autonomy. In *Science and ethics*, ed. R. Haller. Amsterdam: Rodopi.
———. 1988. *The theory and practice of autonomy*. New York: Cambridge University Press.
Elster, Jon. 1983. Sour grapes. In *Sour grapes: Studies in the subversion of rationality*, 110–141. Cambridge: Cambridge University Press.
Enoch, David. 2013. The disorder of public reason. *Ethics* 124: 141–176.
———. 2015. Against public reason. In *Oxford studies in political philosophy*, ed. David Sobel, Peter Vallentyne, and Steven Wall, vol. 1, 112–142. Oxford: Oxford University Press.
———. 2020. False consciousness for liberals, part I: Consent, autonomy, and adaptive preferences. *The Philosophical Review* 129: 159–210.
Eyal, Nir. 2012. The art of manipulation. *Forbes*. https://www.forbes.com/sites/nireyal/2012/07/02/the-art-of-manipulation/?sh=2a0d9f955009. Accessed 24 July 2023.
———. 2014. *Hooked: How to build habit-forming products*. New York: Penguin.
Feinberg, Joel. 1989. Autonomy. In *The inner citadel: Essays on individual autonomy*, ed. John Christman. New York: Oxford University Press.
Formosa, Paul. 2013. Kant's conception of personal autonomy. *Journal of Social Philosophy* 44: 193–212.
Frankfurt, Harry. 1971. Freedom of the will and the concept of a person. *Journal of Philosophy* 68: 5–20.
Frierson, Patrick. Forthcoming. Kant on education. In *The Oxford handbook of Kant*, ed. A. Gomes and A. Stephenson. Oxford: Oxford University Press.
Gillon, Raanan. 2003. Ethics needs principles—Four can encompass the rest—And respect for autonomy should be 'first among equals'. *Journal of Medical Ethics* 29: 307–312.
Govier, Trudy. 2003. Self-trust, autonomy, and self-esteem. *Hypatia* 8: 99–120.
Griffin, James. 1986. *Well-being: Its meaning, measurement and moral importance*, 1986. Oxford: Clarendon Press.
Hurka, Thomas. 1987. Why value autonomy? *Social Theory and Practice* 13: 361–382.
Kant, Immanuel. 1997. *Lectures on ethics*. Trans. Peter Heath, ed. Peter Heath and J. B. Schneewind. New York: Cambridge University Press.
———. 2005. *Notes and fragments*, ed. Paul Guyer. Trans. Paul Guyer, Curtis Bowman, and Frederick Rauscher. Cambridge: Cambridge University Press.

———. 2007a. *Anthropology, history, and education*, ed. Günter Zöller and Robert B. Louden. Cambridge: Cambridge University Press.
———. 2007b. *Critique of the power of judgment*. Trans. Paul Guyer and Eric Matthews. New York: Cambridge University Press.
———. 2008a. Groundwork of the metaphysics of morals. In *Practical philosophy*. Trans. Mary J. Gregor. New York: Cambridge University Press.
———. 2008b. The metaphysics of morals. In *Practical philosophy*. Trans. Mary J. Gregor. New York: Cambridge University Press.
———. 2008c. On the common saying: That may be correct in theory, but it is of no use in practice. In *Practical philosophy*. Trans. Mary J. Gregor. New York: Cambridge University Press.
———. 2008d. *Practical philosophy*. Trans. Mary Gregor, ed. Mary Gregor. New York: Cambridge University Press.
Karremans, Johan C., Wolfgang Stroebe, and Jasper Claus. 2006. Beyond Vicary's fantasies: The impact of subliminal priming and brand choice. *Journal of Experimental Social Psychology* 42: 792–798.
Keeling, Geoff, and Christopher Burr. 2022. Digital manipulation and mental integrity. In *The philosophy of online manipulation*, ed. Fleur Jongepier and Michael Klenk. New York: Routledge.
Kerstein, Samuel J. 2013. *How to treat persons*. Oxford: Oxford University Press.
Khader, Serene J. 2011. *Adaptive preferences and women's empowerment*. New York: Oxford University Press.
Kleingeld, Pauline. 2019. On dealing with Kant's sexism and racism. *SGIR Review* 2: 3–22.
Klenk, Michael. 2022. (Online) Manipulation: Sometimes hidden, always careless. *Review of Social Economy* 80: 85–105.
Knutzen, Jonathan. 2020. The trouble with formal views of autonomy. *Journal of Ethics and Social Philosophy* 18: 173–210.
Korsgaard, Christine. 1986. The right to lie: Kant on dealing with evil. *Philosophy & Public Affairs* 15: 325–349.
———. 1996. *Creating the kingdom of ends*. New York: Cambridge University Press.
Lippke, Richard. 1989. Advertising and the social conditions of autonomy. *Business and Professional Ethics Journal* 8: 35–58.
Mackenzie, Catriona. 2022. Relational equality and the debate between externalist and internalist theories of relational autonomy. In *Autonomy and equality: Relational approaches*, 32–56. New York: Routledge.
Mackenzie, Catriona, and Natalie Stoljar. 2000. *Relational autonomy: Feminist perspectives on autonomy, agency and the social self*. New York: Oxford University Press.
McLeod, Carolyn. 2002. *Self-trust and reproductive autonomy*. Cambridge: MIT Press.
Mill, John Stuart. 1957. *Utilitarianism*. Upper Saddle River: Prentice-Hall.

———. 1988. *On liberty*. Indianapolis: Hackett Publishing Company.
Noggle, Robert. 1995. Autonomy, value, and conditioned desire. *American Philosophical Quarterly* 32: 57–69.
Nussbaum, Martha. 1993. Non-relative virtues: An Aristotelian approach. In *The quality of life*, ed. Martha Nussbaum and Amartya Sen, 242–269. Oxford: Oxford University Press.
———. 2001a. Symposium on Amartya Sen's philosophy 5: Adaptive preferences and women's options. *Economics and Philosophy* 17: 67–88.
———. 2001b. *Women and human development: The capabilities approach*. Cambridge: Cambridge University Press.
———. 2006. *Frontiers of justice: Disability, nationality, species membership*. Cambridge: Harvard University Press.
———. 2011. *Creating capabilities*. Cambridge: Harvard University Press.
Oshana, Marina. 2006. *Personal autonomy in society*. Aldershot: Ashgate Publishing.
Papanek, Hannah. 1990. To each less than she needs, from each more than she can do: Allocations, entitlement, and value. In *Persistent inequalities: Women and world development*, ed. I. Tinker, 162–185. Oxford: Oxford University Press.
Parfit, Derek. 1984. *Reasons and persons*. Oxford: Oxford University Press.
———. 2011. *On what matters*. Oxford: Oxford University Press.
Pettit, Philip. 2014. *Just freedom: A moral compass for a complex world*. New York: W. W. Norton & Company.
Pugh, Jonathan. 2020. *Autonomy, rationality, and contemporary bioethics*. New York: Oxford University Press.
Ramachandran, Nira. 2006. *Women and food security in South Asia: Current issues and emerging concerns*. Helsinki: UNU Institute for Development Economics Research. https://www.econstor.eu/bitstream/10419/63421/1/52139452X.pdf. Accessed 18 July 2023.
Rawls, John. 2005. *Political liberalism*. Expanded ed. New York: Columbia University Press.
Raz, Joseph. 1986. *The morality of freedom*. New York: Oxford University Press.
———. 1990. Facing diversity: The case of epistemic abstinence. *Philosophy and Public Affairs* 19: 3–46.
Rogers, Stuart. 2001. How a publicity blitz created the myth of subliminal advertising. *Public Relations Quarterly* 37: 12–17.
Rubel, Alan, Clinton Castro, and Adam Pham. 2021. *Algorithms and autonomy: The ethics of automated decision systems*. Cambridge: Cambridge University Press.
Scanlon, T.M. 1998. *What we owe to each other*. Cambridge: Harvard University Press.
Schapiro, Tamar. 1999. What is a child? *Ethics* 109: 715–738.
———. 2003. Childhood and personhood. *Arizona Law Review* 45: 575–594.
Schwartz, Barry. 2009. *The paradox of choice: Why more is less*. New York: Harper Collins.

Sen, Amartya. 1990. Gender and cooperative conflicts. In *Persistent inequalities: Women and world development*, ed. Irene Tinker, 123–149. Oxford: Oxford University Press.

———. 1999. *Development as freedom*. New York: Knopf.

Sensen, Oliver. 2009. Dignity and the formula of humanity. In *Kant's groundwork of the metaphysics of morals: A critical guide*, ed. Jens Timmermann. Cambridge University Press.

Smith, Janet E. 1997. The pre-eminence of autonomy in bioethics. In *Human lives: Critical essays on consequentialist bioethics*, ed. David S. Oderberg and Jacqueline A. Laing, 182–195. London: Palgrave Macmillan UK.

Stoljar, Natalie. 2000. Autonomy and the feminist intuition. In *Relational autonomy: Feminist perspectives on autonomy, agency and the social self*, ed. Catriona Mackenzie and Natalie Stoljar, 94–111. New York: Oxford University Press.

———. 2014. Autonomy and adaptive preference formation. In *Autonomy, oppression, and gender*, ed. Andrea Veltman and Mark Piper. Oxford: Oxford University Press.

———. 2018. Feminist perspectives on autonomy. In *The Stanford encyclopedia of philosophy*, ed. Edward N. Zalta. https://plato.stanford.edu/entries/feminism-autonomy/. Accessed 8 May 2023.

Stoljar, Natalie, and Kristin Voigt. 2022. *Autonomy and equality: Relational approaches*. New York: Routledge.

Susser, Daniel, Beate Roessler, and Helen Nissenbaum. 2019a. Online manipulation: Hidden influences in a digital world. *Georgetown Law Technology Review* 4: 1–45.

———. 2019b. Technology, autonomy, and manipulation. *Internet Policy Review* 8: 1–22.

Sweet, Kristi. 2013. *Kant on practical life*. Cambridge: Cambridge University Press.

Terlazzo, Rosa. 2016. Conceptualizing adaptive preferences respectfully: An indirectly substantive account. *Journal of Political Philosophy* 24: 206–226.

Velleman, J. David. 1999. A right of self-termination? *Ethics* 109: 606–628.

Watson, Gary. 1975. Free agency. *The Journal of Philosophy* 72: 205–220.

Wood, Allen. 1999. *Kant's ethical thought*. New York: Cambridge University Press.

———. 2007. *Kantian ethics*. New York: Cambridge University Press.

———. 2014. Coercion, manipulation, exploitation. In *Manipulation: Theory and practice*, ed. Christian Coons and Michael Weber, 17–50. New York: Oxford University Press.

Wyrębska-Đermanović, Ewa. 2021. Kantian moral education for the future of humanity: The climate change challenge. *Journal of Philosophy of Education* 55: 1045–1056.

CHAPTER 3

Mobile Devices and Autonomy: Individual-Level Effects

For getting other human beings' inclinations into one's power, so that one can direct and determine them according to one's intentions, is almost the same as possessing others as mere tools of one's will.
—Kant, *Anth.* 7:271

The thought process that went into building these applications … was all about: "How do we consume as much of your time and conscious attention as possible?" And that means that we need to sort of give you a little dopamine hit every once in a while, because someone liked or commented on a photo or a post or whatever. And that's going to get you to contribute more content, and that's going to get you … more likes and comments. It's a social-validation feedback loop … exactly the kind of thing that a hacker like myself would come up with, because you're exploiting a vulnerability in human psychology.
—Sean Parker, first president of Facebook (This quote from Parker comes from an interview with *The Guardian*. See Solon (*2017a*)).

3.1 Introduction

In this chapter, we lay important groundwork for our coming arguments. We clarify which technologies we are chiefly concerned with, as well as the empirical grounds for thinking that they conflict with our

T. Aylsworth, C. Castro, *Kantian Ethics and the Attention Economy*,
https://doi.org/10.1007/978-3-031-45638-1_3

(individual-level) autonomy.[1] This, we will argue, gives rise to a duty we have to ourselves to be "digital minimalists," that is, to be mindful about our interactions with digital technologies such that they do not conflict with our capacity to set and pursue our ends. But we will not complete that argument—or flesh out the concept of digital minimalism—until the next chapter. This chapter is a clearinghouse for the empirical evidence for the argument for digital minimalism, which also serves as a foundation that we will build upon to justify the empirical premises needed for the arguments for attention ecology (a duty to assist others in being digital minimalists) and group-level digital minimalism (the duty *we* have, as a collective, to engage in digital minimalism).

Throughout the chapter, it will be useful to draw on Esther's example from the introduction. Esther joked about how difficult it is for her to focus on reading a book as she finds herself compulsively unlocking her phone to look at Instagram. Of course, we do not think that Esther's experience is exhaustive; the concerns about problematic smartphone use go well beyond issues with literacy.[2] But it will be a useful lens for exploring some of the concerns that arise in the empirical literature.

3.2 What Are the Devices?

Our main concern in this book is our relationship with what we will call "mobile devices," in the context of the attention economy and a culture of constant exposure to mobile devices.

We will clarify the term and explain why mobile devices are our main concern, but first note that many devices, namely desktop computers, do not qualify as mobile devices. This does not mean that our concerns do not extend to these non-mobile devices. We are *also* concerned with non-mobile, web-connected screens in the context of the attention economy; but this is not our main concern. Now, for clarifications.

By "mobile device" we do not simply mean devices that are portable. If we did, then mechanical watches would qualify. But we are not concerned with people's relationships with mechanical watches. Instead, we mean handheld computers that have access to the internet, such as smartphones,

[1] For the next few chapters, we will be concerned solely with individual-level autonomy. In Chap. 7, we discuss the idea of acting autonomously as a group.

[2] Nevertheless, as we explain below, literacy is among the set of concerns. See Carr (2010) and Wolf (2018).

tablets, and even laptops (though, as we explain in what follows, smartphones are of utmost concern). These devices allow you to be in constant contact with the attention economy, even when you are outside the home and while you are in bed.

Let us now explain why we think the context—the attention economy and a culture of constant contact with it—is important.

We'll begin by explaining what we take the attention economy to be and why we do not take it on its own (or at least without further contextualization) to be an object of concern. Following Castro and Pham (2020), we take *the attention economy* to be the economic market where consumers give media developers their attention in exchange for a service (e.g., a news feed), and where developers sell consumer attention to advertisers.

We do not take the attention economy *alone* to raise the sorts of concerns that animate us here. As Castro and Pham (2020) note, the sorts of exchanges that constitute the attention economy at least go back to the 1830s, when *The New York Sun* transitioned away from subscriptions to advertising as its main source of revenue.[3] To go from the existence of the attention economy to our concerns about autonomy, further elements are needed. In our case those elements are behavior modification devices that can habituate us to having our attention stolen at any moment (i.e., mobile devices), and the normalization of the constant use of these devices. We would now like to say a bit more about these other elements.

When reflecting on the relationships we have with our mobile devices—especially in the context of the attention economy—it is important to consider two important facts about them. First, they are designed to get our attention. They can ring, vibrate, light up, and display notifications. This is for good reason: we do not want to miss important calls, texts, or reminders. Second, these features can be used to instill habits in us. Some of these habitual behaviors appear to be compulsive (and to be products of manipulation).[4]

Consider Nir Eyal's bestselling design guide *Hooked: How to Build Habit-Forming Products*, which opens by boasting about how much we use our phones:

[3] Castro and Pham cite Postman (1993) and Wu (2016) on this issue.

[4] In the next Sect. (3.3) we will explore many ways of understanding the autonomy deficits that result from our excessive use of mobile devices.

> Seventy-nine percent of smartphone owners check their device within fifteen minutes of waking up … one-third of Americans say they would rather give up sex than lose their cell phones. … people check their phones … an astounding 150 [times a day]. Face it: We're hooked. (Eyal 2014, 1)

And what is it, according to Eyal, that keeps us hooked? It's the power of these devices to create and foster *habits*, which he identifies as "'automatic behaviors triggered by situational cues': things we do with little or no conscious thought" (Ibid.).

Eyal sees behavior modification via the use of mobile devices as an important opportunity for companies:

> Forming habits is imperative for the survival of many products. As infinite distractions compete for our attention, companies are learning to master novel tactics to stay relevant in users' minds. (Ibid. 2)

He then goes on to introduce the centerpiece of his book, the Hooked Model: "a four-phase process … to form habits" (Ibid. 8).

The Hooked Model is simple but powerful. As he says, it's what drives us to touch our phone first thing in the morning and then over a hundred times after that; what makes a third of us prefer our phones to sex; what lies behind, "the pull to visit YouTube, Facebook, or Twitter" that keeps you "tapping and scrolling an hour later" even if you just wanted to use one of those services "just a few minutes" (Ibid. 1).

The four phases of the Hooked Model run as follows.

First, *trigger* the user. The idea here is to cue the (potential) user to perform an action ("action" being the next step). This can be done via a notification (e.g., "Someone has visited your profile") or prompt ("To Continue Reading, Register for Free"). As Eyal notes, triggers can be both external (e.g., notifications, prompts) or internal, which happen when "a product becomes tightly coupled with a thought, and emotion, or a pre-existing routine" (Ibid. 47). One of the goals is to use the Hooked Model so effectively that users develop internal triggers, "*the brass ring of habit-forming technology*" (Ibid. 48).

Second, if the first phase went as planned, the user will perform an *action*, or "behavior done in anticipation of a reward" (Ibid. 7). Following up on the examples above, this could include checking your profile or creating an account. The keys to triggers that manifest actions, Eyal notes, are two factors as they relate to action: the underlying motivation to

perform the action and the ease of performing it. The suggestion is to increase motivation and ease. He notes, "To initiate action, doing must be easier than thinking" (Ibid. 61).

Third, the action is to be associated with a *variable reward*, that is, the action is to be associated with an unpredictable possibility of reward:

> Research shows that levels of the neurotransmitter dopamine surge when the brain is expecting a reward … which suppresses the areas of the brain associated with judgment and reason while activating the parts associated with wanting and desire. (Ibid. 9)

Eyal goes on to cite slot machines and lotteries as classic examples of habit-forming products that effectively use variable rewards. He then goes on to give a quick fictionalized example of a Pinterest user, who, in the presence of interesting and uninteresting photos, has her brain dopamine system set "aflutter with the promise of reward"; before she knows it, she has "spent forty-five minutes scrolling" (Ibid. 11).

The fourth, and final, phase of the Hooked Model involves the user making an *investment*, that is, having them put something of value into the product, something that takes time or effort. The idea here is to prime the user so that they are more susceptible to the next trigger sent their way.

As Eyal notes, citing work by Dan Ariely, Michael Norton, and Daniel Mochon, "we irrationally value our efforts" (Ibid. 136). He goes on to say that this, along with a drive to seek consistency in our behaviors and an aversion to cognitive dissonance, can be leveraged to bring users back. This means, Eyal states, that investments can get users to rationalize their use of the product, speeding up the formation of habits: "Rationalization helps us give reasons for our behaviors, even when those reasons might have been designed by others" (Ibid. 141).

An important element of all of this is the fact that it is normal to have the device on your person at all times. Unlike its predecessors—such as the desktop computer (which can run the Hooked cycles, but doesn't fit in your pocket) or print newspapers (whose content could be subsidized through advertising but couldn't run Hooked cycles)—a mobile device can go anywhere we might want to go, making us constantly available for the operant conditioning that Eyal promotes in his book. And as Fabio et al. (2022) note in their study on smartphone use and behavioral and cognitive self-control deficits, "the fact that a smartphone is small, easily handled and portable makes the risks more insidious and pervasive" (2).

Why, then, are we not worried about desktop computers? Again, we are, at least in some cases (such as the problems related to social media, which we discuss in Chap. 7), but they are not our primary concern. People could have damaging relationships with their desktops, absent mobile devices and the attention economy (e.g., gaming or gambling addictions). However, we do not see this as central to the widespread concerns we have about autonomy, as the constant connection, feedback, and monitoring made possible by mobile devices are what create the conditions for the scaled attention economy that worry us.

3.3 How Mobile Devices Undermine Autonomy

We will now discuss the individual-level effects of mobile devices, in terms of the two facets of autonomy discussed in Chap. 2 (viz. capacity and authenticity). We should note that these do not constitute the only empirically oriented concerns of the book; they are concerns that *primarily* play a role in our first major argument of three: our argument for the (individual-level) duty to be digital minimalists. We discuss group-level issues (such as polarization and democracy) in Chap. 7.

We should also note that we have some humility about the results discussed below (and throughout the book). What is discussed below is a snapshot as of this writing of our understanding of the effects of mobile devices on users. While we are confident in our analysis, we should note that studying these effects is complex and often muddled by, for example, the fact that it is difficult (if not impossible) in many cases to find a control group for the study. Furthermore, this is a burgeoning area of psychological and neuroscientific research, so the data are fairly new, and research is ongoing. Nevertheless, we will present the scientific findings that are currently available, and we cite large-scale meta-analyses whenever possible.

3.3.1 Capacity

Recall from the last chapter that one source of autonomy deficits is effects on the capacities required for autonomy: baseline capacities, freedom from external constraints, absence of cognitive inhibitions, and having a sufficiently wide range of options. We begin with the effects of mobile devices on these capacities. In this section, we discuss reasons to think that mobile devices (or our relationships with them) have negative effects on all of them. Having a problematic relationship with your phone has several

negative effects on your baseline capacities. As we explain in the next section, it affects executive function, working memory, cognitive functioning, sleep, and much more.

3.3.1.1 Baseline Capacities

We begin with effects on baseline capacities (the ability to form intentions, make coherent judgments, exercise executive function, etc.). These findings can be sorted into two kinds. We can call one of these direct effects, those which report a direct association between mobile device use and some negative outcome for autonomy, such as the inability to pay attention. The others are indirect, they report on an association between mobile device use and some outcome that isn't itself negative for autonomy as such, but which is associated with a negative outcome. Here, loss of sleep is a good example: loss of sleep *itself* isn't a diminution of a capacity, but we do know that the underslept tend to have diminished capacities.[5]

Following the meta-analysis by Wilmer et al. (2017), we will organize our discussion of direct effects into the following categories: attention, memory, delay of gratification, and general cognitive functioning. While these capacities might not be a complete list of the basic capacities essential for autonomy, they must at least come close. In order to set ends, we need to be able to attend to reasons that speak in favor of or against a choice. In order to pursue those ends, we need to monitor our progress and reconnoiter as obstacles come and go.[6] We need to remember what we have done or what we want to do next. We need to resist temptations that will ultimately set us back, and we need to avoid mundane errors.

Let us begin with a discussion of attention. One of the main concerns in this literature is the impact of mobile devices on focused attention, "the ability to maintain a directed attentional focus over an extended period of

[5] For the effects of smartphone use on sleep, see Cain and Gradisar (2010) or Sohn et al. (2019). For the effects of sleep deprivation on cognitive capacities, see Lim and Dinges (2008).

[6] We would like to avoid overstating our conclusions in this chapter. When we say that smartphone use can interfere with our ability to set and pursue our own ends, we do not mean to suggest that using mobile devices makes us *utterly incapable* in this regard. That would obviously be false. Our conclusion after surveying the empirical literature is far more modest. We simply believe that the excessive use of mobile devices is making us *less* autonomous. When someone has a problematic relationship with her phone, this is likely to make her less capable and less authentic than she would otherwise be. It would be a dramatic overstatement to imply that it renders her totally incapacitated or fully inauthentic. Instead, autonomy should be seen as a normative ideal, and problematic smartphone use is an obstacle that makes it harder for us to live up to the ideal.

time" (Ibid. 4). In their discussion of attention, Wilmer et al. (2017) make the helpful distinction between exogenous and endogenous intrusions on attention. These at least roughly correspond to what Eyal calls "external" and "internal" triggers, with the former—for Wilmer et al. (2017)—occurring when an environmental cue (e.g., a notification) fuels an interruption and the latter occurring when "[one's] own thoughts drift toward a smartphone-related activity, and thereby evince an otherwise unsolicited drive to begin interacting with the device" (Ibid. 4).

As Wilmer et al. (2017) note, both sorts of interruptions can initiate a chain of other distractions:

> Once attention has been shifted to the smartphone for one purpose (e.g., by virtue of a specific notification source), users often then engage in a chain of subsequent task-unrelated acts on the smartphone, thereby extending the period of disruption. Studies exploring these "within-phone" interruptions have found that task completion in one app can be delayed by up to 400% by an unintended interruption from another app (Leiva et al. 2012). And, some evidence suggests that the more "rich" (e.g., including a visual image rather than just text) the information encountered during an interruption, the more detrimental the distraction is likely to be with respect to primary task completion. (Levy et al. 2016). (Ibid. 4)

These are the sorts of interruptions that the Hooked Model aims to make us more susceptible to, and they are emblematic of the sorts of interruptions that concern us from the point of view of autonomy, because, among other things, these are the sorts of interruptions that introduce errors into our thinking and execution of tasks we have set for ourselves. Consider, for instance, *resumption errors*, "errors that arise in task performance that is resumed following an interruption or task-switch" (Ibid. 5).[7] The likelihood of committing such an error increases sharply when an interruption exceeds 15 seconds, which is a threshold that smartphone interruptions—which often snowball in the way described above—commonly exceed.[8] We are even more susceptible to these sorts of endogenous interruptions out of a desire for stimulation when our ongoing tasks fail to entertain us.[9]

Turning now to exogenous interruptions, it is striking how disruptive mobile devices can be. We noted earlier that mobile devices are often

[7] Wilmer et al. cite Monk (2004); Cades et al. (2007); and Brumby et al. (2013).
[8] See Monk et al. (2008) and Leiva et al. (2012). Cf. Wilmer et al. (2017).
[9] See Melcher (2013). Cf. Wilmer et al. (2017).

engineered to grab attention, with alarms, lights, and vibrations. It is worth noting, however, that we seem to be so conditioned to give these devices our attention that their mere presence distracts us and introduces error even when they are off and even when they are not our own device.[10] Screens *themselves* seem to leech attention. Exposure to notifications—let alone responding to them—of course exacerbates these effects. Indeed, it has been shown that exposure to notifications "significantly damages performance on an attention-demanding task" (Stothart et al. 2015, 896). A meta-analysis by Caird et al. (2014) confirms this concern, and they demonstrate the troubling implications for driving, where the introduction of error can be fatal.

Now we will turn from attention, which selects which information to focus on, to working memory, which keeps information readily available for processing. The mere presence of a mobile device has a detrimental effect on available working memory.[11] Ward et al. (2017) demonstrate that the strength of this effect is moderated by how much one uses their phone: "Ironically, the more consumers depend on their smartphones, the more they seem to suffer from their presence—or, more optimistically, the more they may stand to benefit from their absence" (149).

Not only does the presence of a phone impact working memory, behaviors associated with smartphones have negative impacts on working memory as well. For instance, a number of studies show that media multitasking (e.g., scrolling through a feed on one's phone while streaming a show on one's laptop) is linked to diminished working memory,[12] both when external distractions are absent or present.[13]

Taking stock of what we have said so far, it seems, at the very least, that the presence of phones—even if they are turned off—interferes with attention and working memory, two essential features of reasoning to a conclusion about what to do and then doing it. Let us now take a look at another important factor in achieving one's goals: delay of gratification.

In their (2016) paper, Wilmer and Chein demonstrate evidence of an oft-hypothesized connection between mobile device use and delay of gratification. Specifically, they find a significant correlation between device usage and one's discount rate of future rewards, that is, their susceptibility

[10] See Thornton et al. (2014).
[11] See Ward et al. (2017).
[12] See Cain et al. (2016).
[13] See Uncapher et al. (2016).

to immediate gratification. Now, as Wilmer et al. (2017) note, like many studies, it is correlational: it could, on the basis of these findings, just be that mobile device use is evidence of higher discount rates. However, there is at least one experimental result that suggests that the causal arrow flows in the direction that worries critics of mobile devices. Hadar et al. (2015) ran a longitudinal study where participants lacking smartphone experience were tested before and after receiving smartphones. Subjects experienced an increase in impulsivity and a decrease in information processing three months after receiving the devices.

Now that we have reviewed some of the studies that speak to the corrosive effect that mobile devices have on attention, memory, and delay of gratification, let's look at the effects of mobile device use on cognitive functioning more generally.

One way to get a sense of the impact mobile devices have here is by looking at their effect on grades.[14] We can begin by noting that there is a large number of studies showing a negative correlation between mobile device use and academic performance.[15] For instance, Lee et al. (2017) and Dietz and Henrich (2014) show, via random assignments of treatment conditions, that there is a causal connection between mobile device use and the comprehension of material from a lecture. It is also interesting to note that one of these studies—Sana et al. (2013)—showed how laptop use in class not only affects the user but *all* those who could see the screen (Sana et al. 2013, cf. Wilmer et al. (2017)). This shows how the negative effects of mobile devices can flow beyond the individual user. It is not just *your* cognition that is affected but also those around you. There are also reasons to believe that mobile device use affects cognition outside the context of education. A study from Baumgartner et al. (2014) demonstrates a correlation between multimedia multitasking and failures of executive function. Cain et al. (2016) corroborated these links in a laboratory setting (cf. Wilmer et al. (2017)).

All of these effects of smartphones have a *direct* impact on capacities linked to autonomy (executive function, working memory, attention span, etc.). But smartphone use also has an *indirect* effect on our baseline

[14] See Wilmer et al. (2017).

[15] For more information on smartphones and schools, see Beland and Murphy (2016); Dietz and Henrich (2014); Fox et al. (2009); Jacobsen and Forste (2011); Junco (2012a, b); Junco and Cotten (2012); Karpinski et al. (2013); Kirschner and Karpinski (2010); Rosen et al. (2011); Lee et al. (2017); Levine et al. (2007); Rosen et al. (2013); and Sana et al. (2013).

capacities. As we have noted, this happens in the case of sleep deprivation. Although sleep is not one of the capacities that is linked to autonomy, sleep deprivation leads to a variety of cognitive deficits. We will close this section with some brief remarks about the effects mobile devices might have on sleep, anxiety, and depression—factors that have an impact on cognitive functioning.

It is fairly well-known that mobile device use is associated with lack of sleep and that lack of sleep has negative impacts on cognition (Lim and Dinges 2008, cf. Wilmer et al. 2017). For instance, in their meta-analysis of the association between nighttime mobile device use and sleep outcome, Carter et al. (2016) found "a strong and consistent association between bedtime media device use and inadequate sleep quantity … poor sleep quality … and excessive daytime sleepiness" (1203). Further, companies have an obvious interest in keeping us awake, expanding the pool of attention from which they can draw to show us advertising or to sell us products. The profit motive is the driving factor behind this feature of the attention economy.[16] Companies maximize ad revenue by maximizing engagement. The more time we spend looking at screens, the more ads we see. As Johann Hari recounts in *Stolen Focus* from his conversation with physician and sleep researcher Charles Czeisler:

> In a society dominated by the values of consumer capitalism, "sleep is a big problem" … "If you're asleep, you're not spending money, so you're not consuming anything. You're not producing any products." He explained that "during the last recession [2008] … they talked about global output going down by so many percent, and consumption going down. But if everybody were to spend [an] extra hour sleeping [as they did in the past], they wouldn't be on Amazon. They wouldn't be buying things." If we went back to sleeping a healthy amount … Charles said, "it would be an earthquake for our economic system, because our economic system has become dependent on sleep-depriving people. The attentional failures are just roadkill. That's just the cost of doing business." (Hari 2022, 76–77)

[16] Roger McNamee, an early investor in Facebook and former mentor of Mark Zuckerburg, puts this point rather emphatically: "If you want to fix this as quickly as possible, the best way would be for founders of these companies to change their business model away from advertising … We have to eliminate the economic incentive to create addiction in the first place" (Stolzoff 2018).

It is no wonder that mobile devices conflict with sleep. Most of us keep our phones at arm's reach while we sleep, and it can be difficult to resist the pull of the attention economy as it entices us with its triggers to stay up a little longer to consume more content.

Moving on to anxiety and depression, which are themselves associated with diminished cognitive functioning (American Psychiatric Association 2013, cf. Wilmer et al. 2017), a number of recent studies have shown a causal connection between mobile device use and negative mental health outcomes. For instance, after monitoring 143 undergraduates for a week, Hunt et al. (2018) randomly assigned severe limits to social media to some undergraduates but not others for three weeks. During the three weeks, the "limited use group showed significant reductions in loneliness, depression, anxiety, and fear of missing out" (Ibid.). Similar causal results have been reported by Sherlock and Wagstaff (2019), Lambert et al. (2022), and Brailovskaia et al. (2022).

These results should not surprise us. In her breakout book *How to Do Nothing*, Jenny Odell describes what she, in unplugging from the attention economy, wants to escape:

> To me, one of the most troubling ways social media has been used in recent years is to foment waves of hysteria and fear, both by news media and by users themselves. Whipped into a permanent state of frenzy, people create and subject themselves to news cycles, complaining of anxiety at the same time that they check back ever more diligently. The logic of advertising and clicks dictates the media experience, which is exploitative by design. Media companies trying to keep up with each other create a kind of "arms race" of urgency that abuses our attention and leaves us no time to think. (Odell 2019, 59)

Once this "logic of advertising and clicks" is laid bare, it comes as no surprise that, as Hunt et al. (2018) and others have found, social media makes us anxious, depressed, and worse off in other ways. Thus, it is no surprise that mobile device exposure can lead to a diminution of the basic capacities that are necessary for autonomy.

3.3.1.2 External Constraints

Given the negative effects of mobile devices and the attention economy, the idea of unplugging entirely might sound appealing. But research has found that this is harder to accomplish than you might think. As we will

explain later in our discussion of authenticity, people routinely report that they use their devices more than they would like to. One survey found that 62% of US consumers have made efforts to cut back on their smartphone usage, but only half of them had any success.[17]

Indeed, smartphone and social media users often report feeling "addicted" to their devices, and you would have to pay them to quit. As we mentioned in the introduction, a group of economists conducted one of the largest studies of the effects of social media on well-being in 2018. They recruited a sample of 2743 Facebook users, and the researchers offered them various amounts of money to quit using Facebook for four weeks. They wanted to determine how much money it would take to get someone to quit for just one month. The result ended up being higher than most estimates. The mean figure was $180.[18] Ironically, even though users had to be paid to quit Facebook, most of them were happier without it. Allcott et al. (2020) found that those who quit for a month reported increased well-being equal to the jump one would get *from earning about $30,000 more in annual income* (654). During the experiment, they spent more time with family and friends, and they used Facebook much less after the experiment was over.

But in many cases, our relationships with our phones and social media are not built on our own terms. As we saw in the last section, many of us are trapped in the cycles the Hooked Model runs us through, which were deliberately developed so that companies could maximize the amount of time we spend looking at screens. What is more, we are often subject to external constraints that make it so that unplugging is not a live option. In the last chapter, this was presented as the second component of capacities that are required for autonomy. This idea is fairly straightforward. The more you are subject to external constraints, the less autonomous you are. When it comes to our use of mobile devices, perhaps the most obvious instance of external constraints is employment.

It is not uncommon for employers to require employees to be in constant communication by email or through other digital channels (e.g., Slack, text message, WhatsApp). Under these conditions, an employee is never truly off the clock, and this all but necessitates an unhealthy

[17] These figures come from a 2018 survey of 2000 customers. See Deloitte (2018).

[18] Allcott et al. (2020). Other companies have offered much larger sums. One company offered $2000 for a two-month hiatus from social media, and another offered a chance to win $2500 for 25 days of abstention. See Flood (2022) and Rueda (2021).

relationship with one's smartphone. Cal Newport (2021b) argues that these working conditions predictably lead to burnout and high turnover. He suggests that these policies are not in employers' best interests as they undermine productivity rather than promote it. But he is particularly concerned about the effect this has on employees' well-being. When employers require "slavish devotion to in-boxes and chat channels, then this adds up to a whole lot of global miserableness! From a utilitarian perspective, this level of suffering cannot be ignored—especially if there is something that we might be able to do to alleviate it" (Newport 2021a).[19]

Workers often find that they have no choice but to comply with the technological demands of their employers. If a worker is highly skilled and is able to find employment elsewhere, then she may have global autonomy with respect to this issue, but employees typically lack local autonomy in these matters. To fight back against this situation, workers in Europe have won what they call "the right to disconnect." In 2016, this right became part of French labor law; companies with more than 50 employees are now required to negotiate after-hours email policies with employees or their unions. This law was introduced after a report found that many workers were using their work phones at home, and the majority of them wanted more control over their use of technology.[20]

But the majority of workers have much less control over their employer's technology policies. And this is one way that we are subject to external constraints when it comes to our relationships with mobile devices. The situation in schools is quite similar, as the use of devices like tablets and laptops has become a mandatory part of the curriculum for most students. A survey conducted by the US Department of Education found that 94% of public schools used them for the 2022–2023 school year.[21]

We have witnessed this firsthand in the lives of our own children. The Miami-Dade public school system *requires* students to use electronic devices *starting at age 4*. Like 10 million other students in the US, children in Miami-Dade are required to use a digital curriculum called "i-Ready."[22] We were dismayed (but not surprised) to learn that i-Ready includes a reading activity designed to look like Instagram. When a

[19] Mark et al. (2012) offer some confirmation of this hypothesis, showing in an empirical study that limiting workers' access to email both decreased anxiety and improved productivity.

[20] Mettling (2015, 19).

[21] Kuykendall (2022).

[22] Braatz (2022).

five-year-old student correctly answers a question on "Yoop-o-Gram," she is rewarded with a flurry of likes and hearts.

To be clear, we are not simply reactionaries who are opposed to the use of all technology in school. As we will explain in the next chapter, we are not suggesting that we should become Luddites and eschew technology altogether; when used and designed properly, mobile devices can enhance our autonomy. Obviously, students should learn how to use technology, and schools would fail to promote students' autonomy if they did not introduce them to a variety of digital tools. So we are not advocating for the abolition of digital curricula. Our point here (as elsewhere) is simply that we should be thoughtful about how we integrate technology into our lives, and this is especially pressing when the use of technology is mandatory. It is easy for educational administrators to see the upsides of using technology in school. When used appropriately alongside classroom instruction from teachers, digital tools can be indeed useful ways of pursuing learning outcomes. But there are good reasons to be careful about harnessing the addictive components of these devices.

Gamification—the idea behind the hearts and likes on Yoop-o-gram—often looks like a great way to increase engagement, but it has downsides as well. Thi Nguyen (2020, 2021) argues that gamification is alluring insofar as it purports to be a way of getting people to do otherwise boring and repetitive tasks. By pursuing points, people might be more willing to perform those tasks. What makes games appealing, in his view, is that they offer "value clarity." In many domains of our lives, it is very difficult to know which values to promote at which times. Should you be working, cleaning your house, or doing something else entirely? But by simplifying things to a matter of points, games offer clarity. The goal becomes clear and straightforward (e.g., get more hearts, get more likes).

When this is applied to other domains of life (in the form of gamification), it has the effect of flattening out our values. Perhaps students should be interested in education for a variety of reasons: personal growth, edification, enhanced autonomy, etc. But when education is gamified (e.g., through the single-minded pursuit of maximizing GPA), all other values disappear as we become victims of what he calls "value capture." Value capture involves a radical simplification of a wide range of values as it flattens them into something quantifiable. This can have perverse effects. Students who were once interested in learning and taking on difficult

challenges might start taking less challenging, "easy A" courses in pursuit of a high GPA.[23]

This point about gamification has important implications for the use of technology in schools. When it comes to digital technology, we should recall McLuhan's claim that "the medium is the message" (McLuhan and Lapham 1994). In many cases, the content on the smartphone is not the entire story. The fact that the content is on the smartphone (or tablet) is also relevant. When it comes to promoting literacy in children, researchers are unequivocal about the importance of getting children to appreciate and enjoy reading.[24] This often requires that books be present in the child's home, that she sees adults in her life reading books, etc. As neuroscientist and literacy advocate Maryanne Wolf has shown, getting children to become interested in books has become more difficult in a context where they must compete with smartphones, tablets, and YouTube.[25] She explains how reading involves an amazing interplay of neural components—vision, cognition, affect, attention, and language. Reading awakens both hemispheres and all five layers of the brain. Given that our brains did not evolve for reading, acquiring this ability requires neuroplasticity. The brain rewires itself to make deep reading possible. The danger, of course, is that we can undo this as well. So although we see lots of words on screens (news stories, social media posts, etc.), there is a vast neurological difference between this kind of shallow reading and deep reading.[26]

[23] See Nguyen (2021) and Nguyen (2020) for further discussion of gamification and value capture.

[24] Willingham (2015).

[25] Wolf (2018). To be clear, her concern is not limited to children; it extends to adults as well. She even explains her autobiographical interest in this topic as she explores her own relationship with literature and deep reading. Similarly, Johann Hari (2022) describes his own struggle to reclaim his ability for deep reading, and he explains how he regained his capacity while engaging in a three-month-long "digital detox." He writes: "I finished the third volume of *War and Peace*. As I closed its last page, I realized I had been sitting there for most of the day. I had been reading like this, day after day, for weeks. And I thought suddenly: It came back! My brain came back! I feared my brain had been broken, and this experiment might just reveal I was a permanently degenerated blob. But I could see now that healing was possible. I cried with relief. I thought to myself, I never want to go back to email. I never want to go back to my phone" (60–61). For further discussion of children and touchscreen devices, see Rocha and Nunes (2020).

[26] This is one of the central claims of Nicholas Carr's book *The Shallows: What the Internet Is Doing to Our Brains* (2010).

Neuroscientists have argued that this effect may be especially pronounced in "digital natives" who grew up with internet technologies. Loh and Kanai (2014) write:

> With multifaceted affordances, the Internet environment has profoundly transformed our thoughts and behaviors. Growing up with Internet technologies, "Digital Natives" gravitate toward "shallow" information processing behaviors characterized by rapid attention shifting and reduced deliberations. They engage in increased multitasking behaviors that are linked to increased distractibility and poor executive control abilities. Digital natives also exhibit higher prevalence of Internet-related addictive behaviors that reflect altered reward-processing and self-control mechanisms. Recent neuroimaging investigations have suggested associations between these Internet-related cognitive impacts and structural changes in the brain. (506)

For this reason, we may want to be cautious about how we use technology in schools, especially with young children. It is tempting to think that we are enhancing children's ability to read by having them sound out words on a tablet, but there are reasons to worry that the child might learn to sound out words but never develop a desire to read books (whether physical or electronic). Perhaps we should not be surprised when children grow up to find themselves in the same situation as Esther Povitsky. Much as they might like to read a book, they will find it difficult to resist the urge to put it down to look at Instagram or TikTok. In some instances, external constraints limit our options. Some children have no choice but to learn to read on screens.

There are many cases where we are compelled to adopt the ends that others have set for us. We often have to accept the terms that our schools or employers give us. In and of itself, this can undermine autonomy. Given that autonomy involves setting and pursuing your own ends, this capacity is surely undermined when someone else sets the terms of your life for you. What is more, when external constraints require us to have unhealthy relationships with mobile devices, this can have downstream effects in other areas of our lives. As we saw in the previous section, it can weaken some of our baseline capacities. In the next section, we will explore other negative effects, including the ways that mobile devices and social media undermine our self-esteem.

3.3.1.3 Cognitive Inhibitions

The quote from Sean Parker at the beginning of this chapter notes that we are often brought to our mobile devices via the exploitation of vulnerabilities in human psychology, including the drive for social validation. As former Google design ethicist Tristan Harris notes, "Everyone innately responds to social approval … That's why it's so important to recognize how powerful designers are when they exploit this vulnerability" (Harris 2016). Here, we explore one consequence of the power of exploiting this vulnerability: its effects on self-esteem. As we will soon show, there is a tight connection between certain mobile device applications and low self-esteem (i.e., being critical of oneself, downplaying one's positive qualities, judging oneself as inferior to others, using negative words to describe oneself, assuming luck plays a large role in one's successes, blaming oneself when things go wrong, and disbelief in compliments about oneself). In other words, interactions with mobile devices seem to generate a host of cognitive inhibitions of the sort that we think are detrimental to autonomy. Much like Benson's example of the wife in *Gaslight*, we come to question our own capacities in such a way that we rely on others and forfeit some of our agency to them.

Most of us are familiar with the idea that news feeds, especially on image-driven apps like Instagram, present a distorted—if not entirely false—image of others' lives. Indeed, some lives—such as that of Shudu[27] (244,000 followers at the time of writing) and Lil Miquela (3 million followers at the time of writing)—are complete fabrications, entirely computer generated. They are, as one reporter put it, "physically perfect women made of pixels, standing in for women who have long been pressured to become physically perfect" (Tiffany 2019).

While many users are aware of the fact that what we see on Instagram is unrepresentative of real life, this doesn't stop us from comparing ourselves or each other to them. Indeed, a comment on a photo of Lil Miquela posing with a human model reads, "the robot more pretty [sic]."[28] Meanwhile, many other comments express confusion over whether Lil Miquela is a "robot." In a post promoting a makeup brand, Shudu is complimented on her beauty and told that her eyeshadow works well with her

[27] Shudu, a Black woman, is a character created by a White man. For discussion of the issues this raises, see Brucculieri (2018), Square (2018), and Jackson (2018).

[28] See https://www.instagram.com/p/CGc6qt3n1M8/?img_index=1

skin tone. She is asked, "how do I become a model [?]"[29] These comments seem to assume that Shudu is a human. This is understandable. Shudu's representations are photorealistic and easily mistaken for actual photos. Further, the company she is promoting in the post seems—as far as we can tell—to exclusively sell real, physical makeup intended for use on flesh and blood humans.

These are no doubt extreme examples, but the point here generalizes beyond extreme cases. Despite the fact that we know that posts on sites like Instagram are idealizations, when we are exposed to these images they nevertheless enter into our social comparisons. These comparisons include *upwards social comparison*, where we compare ourselves to others that we think are faring better than ourselves (Festinger 1954). Indeed, one study—Fardouly and Holland (2018)—exposed women ages 18 to 25 years old to posts from social media containing images of attractive women, with some participants but not others being shown the images with a disclaimer, such as, "Not real life—I didn't pay for this outfit, took countless photos trying to look hot for Instagram" (4317), and others being shown the images without the disclaimers. Women in both groups left the exposure feeling much worse about their bodies than those in the control group, who were shown images of travel. This suggests that even if we are aware of the distortions of social media, we engage in detrimental upwards social comparisons anyways.

With this in mind, it should come as little surprise that there seems to be a fairly robust association between social media use and low self-esteem. For instance, Kelly et al. (2019) consulted data from 10,904 English 14-year-olds and found that increased "social media use was associated with, among other things, low self-esteem and body image," with this effect being larger for girls than boys (60). Twenge and Farley (2021) corroborate these results, and add to our understanding by showing that the effects not only vary by gender, but also by activity: "Hours spent on social media and Internet use were more strongly associated with self-harm behaviors, depressive symptoms, low life satisfaction, and low self-esteem than hours spent on electronic gaming and TV watching" (207).[30]

[29] https://www.instagram.com/p/Cb-ZoKtIVVo/

[30] See also Barthorpe et al. (2020). Further studies showing a connection for teenagers between social media use and low self-esteem include Woods and Scott (2016), who showed that "those who were more emotionally invested in social media experienced poorer sleep quality, lower self-esteem and higher levels of anxiety and depression" (41).

The findings by Twenge and Farley might raise a challenge to our project: given that it isn't mobile device exposure as such, but what one does on the mobile device that matters, the problem here could be said to lie with social media, as opposed to mobile devices. A few notes are in order here. First, as noted above, mobile devices aren't our *only* concern; they are simply our *primary* concern. Our larger concerns have to do with the effects of web-connected screens on autonomy, but, as we have claimed above, there are reasons to focus on a more narrow target than this, given the power that mobile devices have, in particular, to influence behavior.

Further, we believe that our focus on mobile devices—due to their portability and powers of influence—is only reinforced by the negative effects of social media. Here it is worth noting that most social media access now occurs via phones. Indeed, recent estimates based on Facebook's own data show that more than 98% of users use their phone to access the site, with over 80% *exclusively* accessing the site via phone.[31] The fact that access via phone is so high is not an accident: that we can check in on our social networks anywhere we might be (in bed, at school) in otherwise free moments (going to the bathroom, waiting at a stoplight), and that this checking can become a habit (in Eyal's sense), allows us to be exposed to this material more than we would otherwise be able to.

What we have said so far might give the impression that social media use only has an effect on the self-esteem of young teenagers. But this is not so. For instance, Hanna et al. (2017), in a study of undergraduate men and women, found a positive correlation between Facebook use and self-objectification (how much one is "preoccupied with how their body appears to others" 174), social comparison ("comparing oneself with others" 172), depression, anxiety, and low self-esteem. Several studies mirror Kelly et al. (2019) in finding distinct gender patterns, but in adults. For instance, Miljeteig and von Soest (2022) found that while social media was correlated with stability of self-esteem among both men and women, they also found that, for women, recent low self-esteem was more predictive of current social media use. They further found that social media use was more predictive of current low self-esteem. With this they note that

[31] Here it is worth noting that there is reason to think that in many international contexts, these numbers would only be higher. As part of its controversial "Free Basics" program, Facebook brought free, limited internet service to developing markets—such as Colombia, Ghana, Kenya, Mexico, Pakistan, and the Philippines—via a Facebook mobile app that gives users free access to certain websites (including Facebook) (Solon 2017b).

their findings "support the notion of a reciprocal relationship between social media use and self-esteem for women, where self-esteem level may motivate women to use social media more frequently and social media may be a source of lower self-esteem" (373).

Up until now, most of the studies that we have presented in this section have been correlational. As always, this relationship could just be a result of the fact that lower self-esteem drives users to social media. However, there are studies that show that the causal direction flows in the other direction. Ozimek and Bierhoff (2020), for instance, exposed some subjects but not others to Facebook, with the subjects exposed to Facebook reporting lower self-esteem post exposure. Vogel et al. (2014) found similar results. Sherlock and Wagstaff (2019) found that exposure to beauty and fitness Instagram images significantly reduced self-assessments of attractiveness. Finally, Wolfe and Yakabovits (2022) were able to show that undergraduate women exposed to posts containing edited photos were more likely to edit their own photos when asked to then take a selfie. The same study also showed that "photo editing was associated with adverse changes in perceived attractiveness and mood" (1).

The takeaway here is that mobile devices expose us to content that could instill the sorts of cognitive inhibitions that put a drag on our autonomy. This is not entirely accidental. As we have seen through the testimony of people like Eyal, Parker, and Harris, the architects of services designed to run on our phones—and thus to go with us everywhere we go—will leverage vulnerabilities such as our inborn desire for social validation to get our eyes on the screen. However, in seeking this validation we often engage in activities that lower our opinions of ourselves due to our natural tendency to compare ourselves to those that we take to be faring better than ourselves. This is harmful to us in many ways. Diminished self-esteem negatively affects many aspects of well-being, and, more germane to our present concern, it can be disastrous for personal autonomy.

3.3.1.4 Options

In articulating the importance of options, we gave the example of a child, who, while formally free to do something else, considered being a coal miner as his only live option. No one literally forces the child into coal mining; no one denies him other options. But, he chooses as though he has no other options. This highlights the importance of not only having other options but being alive to them, being able to see them *as* options.

In *On Education*, Harry Brighouse argues for this point forcefully, on the grounds that it is an important facet of autonomy. He concludes that "children have a right to learn about a range of ways of living and to the kind of education that will enable them to reflect on their own way of life in the light of these alternatives, and, ultimately, to revise or reject the way of life their parents would pass down to them" (Brighouse 2006, 2). As Brighouse stresses, this is the only way for children to make informed decisions about which lives work *for them*. Citing Mill approvingly, he stresses the importance of exposing a child to those with convictions that differ from those of her parents and teachers:

> Nor is it enough that he should hear the arguments of adversaries from his own teachers, presented as they state them, and accompanied by what they offer as refutations. That is not the way to do justice to the arguments, or bring them into real contact with his own mind. He must be able to hear them from persons who actually believe them; who defend them in earnest and do their very utmost for them. (Mill 1988, 35)

With this in mind, let us ask whether children—or we, ourselves—have sufficient access to less connected lives, to see if we can see them as suitable *for us*.

We suspect that we do not, and that this is not entirely due to the presence of external constraints mentioned above (though, of course, external constraints only exacerbate the concern). Indeed, despite study after study showing, e.g., that we are happier when we have less access to our email (Mark et al. 2012) or Facebook (Allcott et al. 2020), most of us, in the face of that knowledge, turn back to our devices.

On this last front, Ezra Klein shares an observation that, for us, resonates because we are so familiar with the phenomenon:

> There's a state I get in, less and less these days, but in part because of the way my world works and my phone and my computers. I now associate it with plane flights because nobody can call me, and I don't buy internet. It's a state that I only seem to access when reading, and only when reading without distraction for a long period of time. It's very strange, and it is one of my most loved states. (Klein 2022)

What fuels Klein's ability to access this state in the air? His hypothesis—and ours—is freedom *from* the internet, freedom *from* a fully functional

smartphone. Yet what does he—what do we—do when he lands? Deliver himself back into the clutches of his phone:

> [E]very time I get off of a plane, I say to myself, I'm going to do that more … I'm going to sit, and I'm going to have quiet time with a book … And then I don't. (Ibid.)

And why? To be sure, there are many reasons (Hooked cycles, habits, the chaos of ordinary life, and so on). But among them there is the fact that it is simply, for many, *unimaginable* to live life without a phone.

This fact isn't one that only middle-aged journalists and professors are familiar with. As "Billie," one of the teenage girls Nancy Jo Sales interviewed for her book *American Girls: Social Media and the Secret Lives of Teenagers*, says, teens often feel that they do not have an option to not have a smartphone: "You have to have an iPhone. It's like Apple has a monopoly on adolescence" (Sales 2017, 251; cf. Castro and Pham 2020). Billie is not alone in feeling this way. "Emily," a teen Jean Twenge interviewed for her book *iGen*, hits a similar note: "Everyone uses it. It's a good way to, like, make plans with people. If you don't, you might miss out on plans that you could have gone to" (Twenge 2017, 53, cf. Castro and Pham 2020). And, again, it isn't just teenage girls (or middle-aged journalists and professors) who feel this way. Nearly all of us are familiar with the fact that, now that each of us has the "option" to have a smartphone, but nearly none of us feel as though we have the option to *not* have one.

Simply put: many of us don't really see a less connected life as a genuine option. And this, we think, undermines the thought that the lives we have chosen—ones that are saturated with mobile devices and the attention economy—are done with any genuine understanding of whether they are right *for us*.

3.3.2 Authenticity

Let us now turn to the effects of mobile devices on authenticity, exploring these effects in terms of its four facets, i.e., freedom from manipulation and coercion, consistency with the agent's motivational states (whether they be volitional, evaluative, or long-term plans), freedom from alienation, and not being adaptive.

3.3.2.1 Manipulation

In "How Technology is Hijacking Your Mind," design ethicist Tristan Harris delivers a laundry list of tactics product designers use to "play your psychological vulnerabilities … against you in the race to grab your attention" (Harris 2016). If there is a theme to his list, it is manipulation. Here, we'll look at a few of the items on his list to substantiate the claim that our mobile devices are used to manipulate us.

We'll begin with the second item on the list, "Put a Slot Machine In a Billion Pockets." He begins this discussion—much like Eyal does his—with an estimation of how often we check our phones. His count, like Eyal's, is 150. He adds: "Why do we do this? Are we making *150 conscious choices*?" (Harris 2016). The answer he gives to the first question is variable rewards, and to the second, no. On his telling, variable rewards turn your phone into a slot machine:

> When we pull our phone out of our pocket, we're *playing a slot machine* to see what notifications we got.
> When we pull to refresh our email, we're *playing a slot machine* to see what new email we got.
> When we swipe down our finger to scroll the Instagram feed, we're *playing a slot machine* to see what photo comes next (Harris 2016).

This is a comparison that Eyal himself is happy to make: "Variable rewards are prevalent in many … habit-forming products," such as "slot machines and lotteries" (Eyal 2014, 9).

Turning back to manipulation, the idea here isn't to moralize against playing slot machines. Instead, it's to think about what a slot machine can do to a player, i.e., to get her to engage in "game play" more than is good for her. As Harris, citing NYU professor Natasha Dow Schull, warns, the variable reward aspect of slot machines is highly effective in achieving this effect (Harris 2016). In a 2013 presentation, Harris says, "These are attention casinos because the house always wins" (Harris 2013).

And, as Eyal's playbook demonstrates, users can be made to play these slot machine-like games without being aware of the larger context (i.e., the Hooked cycle) that gets them to use the device automatically, that is "with little or no conscious thought" (Eyal 2014, 1). Here, we have powerful psychological vulnerabilities being exploited to get us to do things in ways that work around our rational agency; that is, we have a prime example of manipulation.

This form of manipulation isn't outright deceptive, but other forms that Harris cites are. Consider list item five, "Social Reciprocity." As we are all well aware, we have a natural inclination to reciprocate others' gestures. If someone says "Hi," you say "Hi" back to them. If someone waves to you, you'll wave back. This is leveraged by product developers to get us on their sites. Harris offers LinkedIn as an "obvious offender" of this tactic:

> When you receive an invitation from someone to connect, you imagine that person making a *conscious choice* to invite you, when in reality, they likely unconsciously responded to LinkedIn's list of suggested contacts. In other words, LinkedIn turns your *unconscious impulses* (to "add" a person) into new social obligations that millions of people feel obligated to repay. All while they profit from the time people spend doing it. (Harris 2016)

He also mentions—in the context of a different but not dissimilar tactic ("Social Approval")—how a similar effect can be achieved by auto-tagging users in photos:

> When I get tagged by my friend Marc, I imagine him making a *conscious choice* to tag me. But I don't see how a company like Facebook orchestrated his doing that in the first place … [W]hen Marc tags me, *he's actually responding to Facebook's suggestion*, not making an independent choice. But through design choices like this, *Facebook controls the multiplier for how often millions of people experience their social approval on the line*. (Harris 2016)

Here, Harris is making the plausible claim that we are sometimes deceived into using a social media service: we are given a false impression that we have a social obligation or have been noticed, driving us to the site to reciprocate or to bask in the warm glow of social approval. As we explained in Chap. 2, this is typical of manipulated behavior. Deceitful manipulation involves getting someone to act on the basis of enticements that she would not endorse if she had accurate information.

The final tactic that we will discuss is number seven: "Instant Interruptions vs. 'Respectful' Delivery" (Harris 2016). Our earlier discussion of attention should leave it as no surprise that messages that interrupt us immediately are more likely to get us to respond. As Harris notes: "Facebook Messenger (or WhatsApp, WeChat or SnapChat for that matter) would *prefer to design their messaging system to interrupt recipients immediately* … instead of helping users respect each other's attention" (Harris 2016). As he further notes, this effect is exaggerated when

messages have read receipts, that is, when the sender is notified in real time when the recipient has read the message. Leveraging the sort of psychology at play in "Social Reciprocity," this heightens the sense of urgency and prompts a response. These forces are, again, working at cross purposes with our best judgments: many of us have accidentally left ourselves logged on and seen a message that we then feel compelled to respond to, knowing all the while that had the message been sent to us offline we would have felt no urgency to respond.

In sum, our mobile devices—often because they reward and monitor us in real time—can, and often are, used to circumvent our rational processes, getting us to jump onto sites when we in fact know that this is not what we should or even want to be doing. More often than not, when our behavior is the result of manipulation, we find ourselves disapproving of our actions or desires. We act in a way that is inconsistent with our higher-order evaluations; we feel alienated from our desires and actions. We will explore these ideas in the following sections.

3.3.2.2 Incoherent Motivational States

In the previous chapter, we explained how the coherence of motivational states is a key component of autonomy. In addition to decision-making capacity, autonomy requires authenticity. Not only must you have the ability to set and pursue ends, the ends must be, in some sense, your own. As we saw, many philosophers defend "coherence" models of authenticity. They suggest that autonomous actions and desires are ones that the agent reflectively endorses. Indeed, this is the central focus of many of the models of autonomy that have been developed in the literature. It will be helpful to recall the three accounts that were discussed in the last chapter.

For Frankfurt and Dworkin, autonomy requires consistency between first-order desires and higher-order reflection. Not only do you want to exercise, you *want* to have that desire. You have a higher-order desire to have the first-order desire. You want that first-order desire to motivate your action. Gary Watson's account differs insofar as it relies on evaluations rather than desires. For Watson, the question is not whether or not you *want* those desires; instead, he thinks you should ask whether or not you *approve* of your desires. Are your desires consistent with what you think is most worthwhile? Finally, Michael Bratman's view tells you to ask whether or not your action is consistent with your long-term plans.

With these models in hand, we are better equipped to discuss the effect of smartphones on authenticity. Perhaps the most natural starting place

would be to ask yourself the following questions: First, take a few seconds to find out how much time you spend on your phone (instructions provided below).[32] Are you content with your usage data? Do you approve of your first-order desire to unlock your phone when you engage in higher-order reflection?[33] Is your usage consistent with your evaluative judgments about what is most worthwhile?

You might be one of the lucky few who can say yes to those questions. But polling data show that the majority of Americans do not feel good about how much time they spend on their phones. A 2022 Gallup poll of 30,000 adults in the US found that 58% of them believe that they spend too much time using their smartphone.[34] This number has risen considerably since the last poll in 2015 which found that only 39% of users held this belief. In 2022, twice as many people reported that their smartphone has made their lives worse overall (compared with 2015 results). A majority of users still believe that their smartphone has improved their lives, but this might have something to do with the fact that half of the respondents said that they cannot imagine life without their smartphone (a topic we discussed above in terms of our limited options).[35]

[32] Both Android phones and iPhones allow you to see your screen time data. With an Android, go to Settings, then tap "Digital Wellbeing & parental controls." You can then go to the Dashboard to see more fine-grained data about which apps you use the most and how many notifications you are getting. With an iPhone, go to Settings and then "Screen Time." Tap "See All Activity" to get a summary of your weekly use. You can do the same with an iPad. If those numbers are not alarming enough, you can plug them into a "lifetime screen time" calculator. By telling it your age and your average daily hours, it will tell you how many years of the rest of your life you will spend looking at your phone. In 2018, Adobe polled 1000 users in the UK and found that the average Gen Z user spends a staggering 10.6 hours per day consuming online content (the majority of which they watch on their phones). For someone who is 20 years old, that adds up to 27 years of their remaining life. See Hebblethwaite (2018).

[33] But we are not likely to reflect on this decision at all. As Eyal points out, the urge to unlock your phone is often unconscious. You do not consciously deliberate about whether or not you would like to use your phone; you get it out "with little or no conscious thought" (Eyal 2014, 1).

[34] See Saad (2022).

[35] It is important to keep the conflicting statistics in mind. Although 65% of respondents believe that their smartphone has made their lives better, 58% said that they use their phone too much. And 83% of respondents said that they keep their phone on them at all times (Saad 2022). This means that some of the people who do *not* think that their phone has improved their lives still keep their phone on them during every waking hour. This seems to indicate a tension in their motivational states. It would seem that their first-order desires are not consistent with their higher-order desires or their evaluative judgments.

We have come to depend on our phones in so many ways that it should come as no surprise that they seem indispensable. But, as we explain in the next chapter, this is consistent with our view. We are not in denial about the many benefits that smartphones have to offer (calendars for keeping appointments, GPS to navigate and get traffic updates, etc.). But for many of us, even if we recognize the ways that our phones help us set and pursue our ends, we use them in ways that weaken our capacities as well. What is more, the behavior often seems compulsive. We find ourselves using our smartphones whether we want to or not. This gives us a good reason to believe that our actions and desires are inconsistent with our higher-order reflection.

Once again, Esther's anecdote is useful. She was unable to achieve the end she set for herself (reading a book) due to her compulsive smartphone use. From what we learned above about the effect smartphones have on our capacities, we should realize now that there is more going on here than the mere opportunity cost of looking at her phone instead of the book. As Eyal explained, when we get hooked on the dopamine surge we get from our phones, we weaken our frontal lobe's executive function, and this makes us worse at sticking to a task like reading.[36] This is only one of the many literacy-related cognitive capacities that is weakened by our phones.[37] The more time we spend engaging with the attention economy, the more our brains prefer shallow content that is less cognitively demanding.

These changes are so drastic that they can even be observed through neuroimaging. In their meta-analysis on this topic, Loh and Kanai (2014) write, "In interrupting the development of deep reading skills, this shift toward shallow information processing may affect brain circuitry necessary for these skills" (516). And in the case of what they call "internet-addicted (IA) individuals" the conclusion is even more emphatic:

> Finally, the rewarding Internet environment also has resulted in the increased prevalence of Internet-related addictive behaviors. IA individuals were worse at inhibiting their responses especially in the face of Internet-related cues and were also highly driven by immediate rewards even in the face of potential losses and uncertainty. These cognitive deficits were further associated

[36] We discussed this above in Sect. 3.2. Cf. Eyal (2014, 9).

[37] As Wolf (2018) explains, reading involves an enormously complex array of neural components, and many of these are endangered by our relationship with technology. See also Carr (2010).

> with alterations in brain networks involved in self-control and reward-processing. (*Ibid*.)

Esther's desire to look at Instagram while she was trying to read a book should come as no surprise. Her brain has adapted (as have ours). She has come to prefer the dopamine-driven content of the attention economy. But this is not a simple case of shifting preferences. This is very different from a change in one's palate.[38] By spending too much time with our phones and the attention economy, we have made ourselves less capable of acting on desires that we still have. We become less capable of setting and pursuing our own ends.

According to the Frankfurt/Dworkin model, Povitsky's first-order desire to check Instagram while reading is inconsistent with her higher-order reflection. She has a second-order desire to act on her first-order desire to read (or, alternatively, she has a second-order desire to be free of the desire to check Instagram while reading). On Watson's characterization, what is distinctive about compulsive behavior is that "the desires and emotions in question are more or less radically independent of the evaluational systems of these agents" (1975, 220). Povitsky's smartphone use is inconsistent with her evaluative judgments about what she ought to be doing, and thus the behavior is compulsive. Finally, her action demonstrates an autonomy deficit on Bratman's model as well. Given what she says, Povitsky, like many of us, would like to read many books over the course of her life and to develop the ability to sit and enjoy reading for long stretches. The action of looking at her phone compulsively is not consistent with her long-term plans.

Given all this information about the seemingly compulsive use of smartphones, it may seem surprising that we have, to some extent, avoided the language of "addiction." In part, we have done this because the topic of addiction remains controversial within the field of psychology. The standard diagnostic tool for mental health professionals is the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (the DSM-5). The language around addiction has shifted considerably over time. The prime example of addiction has long been "substance use disorders" like alcoholism. Substance use disorders were added to the third edition of the

[38] For instance, Enoch (2020) writes, "I used to *just love* quinoa salad, but now it leaves me cold" (169). When our preferences change in this way, it surely does not demonstrate an autonomy deficit.

DSM in 1980. Before that, alcoholism had fallen under the umbrella of personality disorders. For decades, the only "behavioral addiction" that was included in the DSM was "pathological gambling."[39] In 2013, "Internet Gaming Disorder" was added as an appendix to the DSM-5. But no other behavioral "addictions" were included.

Going against the grain, some psychologists have argued that "internet addiction" should be included as well. They point out how similar internet addiction is to both gambling and substance use disorders:

> Control processes are particularly reduced when individuals with Internet addiction are confronted with Internet-related cues representing their first choice use. For example, processing Internet-related cues interferes with working memory performance and decision making. Consistent with this, results from functional neuroimaging and other neuropsychological studies demonstrate that cue-reactivity, craving, and decision making are important concepts for understanding Internet addiction. The findings on reductions in executive control are consistent with other behavioral addictions, such as pathological gambling. They also emphasize the classification of the phenomenon as an addiction, because there are also several similarities with findings in substance dependency… All the findings and clinical implications discussed here have several similarities with other forms of addictive behaviors. They are consistent with neurobiological and psychological models of addiction (Robinson and Berridge 2001; Everitt and Robbins 2005) and with neuropsychological and neuroimaging findings in substance dependency and other forms of behavioral addictions. (Grant et al. 1996; van Holst et al. 2010; Brand et al. 2014, 1, 10)

We do not have a stance on whether or not internet addiction (or smartphone addiction) should be included in the DSM. Obviously, decisions of that kind should be left to the mental health professionals who standardize these terms in order to improve diagnostic and treatment outcomes. This is well outside our domain as moral philosophers.

But we are interested in the psychological literature on addiction because it offers some helpful resources when it comes to separating problematic use from that which is not. After all, many people are capable of gambling or having an occasional drink without developing a clinical

[39] This is the term from the DSM-3. Clinicians no longer use the word "pathological" as it is regarded as pejorative (Petry et al. 2014). It is now referred to as "gambling disorder."

disorder. When assessing whether or not someone's gambling constitutes a disorder, here are two of the questions from the diagnostic screen[40]:

(1_d) Have you tried and not succeeded in stopping, cutting down, or controlling your gambling three or more times in your life?
(2_d) Has your gambling ever caused serious or repeated problems in your relationships with any of your family members or friends? Or, has your gambling ever caused you problems at work or your studies?

As for substance use disorders, the DSM-5 specifies 11 criteria. Here are five that are relevant to smartphone use:

(1_e) Consuming more than intended.
(2_e) Persistent desire to cut down or regulate use.
(3_e) Experiencing craving, a pressing desire to use.
(4_e) Use impairs ability to fulfill major obligations at work, school, or home.
(5_e) Recurrent use in physically unsafe environments.

In both cases, the parallels to problematic smartphone use are clear. First, there are questions about whether or not we have tried to cut down but have struggled to do so ((1_d) and (2_e)). For many of us, this is a familiar experience with our phones.[41] Second, there are questions about whether or not the thing in question interferes with work or school ((2_d) and (4_e)). As we explained above, there is evidence to suggest that smartphones do have a negative effect on academic performance. Third, there are questions about craving (3_e) and about spending more time on these things than we intended (1_e). Smartphone use ticks these boxes as well. Finally, when it comes to using phones in "physically unsafe environments" (5_e), we need look no further than our roads. They are littered with distracted

[40] The diagnostic screen is called the "National Opinion Research Center DSM-IV Screen for Gambling Problems (NODS)." See Petry et al. (2014).

[41] Indeed, there is now an entire cottage industry dedicated to finding ways to cut down on screen time. Tristan Harris started the "Time Well Spent" movement and he now directs a non-profit called "The Center for Humane Technology." A flurry of books have been written on this topic as well with titles like *How to Break Up with Your Phone*; *How to Do Nothing: Resisting the Attention Economy*; and *Stolen Focus: Why You Can't Pay Attention*. We were both inspired by Cal Newport's book *Digital Minimalism*. See Price (2018), Odell (2019), Hari (2022), and Newport (2019).

drivers who look down at their phones in spite of the evidence showing how dangerous this is.[42]

So although we do not wish to take a stand on the clinical and diagnostic issues, we do believe that these tools demonstrate something interesting about our smartphone use. They help us see some of the ways that our devices are at odds with ends we have set for ourselves (e.g., reading books, being able to focus, succeeding in school or at work, not being anxious or depressed, driving safely, etc.). As the polling data show, some of us are already prepared to admit that we have a problem and that we would like to cut back. We would like to have more control over the amount of time we spend on our phones. This shows that our motivational states are not coherent. Much like the unwilling addict that Frankfurt described, we find ourselves doing something that we do not really want to do.

3.3.2.3 Alienated Desires

"The Making of a YouTube Radical" tells the story of Caleb Cain, who described himself as "a liberal college dropout" who was "sucked into a vortex of far-right politics on YouTube." Commenting on Cain's transition, a friend of his said, "I was just, like: 'Wow, what happened? How did you get this way?'" Cain himself, in retrospect says, "I was brainwashed" (Roose 2019).

Assume for a moment that Cain was motivationally coherent after his transformation; that is, when he was—as he used to refer to himself—a "tradcon" (i.e., a traditional conservative). This certainly seems possible. He recounts being "committed to old-fashioned gender norms," dating an evangelical Christian, and fighting with his liberal friends (Ibid.).

[42] Exact data are difficult to come by (as estimates often rely on self-reporting). According to a survey from the US Department of Transportation, there are roughly 660,000 drivers using cellphones while driving at any given moment of daylight hours (NHTSA 2019). Reports from crashes in 2011 indicate that distracted driving contributed to 10% of all fatal crashes. According to the report: "In 2011, 3,331 people were killed in crashes involving distracted drivers and an estimated additional 387,000 were injured in motor vehicle crashes involving distracted drivers" (Ibid.).

Not only does this kind of behavior show that we use our cellphones in situations that are known to be dangerous (a hallmark of addictive behavior), it also demonstrates an important point about our incoherent motivational states. We may have a first-order desire to look at our phones when we get a notification in the car (indeed, most people find it very hard to resist the pull of a notification), but we also have higher-order desires to refrain from acting on any first-order desires that may lead to us being killed or injured in a car accident.

Contrast this with Povitsky or Klein who, while behaving in certain ways—i.e., not reading in the ways they would like to—would be loath to defend those behaviors to their friends. It certainly seems that Cain was more motivationally coherent with respect to his views than they are with respect to their reading habits.

Given that Cain was (at least possibly) motivationally coherent, does this mean that his opinions were held authentically? Not at all. As we see when he refers to himself as "brainwashed," he repudiates the beliefs upon reflection. In such cases, it is especially useful to use Christman's historical test: evaluate whether or not we would have resisted the formation of a desire, belief, or intention if we had reflected on the process that formed it. If the desire, belief, or intention fails the test, then it is one that we are alienated from and, thus, hold inauthentically.

Cain was arguably led to his views via some such process, such as *technological seduction*.[43] As Alfano et al. (2018)—who coined the term "technological seduction"—describe it, seduction occurs in a four-step process:

Signal. The seducer signals to the seducee, "I know what you are thinking" (Alfano et al. 2018, 300)
Affirm. The seducee affirms, "Yes, you do know what I'm thinking" (Ibid., 301)
Suggest. The seducer suggests, "So, let's do …" (Ibid.)
Agree. The seducee agrees.

Alfano et al. (2018) go on to show that this process can be carried out in technologically sophisticated ways. Relevant here is their discussion of *bottom-up technological seduction*, which leverages user data to propose suggestions as to what a user is thinking and what they should do. For instance:

Signal. "[P]redictive analytics will suggest, based on a user's profile and the initial text string they enter, which query they might want to run. For instance, if you type 'why are women' into Google's search bar, you

[43] We should be careful to note here that if something like technological seduction played a role in Caleb's conversion, it was likely a partial role. As we mention below, Cain was "[b]roke and depressed" when he turned to YouTube for help. This suggests that the story of how he went from "liberal college dropout" to "tradcon" is a complex one with many factors.

are likely to see suggested queries such as 'why are women colder than men', 'why are women protesting', and 'why are women so mean'" (308).

Affirm. The user affirms, say, by "selecting why are women colder than men."

Suggest. "[O]ne of Google's top suggestions is a post titled 'Why are Women Always Cold and Men Always Hot'" (309).

Agree. The user then agrees by following the link.

This pattern need not be nefarious. However, as Tristan Harris, quoted in "The Making of a YouTube Radical," says:

> There's a spectrum on YouTube between the calm section—the Walter Cronkite, Carl Sagan part—and Crazytown, where the extreme stuff is … If I'm YouTube and I want you to watch more, I'm always going to steer you toward Crazytown. (Roose 2019)

And this is because, as Guillaume Chaslot—former YouTube engineer, quoted in the same article—states, "it leads to more ads" (Ibid.). Alfano et al. (2021) have been able to empirically confirm certain bits of this hypothesis, showing that "there is a robust pathway [on YouTube] from some seemingly anodyne topics [such as fitness and natural foods] to outright conspiracy theories" (Alfano et al. 2021, 853). As they note, this could be "an effective way to transform people's preferences" and behaviors (Alfano et al. 2021, 838).

Turning back to Cain, one might suspect that he would not have been as receptive to the ideas that ended up seducing him if he had known more about the pathway that he was following. He was in a vulnerable state, given that he was "[b]roke and depressed." He slid quickly from self-help content to conspiracy theories to anti-feminist content to "explicitly racist videos." This wasn't a "natural" path but one that was algorithmically curated to keep him on the site. And the portability of his phone enabled him to expose himself to increasingly radical content, day and night:

> That year, Mr. Cain's YouTube consumption had skyrocketed. He got a job packing boxes at a furniture warehouse, where he would listen to podcasts and watch videos by his favorite YouTube creators all day. He fell asleep to YouTube videos at night, his phone propped up on a pillow. In all, he watched nearly 4,000 YouTube videos in 2016, more than double the number he had watched the previous year. (Roose 2019)

Now, to be sure, we cannot diagnose Cain; we do not know for certain whether he was self-radicalized in the way described above or whether he would have resisted the ideas he accepted were he to better understand the process that served them up to him. But we do think that it is entirely plausible that something like self-radicalization through technological seduction has influenced a great number of users, and some of these users would feel alienated from the views were they to know the full story of how they adopted them. Further, based on Cain's telling of his own story, it is plausible that he is one of them.[44]

Even if this weren't the case, it seems independently plausible that mobile devices play a role in instilling habits, rituals, beliefs, and preferences in us that we would feel alienated from were we to reflect on where they come from. Consider other processes that we have discussed throughout this chapter, some of which are deployed with the explicit aim of instilling "habits" in us: the Hooked cycle, social reciprocity, instant interruption, and so on. These tactics involve getting us to do things, which

[44] Tucker Carlson, the popular right-wing pundit, presents another interesting case of someone who expressed sentiments of alienation when describing how he felt while watching a violent video. Carlson expressed these feelings in a text message that some believe contributed to the series of events that led to his departure from Fox News (Peters et al. 2023). He wrote: "A couple of weeks ago, I was watching video of people fighting on the street in Washington. A group of Trump guys surrounded an Antifa kid and started pounding the living shit out of him. It was three against one, at least. Jumping a guy like that is dishonorable obviously. It's not how white men fight. Yet suddenly I found myself rooting for the mob against the man, hoping they'd hit him harder, kill him. I really wanted them to hurt the kid. I could taste it. Then somewhere deep in my brain, an alarm went off: this isn't good for me. I'm becoming something I don't want to be. The Antifa creep is a human being. Much as I despise what he says and does, much as I'm sure I'd hate him personally if I knew him, I shouldn't gloat over his suffering. I should be bothered by it. I should remember that somewhere somebody probably loves this kid, and would be crushed if he was killed. If I don't care about those things, if I reduce people to their politics, how am I better than he is?" (Peters et al. 2023). In Chap. 7, when we discuss collective autonomy, we will say much more about the kind of content that Carlson was watching. Content that provokes moral outrage spreads very quickly on social media. People enjoy having their worldview confirmed by seeing someone on the other side embarrassed, humiliated, or receiving their comeuppance. Nguyen and Williams (2020) call this "moral outrage porn" and they discuss the many ways that this kind of content makes us worse off. For our present purpose in this section (alienation), the important thing to note is how Carlson came to feel alienated from his desires. He wanted the young man to be badly hurt, but found himself repudiating this desire upon reflection. This kind of radicalization is often the predictable result of consuming too much content from social media. Of course, we are also deeply disturbed by the racism in Carlson's text message; Peters et al. (2023) discuss this as well.

might, like we think technological seduction can, get us to prefer or believe things—and we very well might feel alienated from these beliefs and desires were we to know how we came to have them in the first place.

3.3.2.4 Adaptive Preferences

Earlier, we endorsed an account that holds that a preference is adaptive if it lacks one of the following four features[45]:

- (1_b) the agent reflectively endorsed the preference "at some point in its formation,"
- (2_b) her reflection took place "in the presence of recognized alternatives,"
- (3_b) "at least some of these alternatives were valuable ones," and
- (4_b) "some of these valuable alternative options were live ones (that is, they were ones that X could reasonably see herself exercising, given her current values and ambitions" (Terlazzo 2016, 215).

Here, we will show that at least some of the attitudes we have toward our mobile devices bear the hallmarks of adaptive preferences, and we will do this by exploring the phenomenon of nomophobia (i.e., fear of being away from one's mobile phone).

Let us begin by stepping back and talking about our attachment to our phones. A recent Gallup poll found that over 90% of Americans keep their phones near them "almost all the time during waking hours" and well over 80% "Keep it near at night when sleeping" (Saad 2022). A recent systematic review of the psychological literature on the prevalence of nomophobia found that over half of the studies reviewed (27 out of 53) determined that the rate of nomophobia in the groups they considered was 100% (Jahrami et al. 2022). And nearly all of the studies (45 out of 53) put the rate at over 90% (Ibid.). Putting this all together: it seems to be a safe assumption that many people prefer to have their phones on them.

Might this preference be adaptive? There is a good reason to think that it is.

We can begin by inspecting reasons for thinking that condition (1_b), reflective endorsement, is not met. We began this chapter by talking about some of the motivations that developers have for keeping us on our devices as much as they can and throughout it have disclosed some of the

[45] More precisely, Terlazzo's view is that a preference is adaptive if and only if it fails to meet conditions (2_b) (3_b) or (4_b) or if it fails (1_b) *because* (2_b) (3_b) or (4_b) was not satisfied.

manipulative tactics they use to support the habit of keeping our phones close by. Relevant here is one of Harris's "hijacks" that we have not discussed, Hijack #3: Fear of Missing Something Important. This hijack, which goes hand-in-hand with others—such as instant interruption, putting a slot machine in your pocket, social approval, and social reciprocity, discussed above—works by cultivating the thought that there is a chance that if you are away from an app or device too long, you might miss something important such as "messages, friendships, or potential sexual opportunities" (Harris 2016). As Odell noted above: "Media companies trying to keep up with each other create a kind of 'arms race' of urgency that abuses our attention and leaves us no time to think" (Odell 2019, 59).

It's no wonder that in such an environment, we develop habits and preferences that involve keeping our devices around us at all times. Yet, as we develop these habits and preferences, we don't often stop to think whether we endorse them. This, we think, is often by design. Many of us aren't even aware of the Hooked cycles that are shaping our preferences as we go about our day, responding to seemingly urgent messages or checking up on our accounts to see if there is anything that we have missed. Further, even if we do, it's not clear that we would be able to do much to stop them: being aware that cigarettes contain nicotine does not make it any easier to quit smoking once you have become addicted.

Let us now turn to (2_b), endorsement in the presence of recognized alternatives, (3_b), endorsement in the presence of valuable alternatives, and (4_b), the alternatives are seen as live. Begin with (2_b). We quoted some high school students who did not see getting a smartphone as a choice. Many adults feel the same way too. If you're going to partake in the gig economy, you must have a smartphone. If you want to work in a variety of fields, such as academia or journalism, you might see having a Twitter account as non-optional. Further, if you are able to see not having these as options, it might be hard to see them as valuable. So much happens online, and you might be left out if you aren't part of the conversations happening there.

Quite often, the response to those who are smartphone free or don't have social media accounts is something along the lines of "must be nice." This sort of resignation seems to express the attitude that such an option isn't really on the table *for me*. After all, I'm not independently wealthy, or

stable enough in my career (or whatever it is) for the option to be one that would, at present, add value to my life overall.[46]

As a result, many of us do not see these as live options. It often feels like we don't have a real choice. If you believe that giving up your smartphone would mean an end to your social life, no one would blame you for ruling this out as a live option. So it is no surprise that people want to have their phones on them at all times. But once we understand how this preference was formed in the absence of real alternatives, it is hard to see it as anything other than an adaptive preference. Like those who experience Stockholm Syndrome, we have come to love the devices that are holding us hostage.

3.4 Conclusion

In this chapter, we explored some of the empirical research about the negative effects that smartphones have on us. There is an abundance of evidence in favor of the conclusion that smartphones and the attention economy are undermining our autonomy. Spending too much time on your smartphone weakens your working memory and your executive function. It makes it harder to focus and sustain attention on cognitively demanding tasks. We are frequently interrupted by our phones (prone to both internal and external triggers), and we make more mistakes because of those interruptions. This makes it harder for us to set and pursue our own ends.

Our unhealthy relationship with the attention economy is making us more anxious, depressed, and sleep deprived. It worsens our self-esteem as we constantly compare ourselves to unrealistic representations of how our lives should be. Our actions and desires are being shaped by others. We may want to be the authors of our own life stories, but we often forfeit some of that self-direction as we fiendishly unlock our screens looking for the next hit of dopamine.

We also presented reasons to believe that our problematic relationship with technology was no accident. As those like Parker, Harris, and Eyal explain, tech companies have harnessed the psychology of addiction to engineer these products that we use compulsively. It is in their financial interest for us to maximize the time we spend looking at screens. Some of

[46] Not everyone is in the same position. For instance, Warren Buffett, the billionaire investor, didn't get a smartphone until 2020. He doesn't even use email. His assistant prints relevant emails out for him (Oran 2016). For most of us, living like that would not be an option. Hanin (2021) calls this the "Indispensability thesis".

the successful products from Silicon Valley (like Instagram) came from students of the Stanford Behavior Design Lab, previously known as the Persuasive Technology Lab.[47] Out of charity, we may be inclined to believe that the lab's founder, B.J. Fogg, really does have good intentions. He told an interviewer: "What I always wanted to do was un-enslave people from technology" (Leslie 2016).[48] But he is aware of how his students have used his techniques: "I look at some of my former students and I wonder if they're really trying to make the world better, or just make money" (Ibid.). For the vast majority of products that shape our lives, the answer to this question is fairly obvious.

Once we have an adequate grasp of what smartphones are doing to our brains, our capacities, and our authenticity, we may feel moved to do something about this situation. But what exactly should we do and what kinds of reasons do we have? We turn our attention to those issues in the next few chapters where we will argue that we have a moral duty to restructure our relationship with mobile devices and the attention economy. Given what we learned in this chapter, however, we should recognize that it is likely to be an uphill battle. Unplugging is not merely a question of willpower.

As Tristan Harris points out, "The 'I don't have enough willpower' conversation misses the fact that there are 1000 people on the other side of the screen whose job is to break down the self-regulation that you have" (Singer 2015). But if Kant is right about our moral duties to foster and cultivate autonomy, then we are morally obligated to resist. If we let

[47] Tristan Harris was a student of Fogg's behavior lab (Leslie 2016). Nir Eyal was not formally a student of Fogg's, but he attended a retreat at Fogg's house to learn his methods of persuasion and attended some of Fogg's lectures (see Eyal 2019 and Stolzoff 2018). But there are many notable alumni of Fogg's lab, including Ed Baker, who was the Head of Growth at both Facebook and Uber (Stolzoff 2018).

[48] To Fogg's credit, the Stanford Behavior Lab has a variety of projects that aim to make the world a better place. They are using their behavior modification techniques to address climate change, improve the regulation of negative emotions, and help people flourish. Even Eyal has changed course. Only five years after publishing his best-selling guide to building addictive products, Eyal published *Indistractable: How to Control Your Attention and Choose Your Life*. He claims that there is no inconsistency between his first book and his second. He suggests that *Hooked* was never meant to be used to build addictive products that have a negative effect on users. He wanted the behavior modification techniques to be used for good. He said he was inspired to write *Indistractable* after finding that he was also using his smartphone compulsively in situations where he did not want to use it. "I wrote it because I found that I was becoming distracted by some of these technologies—that I couldn't stop looking at my device even when I was supposed to be spending quality time with my daughter" (Klein 2019).

developers set ends for us and we use products in ways that make us less capable of setting and pursuing our own ends, then Kant would argue that we are failing morally. He believes that our actions should spring from our own autonomy, rather than be driven heteronomously by "foreign impulses" (*G* 4:444).

References

Alfano, Mark, J. Adam Carter, and Marc Cheong. 2018. Technological seduction and self-radicalization. *Journal of the American Philosophical Association* 4: 298–322.

Alfano, Mark, Amir Ebrahimi Fard, J. Adam Carter, Peter Clutton, and Colin Klein. 2021. Technologically scaffolded atypical cognition: The case of YouTube's recommender system. *Synthese* 199: 835–858.

Allcott, Hunt, Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow. 2020. The welfare effects of social media. *American Economic Review* 110: 629–676.

American Psychiatric Association. 2013. *Diagnostic and statistical manual of mental disorders*. 5th ed. Washington, DC: American Psychiatric Association.

Barthorpe, Amber, Lizzy Winstone, Becky Mars, and Paul Moran. 2020. Is social media screen time really associated with poor adolescent mental health? A time use diary study. *Journal of Affective Disorders* 274: 864–870.

Baumgartner, Susanne E., Wouter D. Weeda, Lisa L. van der Heijden, and Mariëtte Huizinga. 2014. The relationship between media multitasking and executive function in early adolescents. *The Journal of Early Adolescence* 34: 1120–1144.

Beland, Louis-Philippe, and Richard Murphy. 2016. Ill communication: Technology, distraction & student performance. *Labour Economics*, SOLE/EALE conference issue 41: 61–76.

Braatz, Kristen. 2022. Is i-Ready dangerous? *Curriculum Associates*. https://www.curriculumassociates.com/blog/is-iready-dangerous. Accessed 22 May 2023.

Brailovskaia, Julia, Verena J. Swarlik, Georg A. Grethe, Holger Schillack, and Jürgen Margraf. 2022. Experimental longitudinal evidence for causal role of social media use and physical activity in COVID-19 burden and mental health. *Journal of Public Health*.

Brand, Matthias, Kimberly S. Young, and Christian Laier. 2014. Prefrontal control and internet addiction: A theoretical model and review of neuropsychological and neuroimaging findings. *Frontiers in Human Neuroscience* 8: 375.

Brighouse, Harry. 2006. *On education*. New York: Routledge.

Brucculieri, Julia 2018. These eerie social media stars aren't even real people. *Huffington Post*. https://www.huffpost.com/entry/some-instagram-influencers-not-real-people_n_5ac4ea12e4b063ce2e580ea5 Accessed 22 May 2023.

Brumby, Duncan P., Anna L. Cox, Jonathan Back, and Sandy J.J. Gould. 2013. Recovering from an interruption: Investigating speed-accuracy trade-offs in task resumption behavior. *Journal of Experimental Psychology: Applied* 19: 95–107.

Cades, David M., Deborah A. Boehm Davis, J. Gregory Trafton, and Christopher A. Monk. 2007. Does the difficulty of an interruption affect our ability to resume? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 51: 234–238.

Cain, Neralie, and Michael Gradisar. 2010. Electronic media use and sleep in school-aged children and adolescents: A review. *Sleep Medicine* 11: 735–742.

Cain, Matthew S., Julia A. Leonard, John D.E. Gabrieli, and Amy S. Finn. 2016. Media multitasking in adolescence. *Psychonomic Bulletin & Review* 23: 1932–1941.

Caird, Jeff K., Kate A. Johnston, Chelsea R. Willness, Mark Asbridge, and Piers Steel. 2014. A meta-analysis of the effects of texting on driving. *Accident; Analysis and Prevention* 71: 311–318.

Carr, Nicholas. 2010. *The shallows: What the internet is doing to our brains.* New York: W. W. Norton & Company.

Carter, Ben, Philippa Rees, Lauren Hale, Darsharna Bhattacharjee, and Mandar S. Paradkar. 2016. Association between portable screen-based media device access or use and sleep outcomes: A systematic review and meta-analysis. *JAMA Pediatrics* 170: 1202–1208.

Castro, Clinton, and Adam Pham. 2020. Is the attention economy noxious? *Philosophers' Imprint* 20: 1–13.

Deloitte. 2018. *Global mobile consumer survey: US edition. A new era in mobile continues.* https://www2.deloitte.com/content/dam/Deloitte/us/Documents/technology-media-telecommunications/us-tmt-global-mobile-consumer-survey-exec-summary-2018.pdf. Accessed 22 May 2023.

Dietz, Stephanie, and Christopher Henrich. 2014. Texting as a distraction to learning in college students. *Computers in Human Behavior* 36: 163–167.

Enoch, David. 2020. False consciousness for liberals, part I: Consent, autonomy, and adaptive preferences. *The Philosophical Review* 129: 159–210.

Everitt, Barry J., and Trevor W. Robbins. 2005. Neural systems of reinforcement for drug addiction: From actions to habits to compulsion. *Nature Neuroscience* 8: 1481–1489.

Eyal, Nir. 2014. *Hooked: How to build habit-forming products.* New York: Penguin.

———. 2019. The science behind how our devices affect our brains and prevent productivity. *Thrive Global.* https://community.thriveglobal.com/the-science-behind-how-our-devices-affect-our-brains-and-prevent-productivity/. Accessed 24 May 2023.

Fabio, Rosa Angela, Alessia Stracuzzi, and Riccardo Lo Faro. 2022. Problematic smartphone use leads to behavioral and cognitive self-control deficits. *International Journal of Environmental Research and Public Health* 19: 1–13.

Fardouly, Jasmine, and Elise Holland. 2018. Social media is not real life: The effect of attaching disclaimer-type labels to idealized social media images on women's body image and mood. *New Media & Society* 20: 4311–4328.

Festinger, L. 1954. A theory of social comparison processes. *Human Relations* 7: 117–140.

Flood, Rebecca. 2022. You can now earn $2k if you quit social media for two months. *Newsweek*. https://www.newsweek.com/earn-2000-quit-social-media-two-months-apps-1676174. Accessed 22 May 2023.

Fox, Annie Beth, Jonathan Rosen, and Mary Crawford. 2009. Distractions, distractions: Does instant messaging affect college students' performance on a concurrent reading comprehension task? *Cyberpsychology & Behavior: The Impact of the Internet, Multimedia and Virtual Reality on Behavior and Society* 12: 51–53.

Grant, S., E.D. London, D.B. Newlin, V.L. Villemagne, X. Liu, C. Contoreggi, R.L. Phillips, A.S. Kimes, and A. Margolin. 1996. Activation of memory circuits during cue-elicited cocaine craving. *Proceedings of the National Academy of Sciences of the United States of America* 93: 12040–12045.

Hadar, A.A., D. Eliraz, A. Lazarovits, U. Alyagon, and A. Zangen. 2015. Using longitudinal exposure to causally link smartphone usage to changes in behavior, cognition and right prefrontal neural activity. *Brain Stimulation: Basic, Translational, and Clinical Research in Neuromodulation* 8: 318.

Hanin, Mark L. 2021. Theorizing digital distraction. *Philosophy & Technology* 34 (2): 395–406.

Hanna, Emily, L. Monique Ward, Rita C. Seabrook, Morgan Jerald, Lauren Reed, Soraya Giaccardi, and Julia R. Lippman. 2017. Contributions of social comparison and self-objectification in mediating associations between Facebook use and emergent adults' psychological well-being. *Cyberpsychology, Behavior and Social Networking* 20: 172–179.

Hari, Jari. 2022. *Stolen focus: Why you can't pay attention – And how to think deeply again*. New York: Crown.

Harris, Tristan. 2013. A call to minimize distraction & respect users' attention. https://digitalwellbeing.org/googles-internal-digital-wellbeing-presentation-transcript-and-slides/. Accessed 17 July 2023.

———. 2016. How technology is hijacking your mind – From a magician and Google design ethicist. *Medium*. https://medium.com/thrive-global/how-technology-hijacks-peoples-minds-from-a-magician-and-google-s-design-ethicist-56d62ef5edf3. Accessed 22 May 2023.

Hebblethwaite, Colm. 2018. Gen Z engaging with 10 hours of online content a day. *Marketing Tech*. https://www.marketingtechnews.net/news/2018/feb/09/gen-z-engaging-10-hours-online-content-day/. Accessed 23 May 2023.

Hunt, Melissa G., Rachel Marx, Courtney Lipson, and Jordyn Young. 2018. No more FOMO: Limiting social media decreases loneliness and depression. *Journal of Social and Clinical Psychology* 37: 751–768.

Jackson, Lauren Michelle. 2018. Shudu gram is a white man's digital projection of real-life black womanhood. *The New Yorker*.

Jacobsen, Wade C., and Renata Forste. 2011. The wired generation: Academic and social outcomes of electronic media use among university students. *Cyberpsychology, Behavior, and Social Networking* 14: 275–280.

Jahrami, Haitham, Khaled Trabelsi, Omar Boukhris, Jumana Hasan Hussain, Ahmad F. Alenezi, Ali Humood, Zahra Saif, Seithikurippu R. Pandi-Perumal, and Mary V. Seeman. 2022. The prevalence of mild, moderate, and severe nomophobia symptoms: A systematic review, meta-analysis, and meta-regression. *Behavioral Sciences* 13: 1–29.

Junco, Reynol. 2012a. In-class multitasking and academic performance. *Computers in Human Behavior* 28: 2236–2243.

———. 2012b. Too much face and not enough books: The relationship between multiple indices of Facebook use and academic performance. *Computers in Human Behavior* 28: 187–198.

Junco, Reynol, and Shelia R. Cotten. 2012. No A 4 U: The relationship between multitasking and academic performance. *Computers & Education* 59: 505–514.

Kant, Immanuel. 2007. *Anthropology, history, and education*, ed. Günter Zöller and Robert B. Louden. Cambridge University Press.

———. 2008. Groundwork of the metaphysics of morals. In *Practical philosophy*. Trans. Mary J. Gregor. New York: Cambridge University Press.

Karpinski, Aryn C., Paul A. Kirschner, Ipek Ozer, Jennifer A. Mellott, and Pius Ochwo. 2013. An exploration of social networking site use, multitasking, and academic performance among United States and European university students. *Computers in Human Behavior* 29: 1182–1192.

Kelly, Yvonne, Afshin Zilanawala, Cara Booker, and Amanda Sacker. 2019. Social media use and adolescent mental health: Findings from the UK millennium cohort study. *EClinicalMedicine* 6: 59–68.

Kirschner, Paul A., and Aryn C. Karpinski. 2010. Facebook® and academic performance. *Computers in Human Behavior* 26: 1237–1245.

Klein, Ezra. 2019. Is big tech addictive? A debate with Nir Eyal. *Vox*. https://www.vox.com/podcasts/2019/8/7/20750214/nir-eyal-tech-addiction-ezra-klein-smartphones-hooked-indistractable. Accessed 1 June 2023.

———. 2022. This is your brain on deep reading. It's pretty magnificent. *The New York Times*.https://www.nytimes.com/2022/11/22/opinion/transcript-ezra-klein-interviews-maryanne-wolf.html. Accessed 1 June 2023.

Kuykendall, Kristal. 2022. New data: Nearly half of schools providing home internet access to students who need it this school year. *The Journal*. https://the-journal.com/articles/2022/09/27/nearly-half-of-schools-providing-home-internet-access-to-students-who-need-it-this-school-year.aspx. Accessed 22 May 2023.

Lambert, Jeffrey, George Barnstable, Eleanor Minter, Jemima Cooper, and Desmond McEwan. 2022. Taking a one-week break from social media improves

well-being, depression, and anxiety: A randomized controlled trial. *Cyberpsychology, Behavior, and Social Networking* 25: 287–293.

Lee, Seungyeon, Myeong W. Kim, Ian M. McDonough, Jessica S. Mendoza, and Min Sung Kim. 2017. The effects of cell phone use and emotion-regulation style on college students' learning. *Applied Cognitive Psychology* 31: 360–366.

Leiva, Luis, Matthias Böhmer, Sven Gehring, and Antonio Krüger. 2012. Back to the app: the costs of mobile application interruptions. *Proceedings of the 14th International Conference on Human-Computer Interaction with Mobile Devices and Services* 12: 291–294.

Leslie, Ian. 2016. The scientists who make apps addictive. *The Economist*. https://www.economist.com/1843/2016/10/20/the-scientists-who-make-apps-addictive. Accessed 24 May 2023.

Levine, Laura E., Bradley M. Waite, and Laura L. Bowman. 2007. Electronic media use, reading, and academic distractibility in college youth. *Cyberpsychology & Behavior: The Impact of the Internet, Multimedia and Virtual Reality on Behavior and Society* 10: 560–566.

Levy, Eilat Chen, Sheizaf Rafaeli, and Yaron Ariel. 2016. The effect of online interruptions on the quality of cognitive performance. *Telematics and Informatics* 33: 1014–1021.

Lim, Julian, and David F. Dinges. 2008. Sleep deprivation and vigilant attention. *Annals of the New York Academy of Sciences* 1129: 305–322.

Loh, Kep Kee, and Ryota Kanai. 2014. Higher media multi-tasking activity is associated with smaller gray-matter density in the anterior cingulate cortex. *PLoS One* 9: e106698.

Mark, Gloria, Stephen Voida, and Armand Cardello. 2012. 'A pace not dictated by electrons': An empirical study of work without email. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*: 555–564.

McLuhan, Marshall, and Lewis H. Lapham. 1994. *Understanding media: The extensions of man*. Reprint ed. Cambridge, MA: MIT Press.

Melcher, Michael. 2013. Mindfulness vs. smartphones. *The Huffington Post*. https://www.huffpost.com/entry/mindfulness-vs-smartphone_b_2856462. Accessed 22 May 2023.

Mettling, Bruno. 2015. *Transformation numérique et vie au travail*. https://www.vie-publique.fr/rapport/35159-transformation-numerique-et-vie-au-travail. Accessed 22 May 2023.

Miljeteig, K., and T. von Soest. 2022. An experience sampling study on the association between social media use and self-esteem. *Journal of Media Psychology: Theories, Methods, and Applications* 34: 373–382.

Mill, John Stuart. 1988. *On liberty*. Indianapolis: Hackett Publishing Company.

Monk, Christopher A. 2004. The effect of frequent versus infrequent interruptions on primary task resumption. *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting*: 295–299.

Monk, Christopher A., J. Gregory Trafton, and Deborah A. Boehm-Davis. 2008. The effect of interruption duration and demand on resuming suspended goals. *Journal of Experimental Psychology: Applied* 14: 299–313.

Newport, Cal. 2019. *Digital minimalism: Choosing a focused life in a noisy world*. New York: Portfolio.

———. 2021a. E-mail is making us miserable. *The New Yorker*. https://www.newyorker.com/tech/annals-of-technology/e-mail-is-making-us-miserable. Accessed 22 May 2023.

———. 2021b. *A world without email: Reimagining work in an age of communication overload*. New York: Portfolio.

Nguyen, Thi. 2020. *Games: Agency as art*. New York: Oxford University Press.

———. 2021. How Twitter gamifies communication. In *Applied epistemology*, ed. Jennifer Lackey, 410–436. New York: Oxford University Press.

Nguyen, C. Thi, and Bekka Williams. 2020. "Moral Outrage Porn." *Journal of Ethics and Social Philosophy* 18 (2). https://doi.org/10.26556/jesp.v18i2.990

NHTSA. National Highway Traffic Safety Administration. 2019. NHTSA survey finds 660,000 drivers using cell phones or manipulating electronic devices while driving at any given daylight moment. https://www.transportation.gov/briefing-room/nhtsa-survey-finds-660000-drivers-using-cell-phones-or-manipulating-electronic-devices. Accessed 1 June 2023.

Odell, Jenny. 2019. *How to do nothing: Resisting the attention economy*. Brooklyn: Melville House.

Oran, Olivia. 2016. Jamie Dimon, Carl Icahn and Warren Buffett all avoid email. *Business Insider*. https://www.businessinsider.com/r-on-wall-street-a-high-ranking-few-still-avoid-email-2016-11#:~:text=Buffett's%20assistant%2C%20Debbie%20Bosanek%2C%20prints,followers%20and%20Buffet%2C%201.2%20million. Accessed 1 June 2023.

Ozimek, Phillip, and Hans-Werner Bierhoff. 2020. All my online-friends are better than me – Three studies about ability-based comparative social media use, self-esteem, and depressive tendencies. *Behaviour & Information Technology* 39: 1110–1123.

Peters, Jeremy W., Michael S. Schmidt and Jim Rutenberg. 2023. Carlson's text that alarmed Fox leaders: 'It's not how white men fight'. *The New York Times*. https://www.nytimes.com/2023/05/02/business/media/tucker-carlson-text-message-white-men.html. Accessed 21 July 2023.

Petry, Nancy M., Carlos Blanco, Marc Auriacombe, Guilherme Borges, Kathleen Bucholz, Thomas J. Crowley, Bridget F. Grant, Deborah S. Hasin, and Charles O'Brien. 2014. An overview of and rationale for changes proposed for pathological gambling in DSM-5. *Journal of Gambling Studies/Co-Sponsored by the National Council on Problem Gambling and Institute for the Study of Gambling and Commercial Gaming* 30 (2): 493–502.

Postman, Niel. 1993. *Technopoly: The surrender of culture to technology*. New York: Random House.

Price, Catherine. 2018. *How to break up with your phone*. New York: Ten Speed Press.

Robinson, Terry E., and Kent C. Berridge. 2001. Incentive-sensitization and addiction. *Addiction* 96: 103–114.

Rocha, Bruno, and Cristina Nunes. 2020. Benefits and damages of the use of touchscreen devices for the development and behavior of children under 5 years old-a systematic review. 2020. *Psicologia, Reflexao e Critica: Revista Semestral Do Departamento de Psicologia Da UFRGS* 33: 24.

Roose, Kevin. 2019. The making of a YouTube radical. *The New York Times*. https://www.nytimes.com/interactive/2019/06/08/technology/youtube-radical.html. Accessed 1 June 2023.

Rosen, Larry D., Alex F. Lim, L. Mark Carrier, and Nancy A. Cheever. 2011. An empirical examination of the educational impact of text message-induced task switching in the classroom: Educational implications and strategies to enhance learning. *Educational Psychology* 17: 163–177.

Rosen, Larry D., L. Mark Carrier, and Nancy A. Cheever. 2013. Facebook and texting made me do it: Media-induced task-switching while studying. *Computers in Human Behavior* 29: 948–958.

Rueda, Jacob. 2021. All Home Connections offering $2,500 to stay off social media. https://www.abc4.com/news/national/att-offering-2500-to-stay-off-social-media/. Accessed 22 May 2023.

Saad, Lydia. 2022. Americans have close but wary bond with their smartphone. *Gallup*. https://news.gallup.com/poll/393785/americans-close-wary-bond-smartphone.aspx. Accessed 23 May 2023.

Sales, Nancy Jo. 2017. *American girls: Social media and the secret lives of teenagers*. Reprint ed. New York: Vintage.

Sana, Faria, Tina Weston, and Nicholas J. Cepeda. 2013. Laptop multitasking hinders classroom learning for both users and nearby peers. *Computers & Education* 62: 24–31.

Sherlock, Mary, and Danielle L. Wagstaff. 2019. Exploring the relationship between frequency of Instagram use, exposure to idealized images, and psychological well-being in women. *Psychology of Popular Media Culture* 8: 482–490.

Singer, Natasha. 2015. Can't put down your device? That's by design. *The New York Times*. https://www.nytimes.com/2015/12/06/technology/personaltech/cant-put-down-your-device-thats-by-design.html?_r=0. Accessed 24 May 2023.

Sohn, Sei Yon, Philippa Rees, Bethany Wildridge, Nicola J. Kalk, and Ben Carter. 2019. Prevalence of problematic smartphone usage and associated mental health outcomes amongst children and young people: A systematic review, meta-Analysis and GRADE of the evidence. *BMC Psychiatry* 19: 356.

Solon, Olivia. 2017a. Ex-Facebook president Sean Parker: site made to exploit human 'vulnerability.' *The Guardian*. https://www.theguardian.com/tech-

nology/2017/nov/09/facebook-sean-parker-vulnerability-brain-psychology. Accessed 22 May 2023.

———. 2017b. 'It's digital colonialism': How Facebook's free internet service has failed its users. *The Guardian*. https://www.theguardian.com/technology/2017/jul/27/facebook-free-basics-developing-markets#:~:text=%E2%80%9CFacebook%20is%20not%20introducing%20people,That's%20digital%20colonialism.%E2%80%9D. Accessed 18 July 2023.

Square, Jonathan. 2018. Is Instagram's newest sensation just another example of cultural appropriation? *Fashionista*. https://fashionista.com/2018/03/computer-generated-models-cultural-appropriation. Accessed 22 May 2023.

Stolzoff, Simone. 2018. The formula for phone addiction might double as a cure. *Wired*. https://www.wired.com/story/phone-addiction-formula/?ref=stanfordreview.org. Accessed 1 June 2023.

Stothart, Cary, Ainsley Mitchum, and Courtney Yehnert. 2015. The attentional cost of receiving a cell phone notification. *Journal of Experimental Psychology: Human Perception and Performance* 41: 893–897.

Terlazzo, Rosa. 2016. Conceptualizing adaptive preferences respectfully: An indirectly substantive account. *Journal of Political Philosophy* 24: 206–26.

Thornton, Bill, Alyson Faires, Maija Robbins, and Eric Rollins. 2014. The mere presence of a cell phone may be distracting: Implications for attention and task performance. *Social Psychology* 45: 479–488.

Tiffany, Kaitlyn. 2019. Lil Miquela and the virtual influencer hype, explained. *Vox*. https://www.vox.com/the-goods/2019/6/3/18647626/instagram-virtual-influencers-lil-miquela-ai-startups. Accessed 22 May 2023.

Twenge, Jean. 2017. *iGen: Why today's super-connected kids are growing up less rebellious, more tolerant, less happy – And completely unprepared for adulthood – And what that means for the rest of us.* New York: Atria Books.

Twenge, Jean M., and Eric Farley. 2021. Not all screen time is created equal: Associations with mental health vary by activity and gender. *Social Psychiatry and Psychiatric Epidemiology* 56: 207–217.

Uncapher, Melina R., Monica K. Thieu, and Anthony D. Wagner. 2016. Media multitasking and memory: Differences in working memory and long-term memory. *Psychonomic Bulletin & Review* 23 (2): 483–490.

van Holst, Ruth J., Wim van den Brink, Dick J. Veltman, and Anna E. Goudriaan. 2010. Why gamblers fail to win: A review of cognitive and neuroimaging findings in pathological gambling. *Neuroscience and Biobehavioral Reviews* 34: 87–107.

Vogel, Erin A., Jason P. Rose, Lindsay R. Roberts, and Katheryn Eckles. 2014. Social comparison, social media, and self-esteem. *Psychology of Popular Media Culture* 3: 206–222.

Ward, Adrian F., Kristen Duke, Ayelet Gneezy, and Maarten W. Bos. 2017. Brain drain: The mere presence of one's own smartphone reduces available cognitive capacity. *Journal of the Association for Consumer Research* 2: 140–154.

Watson, Gary. 1975. Free agency. *The Journal of Philosophy* 72: 205–220.
Willingham, Daniel T. 2015. *Raising kids who read: What parents and teachers can do*, 2015. San Francisco: Jossey-Bass.
Wilmer, Henry H., and Jason M. Chein. 2016. Mobile technology habits: Patterns of association among device usage, intertemporal preference, impulse control, and reward sensitivity. *Psychonomic Bulletin & Review* 23: 1607–1614.
Wilmer, Henry H., Lauren E. Sherman, and Jason M. Chein. 2017. Smartphones and cognition: A review of research exploring the links between mobile technology habits and cognitive functioning. *Frontiers in Psychology* 8: 1–16.
Wolf, Maryanne. 2018. *Reader, come home: The reading brain in a digital world.* New York: Harper.
Wolfe, Wendy L., and Leah Yakabovits. 2022. *I'll see your beautified photo and raise you one: An experimental investigation of the effect of edited social media photo exposure*. Psychology of Popular Media.
Woods, Heather Cleland, and Holly Scott. 2016. #Sleepyteens: Social media use in adolescence is associated with poor sleep quality, anxiety, depression and low self-esteem. *Journal of Adolescence* 51: 41–49.
Wu, Timothy. 2016. *The attention merchants: The epic scramble to get inside our heads.* New York: Random House.

CHAPTER 4

The Duty to Promote Digital Minimalism in Ourselves

"So far from these duties being the lowest, they [duties to oneself] actually take first place, and are the most important of all; for even without first explaining what self-regarding duty is, we may ask, if a man degrades his own person, how can anything else be demanded of him? He who violates duties toward himself, throws away his humanity, and is no longer in a position to perform duties to others."
—Kant (Collins 27:341)

For many people, these distractions can get out of hand, leaving us with a feeling that our decisions are not our own. The fact is, in this day and age, if you are not equipped to manage distraction, your brain will be manipulated by time-wasting diversions … The good news is that we have the unique ability to adapt to such threats. We can take steps right now to retrain and regain our brains. To be blunt, what other choice do we have? We don't have time to wait for regulators to do something, and if you hold your breath waiting for corporations to make their products less distracting, well, you're going to suffocate.
—Nir Eyal, *Indistractable: How to Control Your Attention and Choose Your Life*

4.1 Introduction

Our aim in this chapter is to defend the existence of a moral duty to rebuild our relationship with technology and to be more intentional about how we use it. This is a duty that you owe to yourself. We realize that this

T. Aylsworth, C. Castro, *Kantian Ethics and the Attention Economy*,
https://doi.org/10.1007/978-3-031-45638-1_4

suggestion might provoke immediate resistance. To some readers, the very idea of duties to oneself will seem quite odd—maybe even paradoxical. We typically think of moral obligations as things that we owe to *each other*. It may sound absurd to suggest that you owe something to yourself. After all, we can waive debts when others owe us something, so we should be able to waive whatever obligations we have to ourselves. But we tend to think that moral obligations are *not* something that we can release ourselves from, and this implies that there cannot be any duties to oneself.

An argument along those lines was put forward by Marcus Singer (1959), and many ethicists found it so convincing that, by the end of the twentieth century, duties to oneself had "largely disappeared from the radar of academic philosophers" (Cholbi 2015, 852). The basic idea behind Singer's argument is that duties to oneself rest on a paradox involving three claims. First, duties involve rights. If Iris has a *duty* to refrain from punching Ray, then Ray has a *right* not to be punched by Iris. Second, if Ray has a right against Iris, then Ray can waive that right and release Iris from the obligation. For instance, Ray might agree to enter the boxing ring and thus consent to being punched by Iris. The third claim in the inconsistent triad is that no one can release themselves from an obligation. In order to resist the paradox, one of those three claims has to be rejected.

Within the last decade there has been some pushback against Singer's influential argument. Perhaps the most popular approach for resisting the paradox is to reject the claim that duties to oneself are necessarily connected to waivable rights.[1] One way to do this would be to reject the first claim; we could deny that all duties entail rights. This may be true of *some* moral duties, but not all of them. Alternatively, it could be argued that the second claim is false, that some rights simply cannot be waived.[2] For instance, you might have a moral obligation to refrain from smoking because smoking harms your future self, and you cannot possibly consent to harming yourself in this way. We could imagine that your future self (who has lung cancer) wants you to quit smoking right now. And you

[1] See Muñoz (2022). Muñoz says that this is the most popular response to the paradox; he cites Kading (1959), Wick (1960), Eisenberg (1968), Paton (1990), and Hills (2003). He also attributes this idea to Kant: "So here is the Kantian move. We say Singer is right about *juridical* duties, which really do come with rights and powers of release. There is no such thing as a juridical duty to oneself. But we insist there can still be *non-juridical* duties to oneself, since these do *not* imply rights and powers of release" (Muñoz 2022).

[2] See Schofield (2015).

cannot release yourself from this obligation because you do not presently inhabit the perspective of your future self (Schofield 2015). Other moral philosophers have argued for the surprising conclusion that the third claim is false, that we may be able, in some cases, to release ourselves from a moral obligation (Muñoz 2020, 2021).

But there is at least one ethical tradition in which duties to oneself never went out of fashion.[3] Far from relegating such duties to a second-class status, Immanuel Kant goes so far as to suggest that duties to oneself deserve pride of place, grounding *all* of our moral duties. As Kant sees it, if there were no duties to oneself, "then there would be no duties whatsoever" (*MS* 6:417).[4] Kant even outlined something similar to the paradox of self-release in the *Metaphysics of Morals* about 162 years before Marcus Singer. But unlike Singer, Kant was not troubled by the apparent contradiction.[5] According to Kant's view, our rational agency burdens us with

[3] See Lara Denis (2001).

[4] Kant's argument for this claim is complex. Timmermann (2006) says that it is "as obscure as it is philosophically ambitious" (510). Singer (1959) calls it a "blatant non-sequitur" (138). On Timmerman's reconstruction of the argument, the core claim here is Kant's idea of moral autonomy (which we discussed briefly in Chap. 2). Kant thinks that moral obligation must begin with a recognition of the authority of the moral law and our submission to the law is autonomous precisely because we give the moral law to ourselves. Sanctions of conscience come from within; morality springs from our own practical reason. If we complied with moral rules only because of external punishments and rewards, morality would be, according to Kant, heteronomous. So to have any duties to others (in the Kantian sense) we must begin by recognizing the authority of the moral law as binding on our will. This is a duty to oneself.

[5] There are several reasons for this. First, as Muñoz (2022) points out, Kant would not be troubled by Singer's version of the paradox because he does not believe that all duties entail rights. That would be true only of juridical duties, which Kant calls "duties of right." Juridical duties, which can justifiably be enforced by the state, involve rights that can be waived. But non-juridical duties, which Kant calls "duties of virtue," do not involve legal rights that can be waived. We owe duties of virtue to ourselves in virtue of the fact that we are morally required to respect our own humanity. And there is no right that we can waive to release ourselves from this obligation. The second reason is more complicated. Kant distinguishes between two senses of the human being. He says there is "the human being as a *natural being* that has reason (*homo phaenomeneon*)" and that "same human being thought in terms of his *personality*, that is, as endowed with *inner freedom* (*homo noumenon*)." By distinguishing between two senses of the human being, one as the pure source of obligation and another who experiences morality as a constraint, Kant believes that he avoids the paradox of self-release. He writes, "So the human being (taken in these two different senses) can acknowledge a duty to himself without falling into contradiction (because the concept of a human being is not thought in one and the same sense)" (*MS* 6:418).

obligations that we must respect whether we want to or not. Kant thinks that the value of our autonomy is something that we cannot relinquish, and there is nothing in the world that would permit us to forfeit our rational agency.

As we explained in Chap. 2, we can understand these moral duties in terms of our obligation to respect humanity. We showed why Kant believes that there is something morally significant about the capacity to set and pursue one's own ends. Kant argues that it is morally wrong to undermine someone's capacities, to disrespect their humanity, or to treat them as a mere means. In this chapter, we show how the duty to respect humanity in your own person entails a duty to be a digital minimalist. The empirical evidence from the previous chapter gives us good reason to think that problematic smartphone use undermines autonomy. So if you have an obligation to cultivate your autonomy and protect it from things that threaten to undermine it, then you have a duty to refrain from using mobile devices in such a way. Of course, more needs to be said about the precise content of the duty. What exactly does it mean to be a digital minimalist? We begin in the next section by answering that question. In short, digital minimalism requires us to be thoughtful about how we use mobile devices and to use them in ways that do not interfere with our capacity to set and pursue our own ends. We think of digital minimalism as an end that we ought to pursue. We regard it as a virtue—a robust disposition to what is morally right.

In Sect. 4.3 we provide the Kantian credentials of the duty and we show how it fits in within his broader taxonomy of duties. In the *Metaphysics of Morals*, Kant identifies exactly two ends that we are morally obligated to pursue: our own perfection and the happiness of others. We discuss the latter (e.g., the duty of beneficence) in Chap. 5, but your duty to be a digital minimalist is rooted in the duty to perfect humanity (rational agency) in your own person. As we explain in greater detail below, this means that the obligation should be understood as an imperfect duty rather than a perfect duty. But before we get to that, we must begin with an explanation of digital minimalism.

4.2 Digital Minimalism

To begin our argument, it will be helpful to have a solid grip on the concept of digital minimalism as we understand it. It is easily misunderstood. Cal Newport, who coined the term, defines digital minimalism as follows:

> A philosophy of technology use in which you focus your online time on a small number of carefully selected and optimized activities that strongly support things you value, and then happily miss out on everything else. (Newport 2019, 28)

He says that a *philosophy of technology* is a personal philosophy of technology use that "provides clear answers to the questions of what tools you should use and how you should use them" (Newport 2019, *xiv*). Newport's understanding of digital minimalism outlines an ambitious ideal, but our understanding of the term will differ from his statement in several key ways.

Despite being philosophers, we will avoid understanding digital minimalism as involving a *philosophy*. Instead, we will understand it as a *practice*. That is, we will be less concerned with advocating that users develop clear answers to questions about technology use. Instead, we think it sufficient that users avoid problematic use (i.e., use that conflicts with their ends). So, we will think of digital minimalism in *behavioral* (in contrast to *intellectual*) terms. Some users might find that developing a philosophy is an effective means for sticking with the practice. Indeed, this seems to be what Newport himself thinks. His book is, after all, a work of self-help and not philosophy.

What we then have in mind when we advocate for digital minimalism is a virtue, which—following Aristotle—we conceive of as a habit or stable character trait.[6] The focus on virtue (and Aristotle, for that matter) might seem odd in a book of Kantian ethics, but, as Allen Wood notes, "virtue is at least as important to Kant's ethical theory as it is to any 'virtue ethics'" (Wood 1999, 31). Traditionally, Kant's ethics has been thought of in strictly deontological terms. Most students come away understanding Kant as being focused solely on duties, obligations, and rules. This paints a stark contrast with Mill, who grounds ethics in consequences, and Aristotle, who is principally concerned with character traits that lead to (and are partially constitutive of) flourishing (*eudaimonia*). Recent work has shown how this picture is misleading. As we explained above, Kant *did* care about consequences, and a number of scholars have recently highlighted the many ways that Kant's moral theory is related to virtue ethics.[7]

[6] See *Nicomachean Ethics* II.1-3. Aristotle (2001, 952-56).
[7] See, for example, Denis (2006) and Baxley (2010).

This development in the literature is a welcome improvement in our understanding of Kant's ethics. Older interpretations were often overly influenced by the *Groundwork*. Many readers treated Kant's *first* mature work on moral philosophy as if it were the sole or primary text, definitive of his ethical thought.[8] Recent work has given increased attention to Kant's later moral writings such as the *Metaphysics of Morals* and *Religion within the Boundaries of Mere Reason*. In the *Religion*, Kant defines virtue as "the firmly grounded disposition to fulfill one's duty strictly" (*Rel.* 6:23). This understanding of virtue is helpful for our argument. The duty at stake in our book is grounded in our obligation to respect humanity as an end in itself. As Kant explains, this requires us to cultivate our capacities and to protect them from threats that would undermine our autonomy. In this way, digital minimalism should be thought of as a virtue. It is a stable disposition to fulfill our moral duty.

Understanding digital minimalism as a practice that does not involve developing a philosophy is not our only departure from Newport's official statement. We also take a less demanding approach to digital minimalism than he does. Specifically, our arguments will not require that you "focus your online time on a *small* number of … *optimized* activities" (Newport 2019, 28 emphasis added). We are happy to take a more relaxed approach, where we will understand the underlying duty as one to focus your online time on a *reasonable* number (*or amount*) of activities (*or time*) *in light of your ends*. This, of course, might in some cases require that one follow a more demanding regimen. Perhaps some users need to treat mobile devices or their applications like many of us treat cigarettes or Oreos. Some products simply lend themselves to excessive use.[9]

[8] This is a frequent complaint from Kant scholars who insist that we must read all of Kant's works in order to develop a proper understanding of his ethics. Allen Wood raises this issue in many places. See, for example, the introduction to his translation of the *Groundwork* (Kant 2018).

[9] The details of what the duty looks like in practice will vary from person to person. After reading so much about this topic, both of us tried to cut back our mobile device use by a number of means (changing our phones' screens to black and white, deleting various apps, etc.). But, in the end, we both decided to trade our smartphones in for other, less functional devices. Others may find that they function well enough with less stringent guardrails (deleting social media apps, using app timers, monitoring screen time data, etc.). But this did not work for either of us.

> Putting this all together, we will understand *digital minimalism* as,
>
> Using mobile devices for a number of activities or amount of time that is reasonable, in light of your ends.

Our relaxed conception of digital minimalism might raise the question of whether the view has a right to call itself "minimalism."

In response to this last thought, it will be helpful to contrast digital minimalism with two other habits that run in two opposite directions.[10] We can call one extreme *digital eliminativism*:

> Using mobile devices for a number of activities or amount of time *significantly less than* is reasonable, in light of your ends.

And the other—which we take to be the default—*digital maximalism*:

> Using mobile devices for a number of activities or amount of time *significantly more than* is reasonable, in light of your ends.

The common thread between our conception of digital minimalism and Newport's is the idea that we should take care to use mobile devices in ways that support what we value (or, perhaps, *should* value). This is compatible with using mobile devices to a great degree. Indeed, much of this very book was written on laptops and discussed over video chat, phone, email, and text exchanges in ways that—we think—were perfectly in alignment with our ends. We are not advocating for the elimination of mobile device use. But that is not a reason to think we are not *minimalists*; that is a reason to think we are not *eliminativists*.

Here it is important to note that, as Kantians, we have a particular gloss on what constitutes our ends. Most of our ends are entirely optional. We could commit ourselves to marathon running (or not), baking (or not), and so on. But others are not. For instance, as we have already argued, we are morally required to value autonomy. We will continue to spell out

[10] Newport runs a similar comparison against "Neo-Luddites, who advocate the abandonment of most new technologies" and "Qualified Self Enthusiasts, who carefully integrate digital devices into all aspects of their life with the goal of optimizing their existence" (Newport, 2019, *xiv*).

what this entails in what follows. For now, what is important is that minimalism is not eliminativism, and our brand of minimalism comes with a certain degree of latitude.

This is especially important in light of the fact that mobile devices are an important lifeline for certain vulnerable populations. For instance, despite the fact that mobile devices seem to have particularly negative effects on the young, members of some subgroups might benefit from using them. A recent *New York Times* article—written in response to the surgeon general's warning that "there are ample indicators that social media can … have a profound risk of harm to the mental health and well-being of children and adolescents"—notes that social media might have the opposite effect for some LGBTQ+ youth (Miller 2023).[11] Indeed, Wagaman et al. (2020) found that online activity had a positive effect on how underserved, Southern LGBTQ+ youth feel about their identities. According to one co-author on that study, LGBTQ+ youth frequently describe social media networks in positive terms: "'It's my home,' 'It's my family,' 'It's kept me alive'" (Miller 2023). This must be kept in mind when we make blanket statements, such as the surgeon general's, about the average effect of, say, screens on broader populations (e.g., all youth taken together).

There are members of other populations who likewise receive distinctive benefits from using mobile devices. Indeed, Android and iOS—the two most common operating systems on smartphones—have a suite of features that benefit users with disabilities. They feature screen readers (which, among other things, convert text to audio) and these capabilities can assist the visually impaired in a variety of ways. It has been shown experimentally that these features can be leveraged to assist people with sensory disabilities (such as blindness and hearing loss) in performing daily activities (see, e.g., Lancioni et al. 2017). This, of course, can be a major boon to people for whom completing daily activities is a challenge, as

[11] For the surgeon general's warning, see United States (2023). For the *New York Times* article, see Miller (2023).

being unable to complete the activities can, among other things, affect one's self-confidence and social status (Lancioni et al. 2017).[12]

These two cases help us to highlight the flexibility of digital minimalism. It also helps us foreshadow issues that will arise in the next chapter. Although LGBTQ+ youth may benefit from social media use and certain people with disabilities benefit from using smartphones, this is not the only implication to consider. This point also underscores the need of making these products less harmful. That an LGBTQ+ adolescent needs their social media network is an additional reason to think that the developers of those sites should not leverage that need in order to, say, distract them in excess to what is useful for them. Indeed, as we will soon show, we not only have a duty to ourselves to be digital minimalists, we have duties to lessen the negative effects of mobile device use on others as well.[13]

[12] We are sympathetic to the view of disability that is defended by Elizabeth Barnes in her book, *The Minority Body*. Barnes argues against what she calls the "bad difference" view of disability. That is the view that being disabled should always (or almost always) be regarded as a harm that makes the disabled person worse off than she would be without her disability (even in an ideal society, devoid of ableism). She defends the "mere difference" view, according to which being disabled is a way of being different but not necessarily worse (all things considered). Unless we are ready to ignore the testimony of people who are disabled (or dismiss their claims as adaptive preferences), we must recognize that many of them claim that they do not want their disability to be "cured." They embrace their disabilities and take pride in them. Their disabilities allow them to experience rich goods that they would otherwise lack. This has serious implications for things like the perception of cochlear implants in the Deaf community. Importantly, Barnes is not arguing that there are *no* respects in which a disability could be a harm. She is also not suggesting that there are no persons for whom their disability *is* a bad difference. She agrees, for instance, that being deaf was indeed a harm to Beethoven. The eminently reasonable upshot of her view is that we should make sure that society does not stigmatize disability and that we should provide accommodations for disabilities. There is no doubt that mobile devices (and other technologies) could be useful in this regard. See Barnes (2016).

[13] What's more, this appears to be an adaptive preference. LGBTQ+ youth who find themselves in communities that reject them have no choice but to rely on social media in order to find networks of support. So their reliance on social media is still evidence of an autonomy deficit. If they had other options (such as finding local networks of support and community), then they would not be forced to use social media as their only outlet. Thanks to Archer Amon for helping us make this connection.

This point about the positive effects of mobile devices extends beyond these domains as well. We do not want to overstate our position. We are not suggesting that mobile devices are *always* bad for autonomy or that they undermine autonomy in *every* respect. Although the problematic use of these devices threatens our rational agency, thoughtful use of them can actually *enhance* our capacity to pursue ends. Smartphones have navigation tools that make it easier for us to arrive at our destinations, communication platforms that help us stay in contact with loved ones, weather alerts that keep us prepared for the day, and so on. By using these tools and incorporating them into our agency, we become *more* capable of pursuing certain ends. Ironically, we can even harness their addictive powers for good. For instance, a user who has an autonomous desire to exercise or study German might download an app that uses notifications prompting her to do the activity in question. The app might remind her of a successful streak—harnessing a manipulative tactic to help her pursue an autonomously chosen end.

The danger here is that it is very difficult to avoid the slippery slope. Our relationships with these devices are often decided on terms that are not our own, even when we are pursuing an autonomous end. Someone might begin using social media in order to stay in touch with old friends. This end (and the means to pursue it) was chosen autonomously; it was a product of her rational agency. Over time, however, this autonomous usage might give way to more problematic behavior as she starts to check the app compulsively to see the status of her posts. She begins to use it far more often than she would like to use it. Like Esther, she opens the app while she's trying to read, when she's in class, or even while she's driving. She has fallen prey to manipulative design tactics, and now her relationship with the technology is built on the developers' terms rather than her own.

It is important to remember that the developers' aims are not always consistent with our own. There is an alignment problem here.[14] This applies even in cases where we reflectively endorse our reasons for using the technology in question. Language-learning apps are extremely popular, but critics routinely point out how they almost always fail at developing fluency or even proficiency in the target language.[15] Many of these apps are free, which means that they rely entirely on ad revenue. This, in turn, means that they must dedicate enormous effort to making their apps as addictive as possible. Our aim of learning the language is not perfectly aligned with the company's aim of maximizing our screen time.

After many hours of using Duolingo to learn Italian, journalist David Freedman was unsettled to find just how little he could say or understand in Italy. Freedman then interviewed Luis von Ahn, the co-founder and

[14] In this case, the alignment problem concerns the misalignment between what the tech companies want and what we want. This is a commonly discussed issue in the ethics of artificial intelligence, as many ethicists consider the alignment problem to be one of the most serious worries about AI. In brief, the problem goes like this. Artificial intelligence is very good at finding solutions to well-defined problems. It is fairly simple to train a machine-learning algorithm to identify which pictures have a kitten in them, to identify numbers, etc. One issue with this, however, is that the decision-making procedure is fundamentally obscure—often referred to as a "black box." There is no way for us (or even the developers themselves) to have any idea how the algorithm is making its decisions. It will find a way to accomplish the task, but it cannot tell us how it is doing it. In some cases, we might want to use an algorithm to pursue one end, but it does so by means of a process that we would never have approved. It advances *some* of our aims, but not all of them. For instance, imagine a group of developers who are trying to optimize the disposal of chemical waste at a factory. Without further instructions, the algorithm might determine that dumping all of the waste into the drinking water is the most efficient way to get rid of it. Of course, that would be a disaster. In order to avoid outcomes like this, we must make sure that the AI's aims are aligned with our own. This task can be difficult when trying to get machine-learning systems to represent and incorporate the complex array of human values.

Indeed, the alignment problem is part of the story behind the addictive nature of social media. Researchers at Open AI, Google, and UC Berkeley made this point very clearly: "In ML [machine learning], some platforms maximized clickthrough rates to approximate maximizing enjoyment, but such platforms unintentionally addicted many users and decreased their wellbeing. These cases demonstrate that unintended consequences present a challenging but important problem" (Hendrycks et al. 2021, 10).

[15] For two critical perspectives, see Ravenscraft (2019) and Freedman (2018). For scholarly reviews of the literature on these apps, see Heil et al. (2016) and Kacetl and Klímová (2019).

CEO of Duolingo, and von Ahn was incredibly forthright about these issues. Their aim is to get us hooked: "The biggest problem that people trying to learn a language by themselves face is the motivation to stay with it…That's why we spend a lot of our energy just trying to keep people hooked" (Freedman 2018). Freedman points out how there might be ways of accelerating the learning process and that it could be useful to make the lessons more difficult. But this would be at odds with the aim of getting users to come back for more. And von Ahn admits this to Freeman as well: "We prefer to be more on the addictive side than the fast-learning side…" (Ibid.)

This is not to say that language-learning apps are not helpful in any way. That would overstate the point. Research shows that they are good at building vocabulary, for instance.[16] The point here is simply that we should be cautious about our engagement with addictive technology. Even when we are using it to pursue our autonomously chosen ends, we must be careful to avoid slipping into the kind of compulsive usage that would weaken our ability to set and pursue ends. We must cultivate a robust disposition to use technology in ways that make us more autonomous and refrain from using it in ways that undermine our autonomy. And that is precisely what it means to be a digital minimalist.

4.3 The Kantian Grounds of the Duty

Now that we have a stronger grasp of what it means to be a digital minimalist, we are in a better position to advance our argument. In this section, we will explain why we think there is a moral obligation to be a digital minimalist and what kind of duty this is. At first glance, it may appear that this has already been established or that it follows as a trivial consequence of what was said earlier. In Chap. 2 we defended Kant's claim that autonomy is morally significant. We then showed why Kantians (and other moral theorists) believe that this requires us to refrain from doing things that undermine people's autonomy. For Kant, this is understood in terms of a duty to respect humanity as an end in itself and never treat it as a mere means. Given the empirical evidence presented in Chap. 3, there

[16] See Heil et al. (2016) and Kacetl and Klímová (2019). Although researchers have shown how these apps can build vocabulary, they also show how important it is to integrate this knowledge, contextualize it, and pair it with real interactions in the language.

appears to be a straightforward entailment of the moral obligation to refrain from problematic smartphone use (i.e., to be a digital minimalist).

But the matter is not quite so simple. First, the fundamental obligation to respect humanity leads to different kinds of duties. And they are not equally strict. As Kant sees it, some duties are commanded in *all* circumstances, and you are forbidden from disrespecting the autonomy of rational agents in certain ways. Iago's deceitful manipulation of Othello is a good example of an action that is strictly forbidden. You have a perfect duty to refrain from such behavior. Kant thinks that it is always wrong to perform such an action. He says that perfect duties "admit of no exception in favor of inclination" (*G* 4:421).

In other instances, there are actions that Kant thinks you are morally required to perform, but they are not required in *every* circumstance. These are imperfect duties. Determining when you should perform these actions is a matter of discretion. For instance, Kant claims that we are required to promote the happiness of others—the duty of beneficence. But it is up to each individual to determine whose happiness to promote and to what extent. It would be nice of you to buy a birthday present for your close friend who has been feeling down, but you are not required to buy a birthday present for every person you know. At times you might even wonder whether you should buy a birthday present or donate the money to a charity that fights malaria in sub-Saharan Africa. You must use your moral judgment to decide how to fulfill the duty of beneficence.

This is just one of several important distinctions when it comes to understanding our moral duties concerning our relationship with technology. Thus, in order to get a clearer picture of the duty to be a digital minimalist, it would be helpful to provide more details about Kant's ethical theory and his taxonomy of duties. Kant presents the distinction between imperfect and perfect duties in the *Groundwork* (1785), but he does not put forward a complete division of duties until the publication of the *Metaphysics of Morals* in 1797.

The most fundamental distinction in the *Metaphysics of Morals* is between duties of right and duties of virtue. Duties of right are ones that can justifiably be enforced, whereas duties of virtue cannot. Not only are we morally obligated to refrain from murder and theft, the state can rightly deploy its coercive power to prevent these actions. We discuss duties of right in greater detail in Chap. 6, so we will set them aside for now.

Duties of virtue are then divided between duties to oneself and duties to others. We explore duties to others in Chap. 5 (with a particular

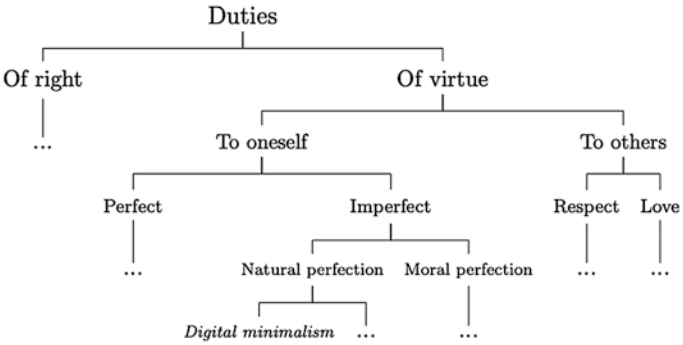

Fig. 4.1 Kant's taxonomy of duties

emphasis on the duty of beneficence). This brings us, at last, to the topic of this chapter: duties to oneself. As previously discussed, these duties are either perfect or imperfect. Finally, they are subdivided once more into duties to ourselves as "rational beings" and duties to ourselves as "animal beings." Duties pertaining to our rational nature require us to pursue "moral perfection" and duties to our animal nature require us to pursue "natural perfection." The entire taxonomy can be represented by Fig. 4.1.[17]

In order to explain the placement of digital minimalism in the taxonomy, we must say more about these last two distinctions.

First, there is the distinction between perfect and imperfect duties. From what was said above, it should already be clear that perfect duties require you to refrain from or perform an action in all circumstances

[17] For brevity, we omitted many of the sub-branches. For instance, we did not list particular duties of love that we owe to others (e.g., beneficence, gratitude, sympathy). We used ellipses to indicate the missing information. We would also like to issue an important caveat. Although our account is broadly Kantian, we are certainly not committed to *all* of Kant's views about particular duties. When Kant discusses duties of natural perfection, he defends duties to refrain from suicide, gluttony, drunkenness, and (perhaps most notoriously) "defiling oneself by lust" (*MS* 6:424). Obviously, we are not adopting his views wholesale. For our purposes, the most useful elements of his moral theory are (1) the absolute value of rational agency (humanity), (2) the humanity formulation of the categorical imperative, and (3) the taxonomy of duties.

Although this diagram can be constructed fairly easily from a careful reading of the *Metaphysics of Morals*, we should acknowledge a debt to Allen Wood for constructing a similar diagram. See Wood (1999, 324).

without any exceptions. Imperfect duties are not so strict. You are not called to perform such actions in every possible moment. But even with this understanding, more needs to be said about what grounds the distinction. How can we distinguish imperfect duties from perfect duties? At first, it would appear that we can tell the two apart by using the formula of universal law and finding different kinds of contradictions. In the *Groundwork*, Kant identifies one perfect duty to the self (refraining from suicide) and one imperfect duty (cultivating talents). He also explains one perfect duty to others (refraining from making a false promise to the banker)[18] and one imperfect duty (beneficence).

He first derives the duties by means of the formula of universal law and he then returns to them once more to show how they can be derived from the formula of humanity. For both duties, when using the formula of universal law, Kant shows why it would be impossible to will these maxims as universal laws. We can show why these actions are impermissible by demonstrating how this leads to a contradiction. If you cannot will the maxim to be a universal law, then you are acknowledging that you are making an exception of yourself when you act that way. You want everyone else to abide by one set of rules while you abide by another. This would be like the child in Chap. 2 who takes from others while insisting that no one take her toys. This behavior demonstrates an inconsistency, and Kant thinks this allows us to see why such actions are morally wrong.

But the formula leads to different kinds of contradictions. When we try to imagine a world where *everyone* makes false promises about repaying loans, we can immediately see why this leads to a contradiction. In such a world, there would be no banks at all. Banks depend on people repaying loans. Indeed, the very institution of promise-making depends on the other person accepting the promise and believing that you will fulfill your commitment. In a world where no one keeps their promises, there would be no promises at all. So we are trying to imagine what it would be like to

[18] In the *Groundwork*, it is fairly clear that Kant wanted to separate the duties in this way: two duties to the self and two duties to others. But this picture is complicated somewhat by the rather surprising fact that Kant identifies "lying" as a violation of a duty to *oneself* in the *Metaphysics of Morals*. See *MS* 6:429.

make a promise in a world where *there are no promises*. This is a contradiction in conception. Such a world is inconceivable.[19]

He contrasts this with the kind of contradiction that results when we try to universalize the maxim of non-beneficence. Can we imagine a world where no one ever helps anyone who is in need? Kant thinks we can. So it does not lead to a contradiction in conception. Instead, Kant suggests that there is a "contradiction in the will." Although we can imagine such a world existing, we cannot coherently will to live in such a world. In the case of beneficence, Kant suggests that there are times where everyone finds themselves in need of help from others. And if this is true, then you cannot coherently will to live in a world where no one helps anyone else.[20]

When it comes to understanding the distinction between perfect and imperfect duties, this is a good place to start. By using the formula of universal law, we can see how perfect duties arise from contradictions in

[19] There are different interpretations of the contradiction in question. Korsgaard (1996) explores three different possibilities. First, there is the "logical contradiction" reading, which suggests that universalized maxim must lead to a logical contradiction. She points out how this might work for certain actions that rely on social conventions (such as promises), but that it works less well for actions like killing. She then discusses "teleological" contradictions. These involve a conflict between (1) the natural end or purpose of some instinct or action and (2) the maxim of the immoral action. This interpretation fares well in the case of the suicide example (where Kant sets up a conflict between suicide and the drive of self-preservation that is contained in what he calls "self-love"). But it does less well with situations like the false promise. Korsgaard favors the "practical contradiction" view. According to this interpretation, the contradiction in question arises when the universalized maxim makes it impossible for the agent to achieve her end, particularly the end specified by her maxim.

[20] This seems to lead to an immediate objection. By making the test depend on what the agent wills, it appears to open it up to a dangerous form of relativism. What would Kant say to a miser who insists that he never wants or needs help from anyone else? Can such a person coherently will the maxim of universal non-beneficence? There are a few options here. We would like to suggest, as Korsgaard does, that there are certain ends that are necessarily connected to all human wills. She writes: "[W]e must find some purpose or purposes which belong essentially to the will, and in the world where maxims that fail these tests are universal law, these essential purposes will be thwarted, because the means of achieving them will be unavailable. Examples of purposes that might be thought to be essential to the will are its general effectiveness in the pursuit of its ends, and its freedom to adopt and pursue new ends" (1996, 96). Kant was reluctant to wed his ethics to anything in human nature as he worried that such an ethics could never yield categorical prescriptions. There is simply too much variance across human beings. But this suggestion looks more plausible if it is maximally thin. For instance, the very idea of a will contains the concept of setting and pursuing ends. This means that it almost analytically contains the idea of being capable of pursuing a variety of ends (which is why Kant thinks we are obligated to cultivate our capacities).

conception and imperfect ones stem from contradictions in the will. But this is not the end of the story.[21] In the *Metaphysics of Morals*, Kant elaborates considerably on the distinction. One thing that is particularly distinctive about Kant's treatment of duties in the doctrine of virtue is his discussion of "ends that are also duties." Indeed, Kant frames the entire section in terms of this idea. As we noted earlier, Kant says that there are only two such ends: one's own perfection and the happiness of others.

This opens up a new way of understanding the distinction between perfect and imperfect duties. You fulfill an imperfect duty whenever your action promotes an end that you have a duty to set (such as the happiness of others). For instance, you act on an imperfect duty when you buy your friend a birthday present or when you make a donation to a charity that treats malaria. You are morally obligated to adopt the end of the happiness of others, and such actions clearly promote that aim. You violate a perfect duty, however, when you set an end that is fundamentally inconsistent with an end that you are required to adopt. So although you are not morally required to get your least favorite colleague a birthday present, it would be strictly forbidden to play a prank on him by putting a snapping mousetrap in his mailbox.

Allen Wood summarizes the distinction in the following way:

> A duty *d* is a duty toward (*gegen*) *S* if and only if *S* is a finite rational being and the requirement to comply with *d* is grounded on the requirement to respect humanity in the person of *S*. A duty is wide or imperfect (or, if toward others, a duty of love) if the action promotes a duty of virtue (that is, an end it is a duty to set); an act is required by a strict, narrow or perfect duty (or a duty of respect to others) if the failure to perform it would amount to a failure to set this obligatory end at all, or a failure to respect humanity as an end in someone's person. An act violates a perfect duty (or

[21] Some commentators, like Simon Hope, take the *Groundwork* to be the definitive text on this issue. See Hope (2023). Hope disagrees with the view, defended by Wood and others, that imperfect duties should be understood in terms of an obligation to adopt an end. Cf. Wood (2009). Hope does not mention Wood by name, but he makes the objection quite clear: "One reason it would be inappropriate is that there is a clear sense in which principles of both perfect and imperfect duty are ends: maxims of duty feature in practical thought as time-general ends, standards that orient one's ongoing and open-ended activity as that plays out across an indeterminate number of possibilities for interaction with an indeterminate domain of agents. In this regard any perfect duty, just as much as any imperfect duty, is a duty to adopt an end that does not come to completion in any specific pattern of action" (Hope 2023, 71). As we explain in this chapter, the view we hold is closer to Wood's.

> duty of respect) if it sets an end contrary to one of the ends it is our duty to set, or if it shows disrespect toward humanity in someone's person (as by using the person as a mere means). (Wood 2009, 233)

When you put a mousetrap in the mailbox, the aim of your action is your colleague's suffering. You have set an end that is inconsistent with an end that you are obligated to set (*viz.* the happiness of others).

Wood's interpretation is a helpful conceptual tool for understanding the duty to be a digital minimalist. According to our view, digital minimalism is a disposition that you have an imperfect duty to cultivate because it promotes an end that you are morally required to adopt. Digital minimalism falls under the umbrella of the duty to promote your own natural perfection.

Kant singled out only a single obligatory end for duties to oneself: one's own perfection. But he believes that we owe different things to ourselves insofar as we are both rational and animal beings. As rational beings, the duty of perfection manifests itself as the requirement to be morally perfect. This requires total purity in our motives. Kant tells us that we should act from the motive of duty; we must do what is right simply because it is right and not because it serves our own self-interest. The command here is to "be holy" (*MS* 6:446). We are also required to be in total compliance with every moral duty. Kant says this is the command to "be perfect" (Ibid.). Of course, Kant recognizes that it is impossible for any human being to live up to this standard. Given the nature of human frailty, we can never expect to be morally perfect (in this life)[22] or to have entirely pure motives. We can only "*strive* for this perfection," so he says that this duty consists "only in continual progress" (Ibid.). It requires us to know ourselves deeply and to subject our actions and motives to the relentless scrutiny of conscience.

As animal beings, Kant understands perfection as the ability to pursue whatever ends we might set for ourselves. He says, "*Natural* perfection is the *cultivation* of any *capacities* whatever for furthering ends set forth by reason" (*MS* 6:391). Here again he defines humanity as the capacity to "set oneself an end" and he says that this is what separates us from nonhuman animals. He then argues that we ought to make ourselves worthy of this capacity by respecting it as the precious treasure that it is. If we let our talents rust and refuse to cultivate any of our abilities, we would

[22] Kant adds this parenthetical aside as a nod to his doctrine of the highest good.

express disrespect for humanity in our own person. What is the point of having the ability to set ends through reason if we render ourselves incapable of pursuing any of those ends? We would end up like Tantalus, forever doomed to desire things that are just out of reach.

In order to demonstrate respect for your capacity to set and pursue ends, Kant thinks that you must make yourself capable of pursuing "all sorts of possible ends" (*MS* 6:444):

> Hence there is also bound up with the end of humanity in our own person the rational will, and so the duty, to make ourselves worthy of humanity by culture in general, by procuring or promoting the *capacity* to realize all sorts of possible ends, so far as this is to be found in the human being himself. In other words, the human being has a duty to cultivate the crude predispositions of his nature, by which the animal is first raised into the human being. It is therefore a duty in itself. (*MS* 6:392)

As he explains later, this requires us to cultivate the "powers of spirit" and the "powers of the soul." He says that the latter includes "memory, imagination and the like" (*MS* 6:445).[23] As we argued in Chap. 2, humanity (understood as the capacity to set and pursue our own ends) is something that matters morally. Kant's ethics requires us to express respect for this capacity in both ourselves and in others. We disrespect the humanity of others by manipulating them, deceiving them, or making them incapable of pursuing their own ends. But we disrespect our *own* humanity if we neglect to cultivate our own capacities or if we fail to protect them from threats that could undermine our autonomy.[24]

With all of this in mind, we are in a much better position to understand how digital minimalism fits in within the broader framework of Kant's ethics. Digital minimalism can be understood as a virtue insofar as it promotes the end of our natural perfection. By refraining from compulsive, heteronomous use of mobile devices and social media, you are protecting

[23] He then concludes by noting that we should care for our bodies as well. He says we must look after "the *basic stuff* (the matter) in a human being" (*MS* 6:445). It is a duty to invigorate the animal in ourselves, as this also plays a role in our ability to realize our ends. Thus, Kant would have seen himself as fulfilling an imperfect duty each morning as he took his daily constitutional through the streets of Königsberg.

[24] The duty to respect humanity in our own person by resisting technological heteronomy is similar to what Carol Hay refers to as the "duty to resist oppression." See Hay 2011. We say more about this similarity in Chap. 8.

your ability to be the author of your own life story. As a digital minimalist, you are refusing to let developers and tech companies determine how much time you spend looking at screens because you value your memory, attention, executive function, self-esteem, and authenticity.

But this is not the only way of understanding the duty. Stefano Lo Re argues for a similar conclusion, which, in our terms, can be understood as a negative duty to refrain from what we above call digital maximalism. In his response to our first paper on this topic,[25] Lo Re argues against our conception of digital minimalism as an imperfect duty, and he defends the existence of a *perfect* duty against "screen overexposure" (2022, 501).[26] He also characterizes it as a duty to promote one's own natural perfection, but he sees it as a duty of strict omission on the grounds that screen overexposure is inconsistent with the duty to make ourselves capable of pursuing a variety of ends. In spite of this apparent contradiction, we believe that his view is actually compatible with ours.

Allen Wood's interpretation of the distinction makes it possible for us to understand how the duty in question could be seen as perfect ("refrain from maximalism") or imperfect ("engage in minimalism"). For every obligatory end (one's own perfection or the happiness of others) there are both perfect and imperfect duties. In our view, the duty to be a digital minimalist is understood as an obligation to cultivate a disposition that *promotes* an end that we are morally required to adopt. In those terms, it should be characterized as an imperfect duty. If, however, the duty is described as a duty to *refrain* from setting an end that is *contrary* to our obligatory end (as Lo Re has done by describing it as a duty to refrain from screen overexposure), then it should be understood as a perfect duty.

[25] See Aylsworth and Castro (2021).

[26] When explaining how his view differs from ours, Lo Re raises a few other issues with our account. He also points out that screen overexposure is broader than our focus on smartphones and mobile devices. This is certainly true, and we explained our reasons for this restricted focus in Chap. 3. Secondly, and more importantly, Lo Re suggests that we are "less concerned with justifying the Kantian credentials" of our account (Lo Re 2022, 501). As evidence of this, he cites our reliance on the concept of autonomy, and he points out how Kant says very little about "autonomy" in the *Metaphysics of Morals*. This is certainly true. The notion of moral autonomy plays a very minor role in the *Metaphysics of Morals*, but the concept of *humanity* looms quite large. Indeed, humanity (which we understand as Kant's concept of personal autonomy) lies at the heart of every duty of virtue, including the duties of natural perfection that ground Lo Re's duty against screen overexposure. In Chap. 2 we took greater care with the distinction between "moral autonomy" and "personal autonomy" (i.e., humanity) and we spent more time explaining our use of those terms.

These are nearly two sides of the same coin; there are only minor differences in the details. The negative duty to refrain from screen overexposure (or digital maximalism) is perfect and thus can never be violated. In this sense, it might seem more demanding than what we focus on here. However, our imperfect duty is *positive*, and, perhaps, in asking us to *do* something, it asks more of us in other respects. In any event, there will be more convergence than divergence between these two accounts. Both recommend that we refrain from being on our screens too much because it is in tension with our duty to set our natural perfection as one of our ends. Lo Re is climbing the same mountain we are, even if he is on a different path to the summit.

4.4 Conclusion

Before moving on to other arguments about technology (duties to others, duties of right, etc.) it would be a good idea to take stock of everything that has been said in this chapter and summarize the argument. The first premise of the argument was established in Chap. 2. We argued for the moral significance of autonomy, and we defended Kant's claim that humanity (the rational capacity to set and pursue your own ends) has a distinctive kind of value—dignity. We then explained Kant's view about the moral requirements that stem from our recognition of the value of humanity. We ought to respect it as an end and never treat it as a mere means.

In Chap. 3, we discussed the empirical research that gives us reasons to believe that having an unhealthy relationship with our mobile devices is bad for our autonomy. In this chapter, we explained the concept of digital minimalism, and we showed how it fits within the broader framework of Kant's ethics. If we value our rational agency, then we must adopt the end of our own natural perfection. One component of fulfilling that project involves a robust disposition to use technology intentionally and in ways that do not conflict with our ability to set and pursue our own ends.

In sum, the argument can be summarized as follows:

P1$_a$ Humanity (i.e., the rational capacity to set and pursue one's own ends) has an objective, unconditional, non-fungible value—dignity.

P2$_a$ Anything that has dignity ought to be respected as an end and never treated as a mere means.

P3$_a$ If humanity ought to be respected as an end and never treated as a mere means, then we ought to adopt the end of our own natural perfection.

Therefore,
$C1_a$ We ought to adopt the end of our own natural perfection (i.e., we should make ourselves capable of pursuing a variety of ends and we should protect our rational agency from things that threaten to undermine it).
$P4_a$ Mobile device use threatens to undermine our capacities of rational agency and our authenticity. The compulsive use of such devices is inconsistent with the end of natural perfection.
$P5_a$ If $P4_a$ and $C1_a$ are true, then we ought to cultivate a disposition to use mobile devices intentionally and in ways that do not undermine our capacities (i.e., we ought to be digital minimalists).
Therefore,
$C2_a$ We ought to be digital minimalists.

It might go without saying, but it should be noted that the "oughts" in the argument are moral "oughts." They are not merely prudential. The empirical literature on mobile devices seems more than sufficient to generate prudential reasons for being more intentional about our relationships with these devices. There is already an overabundant self-help literature, teeming with books about the prudential reasons we have to "detox" from these devices and strategies for accomplishing this aim.[27] We share many of their concerns; we agree that people could promote their self-interest by unplugging from their mobile devices. But Kantian ethics offers even more compelling reasons to adopt the end of digital minimalism.[28] If

[27] We mentioned some of these in the previous chapter. See Price (2018), Odell (2020), Hari (2022), Zahariades (2018), Williams and White (2020), Lanier (2018), and Newport (2019, 2021).

[28] For Kantian ethics, it is very important to avoid conflating duties to yourself with the promotion of your own self-interest. Kant insists that duties to the self do not stem from an obligation to promote one's own happiness (which he considers an absurdity). It is true that he calls this an "indirect duty" in the *Groundwork* on the grounds that we would be tempted by transgressions of duty if we were to utterly neglect our own happiness (4:399). But he thinks that every individual naturally looks after their own happiness, so he thinks that it could not possibly be considered a duty. See *MS* 6:386.

In his lectures on ethics, he also points out how a duty to promote your own happiness would conflict with duties that you owe to others: "It all comes of the fact that people have had no pure concept on which to base a duty to oneself. The thought has been that self-regarding duty consists in promoting one's own happiness, as Wolff also defined it; it now depends on how everyone determines his happiness, and then duty to oneself would consist of a general rule directing us to satisfy all our inclinations and promote our own happiness. This would, however, be a great hindrance thereafter to our duty to others. It is by no means the principle, though, of self-regarding duties, and the latter have nothing to do with well-being and our temporal happiness" (Collins 27:341).

autonomy matters morally, as we believe it does, then we also have moral reasons to be digital minimalists.

In Chap. 2, we briefly explained why other moral theories are also committed to the moral significance of autonomy. Consequentialists may see autonomy as partially constitutive of well-being (or as instrumentally valuable for promoting it). Virtue ethicists and perfectionist theories are likely to include autonomy as a component of human flourishing (as would many ethical egoists). Contractualists ground their moral framework in the idea that we must respect the equal moral status of autonomous agents. And so on. In general, we think that any plausible moral theory will make *some* room for the importance of autonomy.

Any ethical theory that assigns moral weight to autonomy is therefore capable of concluding that there are moral reasons to be a digital minimalist. Of course, the accounts will vary. For instance, the consequentialists who include autonomy in their theory of the good must weigh the promotion of this good against other effects. They will consider digital minimalism a duty if and only if this action (or set of actions) is optimific (i.e., it produces better consequences than any of the alternatives). But we have tried to characterize digital minimalism as a disposition that is variable enough to accommodate this concern. It is not "one size fits all." The application of the duty will vary from person to person, as the aim is to maximize each person's ability to pursue her own ends effectively. Every individual must find the appropriate balance so that they are using mobile devices in ways that promote their ends while refraining from usage that is antithetical to them (e.g., causes depression, anxiety, sleep disturbance, self-esteem issues, etc.). Given that the duty is framed in this way, we believe it is likely to maximize the good (as consequentialists understand it).

Even consequentialists who do *not* value autonomy can find moral reasons to be a digital minimalist. Those who hold monistic theories of value—like hedonism—are likely to acknowledge that our current relationship with mobile devices does not seem to be maximizing overall or aggregate pleasure. By letting ourselves default into a relationship with technology that was not built on our own terms, many people are using them in ways that are not maximizing their well-being.

Naturally, Kantian ethics is stricter about the priority of certain values. If we were to translate Kant's ethics into consequentialist language, we would say that the value of rational agency (humanity) is lexically prior to

values like pleasure.[29] So even a user who derives pleasure from a device would have a moral reason to restructure her relationship with it if she discovers that it is undermining her rational agency.[30]

Although we could have chosen other moral theories to advance our arguments about mobile devices, we believe that Kantian ethics offers the most perspicuous account of what is troubling about this relationship with technology. By foregrounding the concern with autonomy, we can bring the issue into clearer focus. As Kant sees it, humanity is the most precious commodity in our possession. It is in virtue of our humanity that we are capable of morality, the very thing that gives value and meaning to our lives.[31] To wantonly forfeit some of our agency by falling prey to technological heteronomy is to demonstrate a failure to respect this precious capacity as the treasure that it is.

References

Aristotle. 2001. Nicomachean ethics. In *The basic works of Aristotle*. New York: Random House.

Aylsworth, Timothy. 2020. Bolstering the keystone: Kant on the incomprehensibility of freedom. *Archiv für Geschichte der Philosophie* 102: 261–298.

Aylsworth, Timothy, and Clinton Castro. 2021. Is there a duty to be a digital minimalist? *Journal of Applied Philosophy* 38: 662–673.

Barnes, Elizabeth. 2016. *The minority body: A theory of disability*. New York: Oxford University Press.

[29] Some have even developed such accounts. See Cumminsky's "Kantian Consequentialism" (1990).

[30] Although we would prefer to avoid defending all of Kant's views about duties to oneself (including suicide), it would be useful to draw on his treatment of it in order to better understand this point. In Kant's discussion of the suicide example, he says that man's suicidal ideation is driven by the thought that his future life will have more pain than pleasure: "From self-love, I make it my principle to shorten my life when its longer duration threatens more troubles than agreeableness" (*G* 4:422). But given Kant's commitment to the value of rational agency, discarding one's humanity in order to avoid pain is a bad trade-off. For a Kantian discussion of suicide, see Velleman (1999).

[31] Kant's view is that without rational agency (and thus without freedom), "earth would be bereft of value, a meaningless rock floating through the universe" (Aylsworth 2020). In a lecture on ethics, he says "If all creatures had a faculty of choice bound by sensuous impulses, the world would have no value; the inner value of the world, the *summum bonum*, is the freedom to act in accordance with a faculty of choice that is not necessitated. Freedom is therefore the inner value of the world" (VE 27:1482). For more on this point, see Aylsworth (2020). Cf. Guyer (2000, 96).

Baxley, Anne Margaret. 2010. *Kant's theory of virtue: The value of autocracy*. New York: Cambridge University Press.

Cholbi, Michael. 2015. On Marcus Singer's 'On duties to oneself'. *Ethics* 125: 851–853.

Cumminsky, David. 1990. Kantian consequentialism. *Ethics* 100: 586–615.

Denis, Lara. 2001. *Moral self-regard duties to oneself in Kant's moral theory*. New York: Routledge.

———. 2006. Kant's conception of virtue. In *Cambridge companion to Kant and modern philosophy*, ed. Paul Guyer. New York: Cambridge University Press.

Eisenberg, Paul. 1968. Duties to oneself and the concept of morality. *Inquiry: An Interdisciplinary Journal of Philosophy* 11: 129–154.

Freedman, David H. 2018. How to almost learn Italian. *The Atlantic*. https://www.theatlantic.com/magazine/archive/2018/12/language-apps-duolingo/573919/. Accessed 9 June 2023.

Guyer, Paul. 2000. *Freedom, law, and happiness*. Cambridge: New York.

Hari, Jari. 2022. *Stolen focus: Why you can't pay attention – And how to think deeply again*. New York: Crown.

Hay, Carol. 2011. The obligation to resist oppression. *Journal of Social Philosophy* 42: 21–45.

Heil, Catherine Regina, Jason S. Wu, Joey J. Lee, and Torben Schmidt. 2016. A review of mobile language learning applications: Trends, challenges, and opportunities. *The EuroCALL Review* 24: 32.

Hendrycks, Dan, Nicholas Carlini, John Schulman, and Jacob Steinhardt. 2021. Unsolved problems in ML safety. *ArXiv E-Prints* 2109.13916. https://ui.adsabs.harvard.edu/abs/2021arXiv210913916H/abstract. Accessed 9 June 2023.

Hills, Alison. 2003. Duties and duties to the self. *American Philosophical Quarterly* 40: 131–142.

Hope, Simon. 2023. Perfect and imperfect duty: Unpacking Kant's complex distinction. *Kantian Review* 28: 63–80.

Kacetl, Jaroslav, and Blanka Klímová. 2019. Use of smartphone applications in English language learning—A challenge for foreign language education. *Education Sciences* 9: 179.

Kading, Daniel. 1959. Are there really 'no duties to oneself'? *Ethics* 70: 155–157.

Kant, Immanuel. 1997. *Lectures on ethics*. Trans. Peter. Heath, ed. Peter Heath and J.B. Schneewind. New York: Cambridge University Press.

———. 2005. Religion within the boundaries of mere reason. In *Religion and rational theology*. Trans. Allen Wood and George di Giovanni. New York: Cambridge University Press.

———. 2008a. *Groundwork of the metaphysics of morals*. In *Practical philosophy* Trans. Mary J. Gregor. New York: Cambridge University Press.

———. 2008b. The metaphysics of morals. In *Practical philosophy* Trans. Mary J. Gregor. New York: Cambridge University Press.
———. 2018. *Groundwork for the metaphysics of morals: With an updated translation, introduction, and notes*, ed. Allen Wood. New Haven: Yale University Press.
Korsgaard, Christine. 1996. *Creating the kingdom of ends*. New York: Cambridge University Press.
Lancioni, Giulio E., Nirbhay N. Singh, Mark F. O'Reilly, Jeff Sigafoos, Gloria Alberti, Carmen Zimbaro, and Valeria Chiariello. 2017. Using smartphones to help people with intellectual and sensory disabilities perform daily activities. *Frontiers in Public Health* 5: 282.
Lanier, Jaron. 2018. *Ten arguments for deleting your social media accounts right now*. 1st ed. New York: Henry Holt and Co.
Lo Re, Stefano. 2022. The glowing screen before me and the moral law within me: A Kantian duty against screen overexposure. *Res Publica* 28: 491–511.
Miller, Claire Cain. 2023. For one group of teenagers, social media seems a clear net benefit. *The New York Times*. https://www.nytimes.com/2023/05/24/upshot/social-media-lgbtq-benefits.html. Accessed 21 July 2023.
Muñoz, Daniel. 2020. The Paradox of duties to oneself. *Australasian Journal of Philosophy* 98: 691–702.
———. 2021. From rights to prerogatives. *Philosophy and Phenomenological Research* 102: 608–623.
———. 2022. Obligations to oneself. *The Stanford Encyclopedia of Philosophy*. Summer 2022 Edition, ed. Edward N. Zalta. https://plato.stanford.edu/archives/sum2022/entries/self-obligations/. Accessed 6 June 2023.
Newport, Cal. 2019. *Digital minimalism: Choosing a focused life in a noisy world*. New York: Portfolio.
———. 2021. *A world without email*. New York: Random House.
Odell, Jenny. 2020. *How to do nothing: Resisting the attention economy*. Brooklyn: Melville House.
Paton, Margaret. 1990. A reconsideration of Kant's treatment of duties to oneself. *Philosophical Quarterly* 40: 222–233.
Price, Catherine. 2018. *How to break up with your phone*. New York: Ten Speed Press.
Ravenscraft, Eric. 2019. 500 days of Duolingo: What you can (and can't) learn from a language app. *The New York Times*. https://www.nytimes.com/2019/05/04/smarter-living/500-days-of-duolingo-what-you-can-and-cant-learn-from-a-language-app.html. Accessed 9 June 2023.
Schofield, Paul. 2015. On the existence of duties to the self (and their significance for moral philosophy). *Philosophy and Phenomenological Research* 90: 505–528.
Singer, Marcus G. 1959. On duties to oneself. *Ethics* 69: 202–205.
Timmermann, Jens. 2006. Kantian duties to the self, explained and defended. *Philosophy* 81: 505–530.

United States. Public Health Service. Office of the Surgeon General. 2023. *Social media and youth mental health. The U.S. Surgeon General's advisory.* Rockville: U.S. Department of Health and Human Services, Public Health Service.

Velleman, J. David. 1999. A right of self-termination? *Ethics* 109: 606–628.

Wagaman, M. Alex, Keith J. Watts, Vee Lamneck, Sandra A. D'Souza, Lauren B. McInroy, Andrew D. Eaton, and Shelley Craig. 2020. Managing stressors online and offline: LGBTQ+ youth in the southern United States. *Children and Youth Services Review* 110: 104799.

Wick, Warner. 1960. More about duties to oneself. *Ethics* 70: 158–163.

Williams, James W., and Amy White. 2020. *Digital minimalism in everyday life: Overcome technology addiction, declutter your mind, and reclaim your freedom.* Independently published.

Wood, Allen. 1999. *Kant's ethical thought.* New York: Cambridge University Press.

———. 2009. Duties to oneself, duties of respect to others. In *The Blackwell guide to Kant's ethics*, ed. Thomas E. Hill Jr., 229–252. Oxford: Wiley-Blackwell.

Zahariades, Damon. 2018. *Digital detox: The ultimate guide to beating technology addiction, cultivating mindfulness, and enjoying more creativity, inspiration, and balance in your life!* Independently published.

CHAPTER 5

The Duty to Promote Digital Minimalism in Others I: Duties of Virtue

Under the present conditions of human beings one can say that the happiness of states grows simultaneously with the misery of human beings. And there is still the question whether we would not be happier in a raw state, without all this culture, than we are in our present condition. For how can one make human beings happy, if one does not make them moral and wise?

—Kant, Lectures on Pedagogy (UP 9:451) (Kant's lectures on pedagogy can be found in Kant (*2007a*).)

"'On the scale between candy and crack cocaine, it's closer to crack cocaine,' Mr. Anderson said of screens. Technologists building these products and writers observing the tech revolution were naïve, he said. 'We thought we could control it,' Mr. Anderson said. 'And this is beyond our power to control. This is going straight to the pleasure centers of the developing brain. This is beyond our capacity as regular parents to understand.'"

—Nellie Bowles (*2018a*) (This comes from a *New York Times* article about how parents in Silicon Valley have come to appreciate the dangers that screen time poses to their children. It was titled "A Dark Consensus About Screens and Kids Begins to Emerge in Silicon Valley." This quote comes from Nellie Bowles's interview with Chris Anderson, the former editor of *Wired*. See Bowles (*2018a*).)

T. Aylsworth, C. Castro, *Kantian Ethics and the Attention Economy*,
https://doi.org/10.1007/978-3-031-45638-1_5

5.1 Introduction

In the previous chapter, we argued that you have a duty to be a digital minimalist. Our conclusion was that you owe it to yourself to be more intentional about how you use mobile devices and how you engage with the attention economy. You should use these technologies in ways that do not interfere with your ability to set and pursue your own ends. And although we are fully committed to that claim, we do have some reservations about it. There is something troubling about characterizing the issue as if it all boiled down to personal responsibility.[1] After all, we have already explained how several external forces are at work in creating and sustaining our unhealthy relationships with mobile devices.

First, there are the developers and tech companies, who work tirelessly to create products that are addictive. Second, educators and employers compel us to use technology in ways that all but require us to have problematic relationships with mobile devices. And, finally, the young adults of today grew up in a world that was saturated with smartphones, tablets, and social media. They were lured into using mobile devices before adulthood, forming bad habits before they were entirely responsible for their own actions. Technology began undermining their autonomy before it was even fully formed.[2]

So it is important not to overlook the social dimension of this problem. Although we do have duties to ourselves to restructure our relationship with technology, this is not the end of the story. It would be a mistake to think that there is a simple, individualistic solution to a complex, collective problem. It may be tempting to believe that we can solve the problem by going camping without our phones, meditating, or participating in a "digital detox" retreat. But that's not enough. James Williams, a former

[1] We do believe that individuals *can* and *should* take action in their own lives. Our point here is simply that this is not the whole story. We also have duties to others and others owe duties to us.

[2] Indeed, as we noted in various places, the problems we discuss in this book seem to be especially acute among young people. According to a recent meta-analysis, we can estimate that "one fourth of children and young people suffer from problematic smartphone use, a pattern of behavior that mirrors that of a behavioral addiction," which is associated with "depression, anxiety, high levels of perceived stress, and poor sleep" (Sohn et al. 2019, 7).

strategist at Google,[3] uses a helpful analogy to explain the shortcoming in this way of thinking. He says a digital detox is "not the solution, for the same reason that wearing a gas mask for two days a week outside isn't the answer to pollution. It might, for a short period of time, keep, at an individual level, certain effects at bay. But it's not sustainable, and it doesn't address the systemic issues" (Hari 2022, 101).

Addressing these systemic issues will be our aim for the remainder of the book. We begin in this chapter by defending the existence of a duty to promote the autonomy of others with respect to their use of mobile devices. We call this the "duty to be an attention ecologist."[4] In many ways, the argument for this conclusion will resemble the argument of the previous chapter. We will continue to rely on the conclusions of Chap. 2. The moral obligations described in this chapter are grounded, once again, in the idea that autonomy is morally valuable and that we are obligated to respect humanity.

We begin by modifying the argument from the previous chapter. The duty to promote the autonomy of others is not, strictly speaking, identical to the duty to promote your own autonomy. You have control over yourself, but you do not (and should not) have sovereignty over other autonomous adults. So we must present an amended version of the argument in order to extend the duty to others. In the next section we address an obstacle that Kant's ethics presents for this project. Once we have established the existence of a duty to be an attention ecologist, we present three applications of this obligation.

As we will explain, you are not in a position to promote *everyone's* autonomy. Not only would that be impossible, in many cases it would be inappropriate. Imagine you see a stranger working in a cafe. You watch them struggle to complete a work-related task because they keep responding to social media notifications on their phone. It would not be your

[3] Williams is now a member of the Digital Ethics Lab at Oxford and a researcher with the Uehiro Centre for Practical Ethics. Along with Tristan Harris and Joe Edelman, Williams was also a co-founder of Time Well Spent, a movement that is dedicated to reforming the industry. They argue that technology should be more in line with human values and be less concerned with maximizing screen time. Indeed, the aims of Time Well Spent are remarkably similar to our own: "[I]nstead of building products that are designed to keep your face affixed to your phone, it's about making tech that's in line with people's basic human needs and values. The purpose shouldn't be to get you to swipe at more push notifications, but complete your goals" (Baer 2017).

[4] See Aylsworth and Castro (2022).

place to tell them how to conduct their affairs. You have no obligation to promote their autonomy with respect to their phone, and it would be disrespectful of you to interfere. So we must show why the duty to be an attention ecologist applies in some contexts rather than others. We take this on in Sect. 5.3 where we discuss the special obligations that parents and teachers have to promote the autonomy of their children and students. These are two important instances of the duty to be an attention ecologist. There are good reasons to think that parents and teachers have agent-relative obligations to promote autonomy, as the cultivation of autonomy is arguably a constitutive aim of these relationships.

In Sect. 5.4, we transition to a different set of duties to others. In some cases, we have positive duties to cultivate the autonomy of others (such as parents and teachers). In other cases, our duties to others involve negative duties to refrain from undermining someone else's autonomy. In Sect. 5.4, we will discuss obligations that employers and software developers have to refrain from undermining the autonomy of employees and users. The duties in Sect. 5.3 are what Kant calls "duties of love," and the duties discussed in 5.4 are "duties of respect."[5]

5.2 The Argument to Be an Attention Ecologist

In the previous chapter, we gave an argument for a duty to oneself to be a digital minimalist; in this chapter we will give a corresponding argument that we also have a duty to be an *attention ecologist*, one who promotes digital minimalism in others. The philosophical core of the argument for digital minimalism ran as follows:

P1$_a$ Humanity has an objective, unconditional, non-fungible value—dignity.
P2$_a$ Anything that has dignity ought to be respected as an end and never treated as a mere means.
P3$_a$ If humanity ought to be respected as an end and never treated as a mere means, then we ought to adopt the end of our own natural perfection.
Therefore,
C1$_a$ We ought to adopt the end of our own natural perfection.

[5] Duties of love map onto imperfect duties and duties of respect map onto perfect duties. See Wood (2009).

From here, we argued that, given the empirical evidence, $C1_a$ gives rise to a duty to be a digital minimalist.

> It would seem that the corresponding argument for attention ecology would run as follows:
> $P1_a$ Humanity has an objective, unconditional, non-fungible value—dignity.
> $P2_a$ Anything that has dignity ought to be respected as an end and never treated as a mere means.
> $P3_b$ If humanity ought to be respected as an end and never treated as a mere means, then we ought to adopt the end of *others'* natural perfection.
> Therefore,
> $C1_b$ We ought to adopt the end of *others'* natural perfection.

From here, we could, like the argument for digital minimalism, argue that:

> $P4_b$ Mobile device use threatens to undermine *others'* capacities of rational agency and authenticity. The compulsive use of such devices is inconsistent with the end of natural perfection.
> $P5_b$ If $P4_b$ and $C1_b$ are true, then we ought to *promote in others* a disposition to use mobile devices intentionally and in ways that do not undermine *their* capacities (i.e., we ought to be *attention ecologists*).
> Therefore,
> $C2_b$ We ought to be *attention ecologists.*

We ultimately think that this strategy works. However, we cannot accept the argument for attention ecology without further comment.

Kant rightly thinks we stand in a very different relation to our duty to perfect ourselves than to any duty that we might have to perfect others. One major difference—one that we, like Kant, accept as important—is that others' sovereignty over themselves conditions how we might fulfill the duty (if we indeed have it at all). While one might, out of respect for their own autonomy, use a browser extension to limit their access to certain distracting websites, they should not simply log on to their partner's device and do the same. They would, at the very least, need to obtain permission first. Indeed, Kant thought that the asymmetry was so significant that he, at times, expressed it in ways that might be thought to undermine $P3_b$. At times, he seems to imply that it is impossible and therefore not obligatory to adopt the end of perfecting others.

Our immediate goal, then, is to explain Kant's reasons for thinking that taking the end of perfecting others is impossible and to explain why his reasons do not block the above argument for attention ecology. Grappling with Kant on this issue will shed light on complications with pursuing the interpersonal goal of attention ecology that do not arise with the intrapersonal goal of digital minimalism; these complications will, in turn, inform our derivation of the specific duties of attention ecology over the next two chapters.

When Kant discusses obligations that stem from our duty to respect humanity, he identifies just two ends that we must set for ourselves: our own perfection, and the happiness of others. For him, these duties are asymmetrical. That is, he accepts:

> **The happiness/perfection asymmetry**: We *do* have a duty to promote our own perfection, and we *do* have a duty to promote the happiness of others; however, we *do not* have a duty to promote our own happiness, and we *do not* have a duty to promote the perfection of others. (*MS* 6:386)

Kant thinks that we do not have a duty to perfect our own happiness because this is something that we already pursue naturally, making the language of duty a poor fit (*G* 4:414). When it comes to perfecting others, Kant seems to think that this is "self-contradictory" and therefore not required:

> The *perfection* of another human being, as a person, consists in just this: that he *himself* is able to set his end in accordance with his own concepts of duty; and it is self-contradictory to require that I do (make it my duty to do) something that on the other himself can do. (*MS* 6:386)

Taking inspiration from Lara Denis (2001),[6] we will read Kant's argument as follows:

[6] We should note here that Denis offers this reading as the most plausible of her three reconstructions. The other two have "I cannot do anything to promote another person's virtue" and "I so not have it in my power to will the means sufficient to another person's virtue" in the place of ($P2_c$) (see Denis 2001, 143). She identifies these two premises as less promising than ($P2_c$). The first, she says, is "obviously false": we certainly can, among other things, encourage others to be moral, praise them for being moral, and so on. The second is also implausible. As Denis notes, Kant holds that we have a duty to strive toward perfecting ourselves, even though we will never reach it (*MS* 6:446).

$P1_c$ If it is impossible for one to will the indispensable means to φ'ing, they cannot be required to φ.
$P2_c$ It is impossible to will the indispensable means of perfecting another person.
$C1_c$ Thus, one cannot have a duty to perfect another person.

The idea behind $P2_c$ is that an essential element of the virtue in question is guiding *oneself* in accordance with one's own concepts of duty. Further, Denis aptly notes, it is important to remember that $P2_c$ states that we do not have it in our power to *will* (as opposed to, e.g., *wish* or *want*) the operative end: if we guide another, then they are not guiding *themself*. So, we *cannot will* the end in question. If this interpretation is correct—if setting the end of self-perfection is something one must do for *oneself*—then it might seem as though we simply *cannot* perfect another person without, as Kant says, contradiction (MS 6:386).

Now that we have deepened our understanding of Kant's reasons for accepting the happiness/perfection asymmetry, we can explain why this argument poses no threat to our argument for attention ecology. Let us here canvass three strategies, each of which is compatible with the others and each of which saves $P3_b$ from any threat posed by Kant's argument. One approach simply rejects $P2_c$, ultimately concluding that Kant was wrong to reject the symmetry. Another strategy works around the asymmetry by grounding attention ecology in the duty to promote others' happiness. Another shows that attention ecology can be conceived of as a negative duty, grounded in respect, to refrain from "giving scandal," that is, tempting others to do what is wrong.

Despite the fact that any of the three responses could suffice as an answer to the challenge, we choose to work through all three because they shed light on different aspects of attention ecology that we would like to explore. For instance, seeing that we can conceive of attention ecology as an imperfect duty (via the first two strategies) sheds light on the positive aspects of the duty and assists in its applications to, e.g., students and children. Seeing that we can conceive of it as a negative duty will help when we talk about how it applies in more distant relationships, e.g., employer/employee, and whether legislators have grounds to codify aspects of the duty in the law.

In pursuing the first strategy—i.e., denying $P2_c$.—Denis notes that it is not impossible to will that we put another "in a position in which they can exercise either self-governance and correct their errors" (Denis 2001,

147). Korsgaard expresses a similar sentiment, stating that the happiness/perfection asymmetry is "overstated":

> [I]t is clear that we have a duty to provide for the moral education of our children, and, Kant himself insists, our intimate friends (MPV 387). Choosing ends on another's behalf is as impossible as it would be disrespectful, but putting others in a good position to choose ends for themselves, and to choose them well, is the proper work of parents, teachers, friends, and politicians; providing for someone's moral education as well as nurturing her self-respect is an important part of the way we do this. (Korsgaard 1996, 220)

This seems right. As Denis notes, our duty to promote the perfection of others would be one where we have latitude, and one that we could choose—as Korsgaard suggests—to fulfill in the case of our intimates (our children, close friends, students, and so on). O'Neill offers further suggestions on this front, stating that we can support the rational self-governance of others by, for example, ameliorating poverty, hunger, and political domination (O'Neill 1980; cf. Denis 2001).

In many cases, we are in a special position to know what our intimates have chosen for themselves or what they would need in order to be in a position to choose for themselves and could take it upon ourselves to provide this for them. Here Korsgaard—like Denis—lists a number of things we can certainly will in relation to the perfection of others. In many cases, our reach is conditioned by the relation in which we stand with the person whom we are interested in helping, but this is not always an obstacle. Sometimes it enables us to support them in the pursuit of their perfection.

But let us now suppose that this first strategy doesn't work; suppose that we insist that $P2_e$ is unassailable because we cannot, in some narrow sense, perfect others *as persons* (*MS* 6:386). Robert Johnson pursues this strategy in defending Kant from charges of inconsistency here. He states that one can be perfected as *a person* "only if and to the degree that it represents the exercise of that individual's own will" (Johnson 2011, 146.) While we are not inclined to accept this (we think that the initial responses from Denis et al. to $P2_e$ are successful) this still doesn't block the argument for attention ecology. As Johnson is happy to admit, while we cannot perfect others *as persons*, we can perfect them in other ways: we can perfect them *as pianists*, or *Spanish speakers*, or *marathon runners,* or whatever skill it is that they are working on (other than rational agency *as such*).

As Wood (2004) reminds us, Kant claims that our duty of beneficence includes an obligation to promote exactly these sorts of things, including the "moral well-being" of others (MS 6:394). As he puts it:

> Kant's point could … be put this way: I do have a duty to promote my own happiness, but only insofar as my happiness falls under the heading of my perfection; and I do have a duty to promote the perfection of others, but only insofar as it falls under the heading of their happiness. (2004, 148)

So, on pain of inconsistency, it seems that *even if* the asymmetry is in some sense true, it must be limited in scope: it must leave room for a duty to promote the moral well-being of others out of our duty to be beneficent.[7] Regardless of Kant's position, this view became more popular in the decades after he wrote about it. Later German idealists, such as Fichte, held the position that we should indeed take on the project of morally perfecting others.[8]

Finally—even if we leave the aforementioned strategies on the table—there is another move available to us, and this is to ground the duty in a perfect duty to respect others, that is, to assist them in, as O'Neill puts it, "preserving them from temptation" (O'Neill 1975, 91), or as Kant would say, assisting them in refraining from "giving scandal" (MS 6:464). This would make attention ecology a perfect duty, one that would involve, say,

[7] Mary J. Gregor makes a similar point, and she emphasizes what Kant says about teachers and children. She writes: "Against Kant's exclusion of the perfection of others from our obligatory ends it might be said that, while we admittedly cannot perform a moral action for another person, we can still provide the conditions which will help, or at least not hinder, others in their self-imposition of duty. In fact, Kant does not fail to recognize that our actions and attitudes can influence the moral development of others: his discussions of practical anthropology presupposes this, and when he descends into the relations of an adult to a child or youth, or a teacher to his pupil, he discusses at some length what the adult can do to assist his charge in developing moral character and to prepare him for his adoption of moral purposes" (1963, 186).

[8] Fichte argues in his *System of Ethics* that we are morally required to promote morality both in our own person and in others. Stefano Bacin (2021) highlights this contrast between Kant and Fichte, citing §18 of Fichte's *System of Ethics:* "What I will is morality as such" and "it does not matter in the least whether this is in me or is outside me," "my end is achieved if the other person acts morally" (Fichte 2005, GA I/5: 210; SL 4: 232). Allen Wood explains how Fichte saw himself as simply trying to be a consistent Kantian on this point. Fichte stressed the ethical importance of institutions that would develop the morality of others (Wood 2016, 214).

not developing addictive applications or forcing workers and students to use smart devices in ways that are bad for them.

In sum, the happiness/perfection asymmetry, and Kant's reasons for accepting it, do not seem to pose much of a challenge for us here. We can summarize our response to the apparent obstacle in the following way. The view we defend in this chapter is that we have a moral obligation to promote the autonomy of others with respect to their use of technology. Someone could object that this is inconsistent with what Kant says in the *Metaphysics of Morals* and is thus at odds with his considered view about our duties to others. We have outlined three possible responses to this objection. First, we could concede that our view is inconsistent with Kant's and claim that he is simply wrong about morally perfecting others. Denis (2001) provides some reasons for doubting the soundness of Kant's argument. Second, we could argue that our position contradicts what he says in the Doctrine of Virtue but that he overstates his position and fails to express his considered view. This is the strategy suggested by Wood. Finally, we could suggest that Kant is indeed expressing his considered view, but that our position is perfectly consistent with what he is saying. His reservation about morally perfecting others has a very narrow scope, and he simply means it would be impossible for us to autonomously set ends for others as this is something that they must do for themselves.

Note that these strategies open up different options to us. The first two pave the road for applications of attention ecology that are more positive, i.e., taking the end of perfecting another. The last is more negative: a duty to *refrain* from, e.g., tempting others to do what is wrong. This nicely corresponds to two sides of the formula of humanity ("act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means" (*G* 4:429)). One part of the formula asks us to treat others as ends (as something to promote), another to never use others merely as a means. This division nicely maps on to Kant's distinction between duties of love (which ask us to promote others' permissible ends) and respect (which ask us never to degrade another) (*MS* 6:450). This chapter will respect that division by first exploring the duty as it pertains to those that we seem to have a duty to perfect: our children, students, and close friends. Then, we will present duties of respect: employers to their employees and software developers to users. Finally, we will revisit respect as it applies to children, students, and close friends.

5.3 Applications of the Duty I: Always as Ends in Themselves

In the last chapter, we saw how imperfect duties involve moral discretion. They compel us to adopt some end, such as the happiness of others, but this does not tell us exactly what we must do to promote that end. You might promote the happiness of others by treating your lonely friend to dinner, volunteering at a homeless shelter, giving money to famine relief, or fighting climate change.[9] When Kant says that imperfect duties leave us some latitude or "playroom," he is not suggesting that imperfect duties are somehow less strict than perfect duties.[10] You are strictly required to adopt the end in question, but you must determine for yourself how to promote that end.

In the previous section, we explained how the duty to be an attention ecologist can fall within the domain of beneficence. Promoting the happiness of others means pursuing their ends as if they were your own.[11] When a good student requests a letter of recommendation for law school, we

[9] Consequentialists sometimes fault Kantians for having less to say on these matters (e.g., Jamieson 2007), but recent work has shown how Kant's ethics can be fruitfully applied to these large-scale issues. For instance, Albertzart (2019) shows how the duty of beneficence could entail a duty to fight climate change: "To adopt the happiness of others as an end implies willing the necessary means for achieving this end. Given the negative impact climate change is expected to have on human happiness, combating climate change qualifies as one of these necessary means" (844). Many others have addressed Kantian responses to the issue of famine relief. See O'Neill (1980), Atterton (2007), and Sensen (2022).

[10] He writes, "[F]or if the law can prescribe only the maxim of actions, not actions themselves, this is a sign that it leaves a playroom (*latitudo*) for free choice in following (complying with) the law, that is, that the law cannot specify precisely in what way one is to act and how much one is to do by the action for an end that is also a duty. — But a wide duty is not to be taken as permission to make exceptions to the maxim of actions but only as permission to limit one maxim of duty by another (e.g., love of one's neighbor in general by love of one's parents), by which in fact the field for the practice of virtue is widened" (*MS* 6:390).

This could be used to address questions about the limits of beneficence when it comes to things like famine relief or other forms of effective altruism. Just how much are we required to give? Kant's answer is that it depends on the agent and her needs: "I ought to sacrifice a part of my welfare to others without hope of return, because this is a duty, and it is impossible to assign determinate limits to the extent of this sacrifice. How far it should extend depends, in large part, on what each person's true needs are in view of his sensibilities, and it must be left to each to decide this for himself" (*MS* 6:394).

[11] Of course, we must add a crucial proviso. You should promote *only* their morally permissible ends. You may help your friend get into law school by studying for the LSAT with her, but you should not help her rob a bank. Kant writes, "The duty of love for one's neighbor can, accordingly, also be expressed as the duty to make others' ends my own (provided only that these are not immoral)" (*MS* 6:450). Cf. *MS* 6:388.

write one in an effort to help them pursue their end. But what should we do if we see them struggling to study for the LSAT because, like Esther, they cannot resist the temptation to look at their phone? Surely we could help them pursue their end by helping them become a digital minimalist.[12] But as we pointed out in the introduction, it is not always our place to interfere with the affairs of others. How, exactly, can we determine when to act on the duty to be an attention ecologist?

In this section, we will discuss three contexts where it seems perfectly appropriate to promote digital minimalism in others. We will argue that parents are morally obligated to promote the autonomy of their children and we will defend a similar view with respect to teachers and their students. We believe (as Kant did) that fostering autonomy should be seen as a fundamental aim of parenting and teaching. We then extend the duty to close friendships. If you are close enough to someone to promote their ends as if they were your own, then you are probably in a position to help them have a healthier, more autonomous relationship with their mobile devices and the attention economy.

5.3.1 Implications for Parents and Teachers

It should come as no surprise that Kant championed autonomy as the principal aim of both parenting and education. After all, his thoughts on these issues were famously influenced by Rousseau's *Émile*. Like Rousseau, Kant argues that the cultivation of autonomy ought to be the aim of the entire educational enterprise. In the end, Kant believes that this project must go beyond merely building capacities. He believes that education should ultimately culminate in the proper exercise of rational agency:

[12] As we explained in the previous chapter, the details of the advice will vary. For some people, simply turning off notifications might be enough. Others may want to change their screen to black and white or delete social media and news apps. Like us, they might opt for a more extreme approach by replacing their smartphone with a less distracting device.

morality.[13] But even if we do not share Kant's particularly moralistic vision of education, we are likely to find something very plausible about his claim that parents and teachers ought to adopt the aim of cultivating autonomy.

For the most part, parents already see this as an important part of their job. They want to help their children become adults who are capable of setting and pursuing their own ends. Similarly, teachers tend to see themselves as developing their students' capacities in such a way that they will be free to pursue a variety of paths in life. The two of us are parents and teachers, and we are certainly committed to the project of fostering the autonomy of our respective children and students.

We will present Kant's thoughts on parenting and education alongside the work of contemporary philosophers.[14] Luckily, there is a great deal of convergence on these issues. Many contemporary theorists of education share Kant's commitment to the centrality of autonomy, as do those who write about parenting.[15] And in both cases, we believe that attention ecology fits in very well with these projects. We will revisit some of the empirical conclusions presented in Chap. 3, but we will place a particular emphasis on the research that has looked at the effect of smartphones and the attention economy on children. We will begin with education and then turn to parenting.

[13] In his lectures on pedagogy, Kant says that education should begin with discipline. This is when we learn to tame our impulses (animality) and to restrain ourselves through reason (humanity). Second, he says that the "human being must be cultivated" (UP 9:449). Here the project is not strictly negative. We must cultivate skills to pursue various ends. For instance, Kant sees reading and writing as essential skills. Third, Kant identifies prudence as a necessary component of becoming integrated into civilized society. And finally, there is moralization. Kant says that we must not educate pupils to pursue any ends whatsoever; we should cultivate the "disposition to choose nothing but good ends" (Ibid.). For more detailed discussions of Kant's views on education, see Herman (2007), Louden (2000), Cohen (2016), Roth and Formosa (2019), and Frierson (forthcoming).

[14] Our discussion of Kant's views on education is greatly indebted to Frierson (forthcoming).

[15] On education, see Brighouse (1998, 2006). See also Schouten (2018), Levinson (1999), Barrow and Woods (2006), Gutmann (1987), Macedo (1990), and Callan (1997). Although many of these discussions are about civic education and liberal legitimacy, an incredibly diverse set of perspectives converges on this conclusion about autonomy. Patrick Frierson has written about how Maria Montessori, one of the most influential philosophers of education, made autonomy a central aim of education. See Frierson (2016) and (2021). For parenting, see Brighouse and Swift (2014), Callan (2002), Feinberg (1980), Austin (2007), and Bluestein (1982). As Paul Smeyers puts it, "In its most general terms, education and child rearing is centrally concerned with autonomy" (Smeyers 2012, 1).

When it comes to thinking about the purpose of education, a variety of things should come to mind. This is an important question—one that should be discussed far beyond the confines of academic philosophy. We need to be able to provide justifications for education, especially given the fact that almost every government in the world has laws requiring children to go to school.[16] If the state is going to use its coercive power to make education compulsory, then it is incumbent on us to justify this constraint on people's liberty.

For democratic countries, one potential justification would be the importance of an educated electorate. If citizens are going to enact their own legislation (or going to elect representatives who do that for them), then education surely has some value in this regard. We want our fellow citizens to be educated so that they can understand the policies that are being proposed and engage in rational deliberation about them. Indeed, the benefits of an educated citizenry are so vital to democracy that even libertarians like Milton Friedman believe that we can justify requiring the state to pay for education.[17]

[16] There are *very* few exceptions. Among UN members, the only countries that do not have compulsory education are Solomon Islands, Oman, Bhutan, and Papua New Guinea. See Besche-Truthe (2022).

[17] Libertarians are typically committed to a theory of negative freedom and property rights that precludes redistributive efforts by the state. They argue that it is unjust for the state to take your money to pay for benefits to someone else. Indeed, Nozick (1974) famously argues that the taxation of earnings is tantamount to slavery (169). Nozick's vision of a minimal "night-watchman" state rules out the permissibility of compulsory education. Jonathan Wolff points out how the lack of education in Nozick's libertarian utopia would be problematic, however: "Nozick realizes that there are problems with the framework. Children present one such. At what age, for example, should they be able to leave? Do they have a right to be informed in a balanced way about alternative ways of life? If a society shares the belief that to discuss the theory of evolution is a sin, may they indoctrinate this belief in their children? If so, how can their children make free choices? If not, how is this to be regulated" (1991, 135)?

Friedman's view is less stringent than Nozick's. In *Capitalism and Freedom*, he argues that the benefits of education extend far beyond the individual student who receives it. He writes, "A stable and democratic society is impossible without a minimum degree of literacy and knowledge on the part of most citizens and without widespread acceptance of some common set of values. Education can contribute to both. In consequence, the gain from the education of a child accrues not only to the child or his parents but also to other members of society. It is not feasible to identify particular individuals (or families) benefitted and so to charge for the services rendered. There is therefore a significant neighborhood effect" (1975, 86). Cf. Brighouse (2000, 26). Of course, Friedman goes on to argue that although the state should *pay* for education, he believes that it is not maximally efficient for them to administer the education. It is here that he presents the first well-known argument for school voucher programs.

And in addition to the positive externalities for democracy, education has undeniable economic benefits. We all have a stake in the existence of a thriving economy, and education can play an important role in enhancing the value of labor. By educating our citizens, we produce a more highly skilled workforce, and this benefits us all. Harry Brighouse refers to this as the "human capital theory approach" (2006, 27). He writes, "[T]he imperative is developing a strong and competitive economy, and the means is educating children to be productive workers. This benefits everyone; we all gain from higher Gross Domestic Product, and children gain from the fact that they are more able to operate well in the workplace" (Ibid.). To some extent, Brighouse is sympathetic to this aim. If we do not equip children with the tools they need to participate in the economy, then we would be depriving them of the basic tools they need in order to flourish. But he rejects the idea that this should be seen as the sole or primary aim of education.

Brighouse, like many other philosophers who discuss education, argues that we should see the cultivation of autonomy as the principal aim and justification of education.[18] He does this on the grounds that autonomy is essential for flourishing. Many discussions of education (and parenting) make reference to the famous US Supreme Court case of Wisconsin v. Yoder.[19] In the 1970s, three Amish students in New Glarus, Wisconsin stopped attending school after eighth grade. This was a violation of the state's law on compulsory education (which required schooling until the age of 16).[20] The state argued that it should continue to require the students to attend school, but the representatives of the Amish claimed that further schooling would undermine the children's religious upbringing and was thus at odds with the First Amendment's protection of the free exercise of religion.

The court ultimately decided in favor of the Amish, as they agreed with the claim that further education would "gravely endanger if not destroy the free exercise of their religious beliefs" (Wisconsin v. Yoder, 406 U.S. 215–219). We do not have a particularly strong stance on the court's decision. As Feinberg (1980) points out, there is not much difference

[18] There are so many thinkers in agreement on this point that it would be impossible to provide an exhaustive list. See the footnote above for a handful of references.

[19] Brighouse (2006) discusses the case, as do Feinberg 1980, Schouten (2018), Callan (1997), Fernandez (2010), Gutmann (1995), and Galston (1995).

[20] The age requirement varies from state to state. In the US, the range of ages at which a student can drop out of school is between 14 and 18.

between compulsory education ending at the age of 14 rather than 16. So the decision in the case might not be the most philosophically interesting issue. For many philosophers, it is the opinions of Justice Douglas and Justice White that have sparked the most interest.[21] Justice White points out how the case would have been radically different if the Amish were trying to withdraw from compulsory education altogether. He argued that such an action would be unacceptable on the grounds that it would undermine the children's autonomy to choose their own path in life:

> In the present case, the State is not concerned with the maintenance of an educational system as an end in itself; it is rather attempting to nurture and develop the human potential of its children, whether Amish or non-Amish: to expand their knowledge, broaden their sensibilities, kindle their imagination, foster a spirit of free inquiry, and increase their human understanding and tolerance. It is possible that most Amish children will wish to continue living the rural life of their parents, in which case their training at home will adequately equip them for their future role. Others, however, may wish to become nuclear physicists, ballet dancers, computer programmers, or historians, and for these occupations, formal training will be necessary. There is evidence in the record that many children desert the Amish faith when they come of age. A State has a legitimate interest not only in seeking to develop the latent talents of its children, but also in seeking to prepare them for the lifestyle that they may later choose, or at least to provide them with an option other than the life they have led in the past. (Wisconsin v. Yoder, 239–40)

Justice White argues that education plays a vital role in equipping children with the capacities they need in order to *choose* their own path in life and to *pursue* it effectively. In other words, Justice White is concerned with what we have been describing as Kantian respect for humanity. He wanted to protect the children's ability to set and pursue their own ends, and he recognized that education is indispensable in that regard.

The dissenting opinion of Justice Douglas is similar, but in some ways he was even more forceful about the rights of the children:

> On this important and vital matter of education, I think the children should be entitled to be heard. While the parents, absent dissent, normally speak for

[21] Justice White joined the majority, but he wrote a separate, concurring opinion that was endorsed by Justices Brennan and Stuart. Justice Douglas wrote another opinion, dissenting in part.

> the entire family, the education of the child is a matter on which the child will often have decided views. He may want to be a pianist or an astronaut or an oceanographer.
>
> To do so he will have to break from the Amish tradition.
>
> It is the future of the student, not the future of the parents, that is imperiled by today's decision. If a parent keeps his child out of school beyond the grade school, then the child will be forever barred from entry into the new and amazing world of diversity that we have today. The child may decide that that is the preferred course, or he may rebel. It is the student's judgment, not his parents', that is essential if we are to give full meaning to what we have said about the Bill of Rights and of the right of students to be masters of their own destiny. (Wisconsin v. Yoder, 244–45)

Both of these opinions highlight something important about education and autonomy. Many children who grow up in the Amish community may find that they are perfectly suited for such a life. Living in such a tight-knit, religious community can provide them with profound goods that are difficult to find in other contexts. Some Amish children may see the value in that way of life and autonomously choose it for themselves. But others may find that this is not the life for them. And if those children are deprived of an education, then, as Justice Douglas says, they will be forever barred from alternative life paths.

This is why many philosophers see autonomy as the central value of education. We all want children to flourish. And, as Brighouse points out, we must recognize that not every child is lucky enough to be perfectly suited for the lifestyle, values, and commitments that they inherit from their parents.[22] Accordingly, education must do more than teach children to read, write, and do arithmetic. It should give students a sufficiently broad view of the world so that they can choose their own path in life. We must give them the tools they need to author their own life stories.

As for parenting, we should be able to reach the same conclusion. Like teachers, parents are obligated to cultivate the autonomy of their children. The only difference is the grounding of the duty. Teachers are obligated to promote the autonomy of their students insofar as building capacities is

[22] He writes, "One purpose of delivering the resources and liberties that justice requires is to enable people to live well by their own judgement. But to live well, one needs more: one also needs some sense of what constitutes living well. So providing the opportunity to enter ways of life requires that the state educate children in the skills of rational reflection and comparison usually associated with autonomy" (Brighouse 2006, 18–19).

a constitutive aim of education. Teachers voluntarily take on this role, so it is incumbent on them to adopt this end. The situation is similar for parents, but the conditions vary slightly. Obviously, there is (and ought to be) less freedom of exit for parents than there is for teachers. But, like teachers, parents owe something to their children when it comes to their autonomy. Imagine a set of parents who do nothing more than satisfy their child's biological needs. The child has food, water, and a place to sleep, but her parents do nothing to cultivate her capacities. They do not read to her; they do not allow her to attend school (or educate herself); they do nothing to promote her basic physical abilities; etc. It is obvious that her parents have wronged her.[23] Parents are morally obligated to do more than just keep their child alive.

We may wonder what grounds parental obligations. Why should we think that parents owe it to their children to cultivate their autonomy? One way to answer this question is to point out that parents, like teachers, have (typically) taken this role voluntarily. But the child did not. Children are, obviously enough, brought into existence without their consent. In the *Metaphysics of Morals*, Kant says that parents have obligations to their children because they "brought a person into the world without his consent" and so they "incur an obligation to make the child content with his condition so far as they can" (*MS* 6:280).[24]

But even if this tells us something about the grounds of parental obligations, it does not tell us enough about the content of those obligations. For that, it would be useful to draw on what Joel Feinberg calls "the child's right to an open future" (Feinberg 1980). Of course, children have certain rights in common with adults. Both adults and children have a right to life, for instance. But Feinberg explains a special class of rights that

[23] Brighouse and Swift go so far as to claim that state intervention would be justified in such a case: "For us, children have a vital interest in developing the capacity for autonomy, and parents harm children—in ways that the state may legitimately seek to prevent—when they deny them the kind of upbringing that develops that capacity" (Brighouse and Swift 2014, 12). Everyone recognizes the value of keeping children and parents together, but it is also widely recognized that children should be protected from abuse and neglect. Brighouse and Swift defend a "thicker sense" of what is required of the state when it comes to protecting the child's interests (Ibid.).

[24] For more on the Kantian justification, see Puls (2016). Contemporary scholars have presented similar arguments. Austin (2007) discusses an account of parental obligations according to which parents incur "custodial obligations" to their child by agreeing to take on the job (Austin 2007, 34). Austin notes that Blustein (1982), O'Neill and Ruddick (1979), and Vallentyne (2002) all defend for the "consent view."

pertain only to children. We can think of these as "rights-in-trust" because they involve safeguarding future possibilities that the child cannot pursue at present. As Feinberg puts it:

> When sophisticated autonomy rights are attributed to children who are clearly not yet capable of exercising them, their names refer to rights that are to be *saved* for the child until he is an adult, but which can be violated "in advance," so to speak, before the child is even in a position to exercise them. The violating conduct guarantees *now* that when the child is an autonomous adult, certain key options will already be closed to him. His right while he is still a child is to have these future options kept open until he is a fully formed, self-determining adult capable of deciding among them. (Feinberg 1980, 76–77)

By refusing to cultivate any of the child's capacities, the parents have arguably violated the child's rights. No one chooses to be born, but everyone should have the right to choose what to do with their life. This means that parents must cultivate the capacities of their children so that they are able to set and pursue their own ends.[25]

Of course, Kant wholeheartedly agrees. In the previous chapter, we saw how Kant viewed the cultivation of humanity as a moral obligation that you owe to yourself. He says that we have a duty "to make ourselves worthy of humanity by culture in general, by procuring or promoting the *capacity* to realize all sorts of possible ends, so far as this is to be found in the human being himself" (*MS* 6:392). But this is not the end of the story. Although you do owe this duty to yourself, Kant also claims that we owe this duty to one another and he believes that education and parenting are essential in this regard.[26] In his lectures on pedagogy, he says, "The human

[25] Eamonn Callan argues that some liberal theories of child-rearing and autonomy have failed to appreciate the second half of this statement. He worries that too much focus has been placed on *choosing* one's own ends and not enough attention has been given to *pursuing* those ends. He writes, "A one-sided focus on the development of capacities for revision of conceptions of the good should be corrected by attention to the value of developing capacities permitting a rational adherence to a conception of the good" (Callan 2002, 118).

[26] As Patrick Frierson notes, "Kant's *Anthropology* culminates in articulating an ideal of autonomy for humanity, a 'destin[y] … to cultivate himself, to civilize himself, and to moralize himself' by resisting a life of 'passively giv[ing] himself over to impulses' and instead 'actively struggling with obstacles (*Anth.* 7:324). Crucially, however, this ideal is not offered as one for individual moral effort; rather, 'the human being must … be *educated* to the good' (*Anth.* 7:325)" (Frierson forthcoming).

being can only become human through education. He is nothing except what education makes out of him" (UP 9:433). And in a lecture on anthropology, he makes it clear that freedom should be the aim of education (and child-rearing in general): "The child must be reared [to be] free, but in a way that it allows others [to be] free and does not, with its freedom, become detrimental to itself. Freedom is the sole condition where the human being can do something good [based] on his own disposition (AF 25:725).

This kind of freedom requires us to rationally deliberate about the value of ends, to choose ends on the basis of reflectively endorsed commitments, and to pursue those ends effectively.[27] In order to achieve this goal, Kant thinks that education (and child-rearing in general) must begin by teaching children to restrain the force of their inclinations. This is an essential component of what Kant calls "culture." In the *Critique of the Power of Judgment*, he defines "culture" as the "production of the aptitude of a rational being for any ends in general (thus those of his freedom)" (*KU* 5:431).

In order to achieve the aim of culturing humanity, Kant identifies two components. First, there is "skill." For instance, in order to pursue the aim of reading books, you must learn how to read. Second, there is "will," which involves the ability to effectively pursue the ends that you are capable of setting. It is here that Kant sees untamed inclinations (or desires) as obstacles that can prevent us from achieving our goals. In a particularly lovely passage, he writes:

> The latter condition of aptitude, which could be named the culture of training (discipline) is negative, and consists in the liberation of the will from the despotism of desires, by which we are made, attached as we are to certain

[27] This goal is ambitious, but it is much less lofty than "unencumbered self" that Michael Sandel puts at the core of Rawls's original position. See Sandel (1984). We agree with Gina Schouten on this point: "While some discussion of autonomy education reads as if the goal is to enable unencumbered choice among a menu of options for how to live, this is neither a goal that we should aim for nor a goal that a liberal education will serve well. What we're after is not unencumbered choice based purely on rational deliberation over personal preferences, but a capacity for independent reflection and judgment" (Schouten 2018, 1074–75). She approvingly cites Eamon Callan, who writes, "The truth is surely that whatever reflection autonomy requires does not demand that we can detach ourselves from all our ends. The requirement is only that we be capable of asking about the value of any particular end with which we currently identify and able to give a thoughtful answer to what we ask" (Callan 1997, 54).

> things of nature, incapable of choosing for ourselves, while we turn into fetters the drives that nature has given us merely for guidance in order not to neglect or even injure the determination of the animality in us, while yet we are free enough to tighten or loosen them, to lengthen or shorten them, as the ends of reason require. (*KU* 5:432)

Even someone who is literate (like Esther) might find herself unable to achieve the end of reading because she is fettered by the despotism of her desires.[28]

This is why Kant, like Rousseau, argues that the education of young children requires us to rein in their desires and to avoid creating bad habits.[29] It is much easier to avoid forming bad habits than it is to break one after it is established (something that any former smoker will confirm):

> Habit is an enjoyment or action which has become a necessity through frequent repetition of the same pleasure or action. There is nothing to which children can accustom themselves more easily, and nothing which one

[28] Kant's point about nature applies directly to our discussion of technology. Nature has given us certain impulses, like hunger or a sex drive, and these are conducive to the preservation of ourselves as an animal species. We need to eat and reproduce in order to preserve ourselves as animals. But Kant worries that these natural inclinations can lead us to heteronomy (and even immorality) if we fail to restrain our desires. This is perfectly in line with our discussion of smartphones and dopamine. In popular culture, dopamine is often described as "the pleasure molecule," but this is a mischaracterization. Although surges of dopamine are often present in pleasant experiences, dopamine has a wide variety of functions in the brain. One of its main jobs is motivating behaviors. As Vaughan Bell explains, "Dopamine is indeed involved in addiction, but it isn't a 'pleasure chemical'. In fact, dopamine has lots of functions in the brain—being involved in everything from regulating movement to the control of attention" (Bell 2013).

Far from being a simple "pleasure chemical," dopamine has even been associated with aversive reactions. See Wenzel et al. (2015). One of dopamine's most important jobs is producing reward-motivated behavior. From an evolutionary perspective, we can see why nature found it fitting for us to get dopamine from sugary or fatty foods. Dopamine is often novelty driven. For our early ancestors, sugar and fat were hard to come by. Now, they are superabundant, and the brain's natural reward system is no longer aligned with our food environment. Instead of helping us survive, the hijacking of the brain's dopaminergic activity can lead to unhealthy, addictive behaviors (as it does in cases of drug addiction, for instance). See Wise and Robble (2020). The same goes for smartphones and the attention economy. The natural reward system that was built to keep us alive is now keeping us glued to our screens.

[29] On this matter of "negative education," Frierson says that "Kant's pedagogy is at its most Rousseauian" (Frierson forthcoming).

> should therefore give to them less, than piquant things; for example, tobacco, brandy, and warm drinks. Breaking the habit later is very difficult and is connected with hardship at first, because through the repeated enjoyment a change in our bodily functions has occurred. The more habits someone has, the less he is free and independent. It is the same with the human being as with all other animals: they always retain a certain propensity for that to which they were accustomed early. (UP 9:463)[30]

Much of what Kant says in the lectures on pedagogy is perfectly in line with our aim in this section. Our concern about children becoming habituated to mobile devices and the attention economy is identical to Kant's concern about other habits (like tobacco).[31]

He even singles out distraction as a worrisome habit that threatens to undermine the child's education:

> Distractions must never be tolerated, least of all at school, for they eventually produce a certain tendency in that direction, a certain habit. Even the most beautiful talents perish in one who is subject to distractions…Then they only hear half of everything, answer wrongly, do not know what they are reading, and so forth. (UP 9:473–74)

He goes on to argue that we must work to preserve our capacity for paying attention: "As concerns the strengthening of attention, it should be noted that this must be strengthened in general…Distraction is the enemy of all education" (*UP* 9:476).

Of course, for Kant, there is an undeniable moral drive behind this kind of education. By weakening the force of inclinations, parents and teachers make it easier for their children to subordinate their self-interested desires

[30] This is quite similar to what he says earlier in the lecture: "Now by nature the human being has such a powerful propensity towards freedom that when he has grown accustomed to it for a while, he will sacrifice everything for it. And it is precisely for this reason that discipline must, as already said, be applied very early; for if this does not happen, it is difficult to change the human being later on" (UP 9:442).

[31] Indeed, when we presented some of these ideas at a conference, there was an undergraduate student (from Generation Z) who asked us whether or not it would be possible for him to build an autonomous relationship with technology given that he grew up with it and that his phone "had already broken his brain."

to the moral law.[32] But it is easy to see how this also promotes personal autonomy more broadly. When it comes to their use of technology, if parents and teachers fail to teach children good habits, they do them a disservice with respect to their future autonomy. Like Esther, they may find themselves setting ends that they cannot successfully pursue.

It is for precisely this reason that so many parents have already become conscientious about their children's use of mobile devices. Ironically, this trend is particularly strong in places like Silicon Valley. Journalist Nellie Bowles addressed this in an article titled "A Dark Consensus About Screens and Kids Begins to Emerge in Silicon Valley." Bowles interviewed parents who work in the tech industry about their household screen rules, and she found that many of them had very harsh words about the effects of screens on children. Chris Anderson, a former editor of *Wired* magazine, compared screen time to addictive drugs like crack cocaine, and another said, "I am convinced the devil lives in our phones and it's wreaking havoc on our children" (Bowles 2018a). Anderson lamented the mistakes he had already made with his own children: "I didn't know what we were doing to their brains until I started to observe the symptoms and the consequences…This is scar tissue talking. We've made every mistake in the book, and I think we got it wrong with some of my kids," Mr. Anderson said. "We glimpsed into the chasm of addiction, and there were some lost years, which we feel bad about" (Bowles 2018a). Bowles said that many of those who work in Silicon Valley have become "obsessed" with what mobile devices are doing to our brains. She says that "no tech homes are cropping up" in the area, and nannies are routinely asked to sign "no phone contracts."[33]

[32] As Frierson (forthcoming) explains, education begins with the negative project of weakening the force of inclinations. It then moves to "discipline," which involves learning how to adhere to rules. Afterwards, Kant tells us to inspire the child with moving stories about moral exemplars, and the child's education culminates finally in what Kant calls the "moral catechism." Frierson stresses the importance of respecting the child's humanity at every stage in the process. When the child is young enough, we do not disrespect their humanity, because they lack the relevant capacities. Later on, we must avoid setting ends for them, and this is why the moral catechism involves something akin to Socratic dialogues which aim to get the pupil to find the moral law through his own use of reason.

[33] Bowles wrote a second piece focusing specifically on the nanny contracts. She writes, "The fear of screens has reached the level of panic in Silicon Valley. Vigilantes now post photos to parenting message boards of possible nannies using cellphones near children. Which is to say, the very people building these glowing hyper-stimulating portals have become increasingly terrified of them" (Bowles 2018b).

Given what we said earlier, this should come as no surprise. Of course the parents in Silicon Valley are concerned about their children using these products. They know how addictive these products are; that's precisely how they were designed. But software developers are not the only ones who have taken notice. Organizations like the American Academy of Pediatrics and the World Health Organization have recommended that children under age two get no screen time at all (except for video calls) and children from two to five should get less than an hour a day (Pappas 2020). In addition to the risk of developmental delays that are associated with screen usage at young ages (speech, socialization, etc.), the Canadian Paediatric Society points out that time away from screens is "critical for developing essential life skills such as self-regulation" (Ponti 2023).

Pediatricians and psychologists are alarmed about children's use of mobile devices for many of the same reasons that we discussed in Chap. 3. They are concerned about the effects that these devices have on, e.g., attention, working memory, and executive function. They are also worried about the effects that these devices have on the mental health of teenagers. This is why the surgeon general recently issued a warning about how social media can "have a profound risk of harm to the mental health and well-being of children and adolescents" (United States 2023). All of the effects described in Chap. 3 are of concern when it comes to children, but there are good reasons to believe that technology poses an even graver threat to children, as their developing brains are especially vulnerable.[34]

[34] In their meta-analysis of 41 studies on the effects of problematic smartphone use (PSU) among children and adolescents, Sohn et al. found troubling associations with many of the harms we described in Chap. 3: "PSU has been consistently associated with measures of poor mental health, in particular relating to depression, anxiety, stress, poor sleep quality, and day to day functional impairment demonstrated by poor educational attainment…Younger populations are more vulnerable to psychopathological developments, and harmful behaviours and mental health conditions established in childhood can shape the subsequent life course" (Sohn et al. 2019, 4, 7). They found a disturbing prevalence of problematic smartphone use among children and young people, describing it as a "widespread problem" (Ibid. 4–5). They claim that nearly one in four young people has a relationship with their smartphone that "mirrors that of a behavioural addiction" (Ibid. 7).

That study was published in 2019, and recent work has shown that the problem is getting worse. A 2022 meta-analysis found an increase in "digital addictions" across the globe. Among other things, they looked at "smartphone addiction" and "social media addiction," and they found that these problems were exacerbated by the COVID-19 pandemic (Trott et al. 2022).

Excessive screen time has been linked to developmental delays in young children and it leads to worse academic performance in older children (Madigan et al. 2019, Sohn et al. 2019). When they are given too much screen time, children exhibit behavior (and brain activity) that "resembles substance use behavior" (Lissak 2018, 149).[35] And just as our knowledge of this threat is growing, children's usage is spiking. A report in 2019 found that children between the ages of 13 and 18 were on screens for an average of 7 hours and 22 minutes per day (and that is excluding the time spent doing schoolwork or homework on screens).[36] Usage subsequently increased during the COVID-19 pandemic.[37] So it is safe to say that we should be concerned about children spending too much time using mobile devices and engaging with the attention economy.

On the flip side, we must recognize, as we did in the previous chapter, that the thoughtful, deliberate use of technology (including mobile devices) can enhance our autonomy. It would be a mistake to raise children who are computer illiterate or who do not know how to use a mobile phone at all. The potentially difficult task for parents and teachers is to promote the kind of thoughtful use of these tools that is conducive to autonomy, while guarding against the compulsive or addictive usage that undermines autonomy.

It may seem difficult to thread this needle, but this is precisely what the research suggests we ought to do. Sziron and Hildt (2018) present empirical evidence arguing for precisely this conclusion, and they link their conclusion to Feinberg's concept of the child's right to an open future:

> A child's right to an open future requires parents, software engineers, designers, policy makers, and marketers understanding the ramifications of digital media for ages 0–5. Children's first interactions with digital media need to be supplemental for developing autonomy, digital citizenship, and culturally responsible use of technology, while avoiding negative privacy implications, the risks of digital divide, and "being like adults." Ethical development of future digital media technology should take into consider-

[35] Lissak (2018) writes, "[A]ddictive screen time use decreases social coping and involves craving behavior which resembles substance dependence behavior. Brain structural changes related to cognitive control and emotional regulation are associated with digital media addictive behavior" (149).

[36] Pappas (2020).

[37] See Trott et al. (2022).

> ation the child, as a child with an open future, in the hopes of maintaining that open future. (2018, 3)

There are various ways that parents and teachers can try to strike this balance. First, the recommendations about screen time vary by age. Although the rules are highly restrictive at young ages (especially for kids two and under), they are more lax for older children. Second, the empirical findings suggest that there are significant benefits when parents and teachers are present during the screen time (co-viewing), as they can engage with the child about the content and supplement their learning with further discussion (Pappas 2020). Third, the quality of the content matters. It has been shown that slow-paced content (like Sesame Street or other PBS programs) can foster learning in young children (Ibid.).

But there is a stark contrast between the kind of content that was made to be slow and educational versus the kind that is meant to maximize engagement. For years, Netflix has kept records of its top ten most popular programs. And there is one program that has consistently broken every record and remained on the most-watched list longer than any other program: CoComelon.[38] Parents who have struggled with the effect this program has on their children are fighting an uphill battle. CoComelon was meticulously designed to capture the attention of young children. The company who produces it runs a lab that tracks the eyeball movement in young children, making notes every time that a child looks away from the program.[39] This should remind us once more of the alignment problem. When content is designed to maximize engagement, it may fail to promote other values that matter to us (such as producing content that is edifying or, at the very least, content that does not stoke the flames of the dopamine-fueled forest fire that is undermining children's capacities).

Thus, the first and most important thing we can do as parents and teachers is to take notice of this issue and realize that it requires us to be more deliberate about how our children and students engage with technology. If we allow them to default into whatever relationship with it they want, we are giving free rein to the software developers and tech companies whose main goal is to maximize the time our children spend looking at screens. If parents and teachers are morally obligated to cultivate autonomy, then they must do more when it comes to mobile devices and the

[38] See Bean (2021).

[39] See Segal (2022). We are grateful to Hanna Gunn for reminding us of this example.

attention economy. Sometimes beneficence means *not* giving someone what they want (or what they *think* they want).[40] This is especially true of children who are not capable of fully understanding the consequences of their choices.

Let us close the section by briefly noting that we, as parents and teachers, acknowledge that one implication of what we have said here is that parents and teachers, who are too often undersupported and overburdened, have yet another burden to bear. But let us also note a larger theme of this chapter and coming chapters: we are not in this alone. Just as parents and teachers have obligations to their children and students to promote their autonomy, others—individually and collectively—have obligations to support them in achieving this end.

5.3.2 Implications for Close Friends

Before moving on, we would like to comment briefly on one last context where duties of beneficence may include a duty to promote digital minimalism in others. Parenting and education are special cases. These are two domains where the cultivation of autonomy is a constitutive aim of the activities in question. But we also believe that the duty to be an attention ecologist could extend to close friendships.[41]

Kant concludes the Doctrine of Virtue with a discussion of friendship.[42] Like Aristotle, Kant thinks that the highest form of friendship involves a moral component.[43] It cannot be a relationship that simply boils down to mutual advantage. His ideal of friendship is "the union of two persons through equal mutual love and respect" (*MS* 6:469). He thinks that friendship should involve the moral perfection of one another: "From a moral point of view it is, of course, a duty for one of the friends to point

[40] When describing duties of beneficence, Kant is explicit about this limiting condition: "It is for them to decide what they count as belonging to their happiness; but it is open to me to refuse them many things that they think will make them happy but that I do not, as long as they have no right to demand them from me as what is theirs" (MS 6:388).

[41] Although we use the word "friend" here, we mean this to apply more broadly to any close relationships (including familial relations).

[42] There is also an extensive discussion of friendship in the Collins lectures. See VE 27:425–31. His discussion in the *Metaphysics of Morals* is more focused on his ideal of friendship, but in the lectures he spends more time discussing the other forms of friendship (friendships of need, taste, or disposition).

[43] Kant's view of friendship is similar to Aristotle's in several ways, but there are some important differences. For further discussion, see Sherman 1987.

out the other's faults to him; this is in the other's best interests and is therefore a duty of love" (*MS* 6:470). Earlier in this chapter, we explained how duties to morally perfect others can be understood in terms of beneficence. Kant understands beneficence as furthering someone else's ends as if they were your own. And given that other people are (or ought to be) committed to morally perfecting themselves, we can promote their ends by helping them along in their journey of perpetual moral progress.

But Kant recognizes that this can be a delicate matter. The other friend might take offense when you point out their flaws. They may see this as a failure of respect. So it is not as if we should all go around telling all of our friends that they use their phones too much. But there are indeed moments where we may be in a position to help. For instance, if someone comes to you complaining about how their phone is interfering with their productivity or sleep, you could help them by showing them ways of reducing their usage (turning off notifications during work, keeping their phone outside the bedroom at night, etc.).

Once again, as a duty of love, you must use discretion when it comes to being an attention ecologist. In some cases, it will be appropriate to help a close friend restructure their relationship with their phone. In other instances, you might be overstepping a boundary and failing to respect their autonomy. This brings us to the topic of the next section. In addition to duties of love, we also have duties of respect. And this will require us to extend our discussion beyond the domains of teachers, parents, and close friends.

5.4 Applications of the Duty II: Never as a Mere Means

We focused above on positive duties associated with attention ecology, specifically as they play out in our relations to those who may not have yet developed their capacities for autonomy (e.g., children), cases where we are in a role that has developing the capacities for autonomy as a legitimate constitutive aim (e.g., teacher), and in cases where we bear an intimate relationship to someone and might be in a special position to legitimately take on the promotion of their self-governance (e.g., close friends).

But the demands of attention ecology do not apply to *only* these relationships which, as Kant would have characterized them, fall under the

banner "love,"[44] that is, duties which consist of adopting others' permissible ends as our own (*MS* 6:450). Indeed, there are relationships that are in some sense thinner than these that still activate the negative side of the duty, which would fall under the banner of "respect." Respect is a duty to refrain from using others as mere means (*MS* 6:450). Importantly, what distinguishes this duty (which falls under the banner of "virtue") from others we will explore in the next chapter (which fall under the banner of "right") is that they must, like duties of love, be accompanied by a practical feeling. That is, the duty includes recognizing others' status as moral agents.

5.4.1 Implications for Employers

Like parents and teachers, employers are in a privileged position to set ends for their employees. They do this all the time by giving them tasks, setting deadlines, telling them when they should arrive at work, when they can leave, etc. But unlike parents and teachers, employers are not necessarily obligated to cultivate the autonomy of their employees. Like everyone else, however, employers *are* morally required to respect the autonomy of their employees. Kant's moral theory requires us to always refrain from using others as mere means. When describing such duties of respect in the *Metaphysics of Morals*, Kant says, "The duty of respect for my neighbor is contained in the maxim not to degrade any other to a mere means to my ends (not to demand that another throw himself away in order to slave for my end)" (*MS* 6:450).

This does not mean that Kant objects to wage labor as such.[45] But his ethical theory does provide us with the resources to condemn conditions

[44] By "love" Kant means love "in a practical sense," that is, the duty here is *not* one that involves the feeling of love, but, as we state above, the taking of others' ends as our own (*MS* 6:448, cf. Gregor 1963, 182). This is because, as Gregor writes, "What the [moral] law commands must always be within our power, and we cannot at will summon up a certain emotion" (Gregor 1963, 182). In the *Critique of Practical Reason*, Kant ties this concept of love to the Gospel's prescription to "Love God above all, and your neighbor as yourself" (5:83). He says that to love God is "to do what He commands *gladly*" and "to love one's neighbor means to practice all duties toward him *gladly*" (Ibid.). Kant recognizes that we do not have perfect control over our feelings. So he describes this as an ideal of holiness that we should strive for, even if we know that we will always fall short of perfection.

[45] In a paper comparing Kant and Marx, Allen Wood explains how wage labor contracts are valid under Kant's account in the Doctrine of Right (*MS* 6:285). But Wood goes on to explain how a Marxist objection could be raised about the potentially exploitative conditions under which a worker agrees to a labor contract. See Wood (2017).

of employment in which the employee is subject to unjust domination. In *Private Government: How Employers Rule Our Lives (and Why We Don't Talk about It)*, Elizabeth Anderson argues that employers' power has become tyrannical as their control over their employees' lives now extends far beyond reasonable boundaries. She points to examples where employers regulate their "employees' diet, clothing, political speech, and use of drugs" (Anderson 2017, 39). She even cites extreme cases where employers deny employees adequate bathroom breaks, forcing them to wear diapers (Anderson 2017, 135).[46]

There is something deeply objectionable about employers having this much control over the lives of their employees.[47] It is one thing for employers to set deadlines for employees, and it is quite another for them to restrict their use of bathrooms or their freedom of speech. To use Kant's definition of respect, we could say that these employers have degraded their employees to mere means; they are being asked to "throw themselves away" to slave away for someone else's ends.

[46] Anderson was referring to workers in the poultry industry, and she also points out that employers threatened to fire workers if they complained about the abysmal conditions. Once again, the complaint here is not about wage labor in and of itself. It is about the degrading or dehumanizing conditions of certain forms of labor. In his book on Marx, Allen Wood offers a somewhat Kantian interpretation of exploitation, as he suggests that wage labor can disrespect and degrade human dignity by taking advantage of workers' vulnerability. See Wood (2020). More recently, Nicholas Vrousalis has used neo-republican concerns with freedom to argue that domination can explain what is morally objectionable about exploitation. See Vrousalis (2013, 2020, 2023). His view differs from Pettit's, however, as Vrousalis takes issue with the vagueness of the "arbitrariness" component of Pettit's definition. Instead, Vrousalis argues that domination involves taking advantage of someone's vulnerability in a way that is disrespectful to them.

In both cases, the objections to exploitation may appear to be divorced from traditional Marxist concerns about distribution or the extraction of surplus value. But as Paul Warren argues, both accounts involve distributive components, so he thinks it would be a mistake to exclude the distributive dimension from the theory of exploitation. See Warren (2015).

Our account of workplace domination does not require a general view of exploitation, however. Instead, all that we need is an explanation of why it is wrong for employers to undermine the autonomy of their employees by weakening their capacities (i.e., disrespecting their humanity). As we explain in our 2022 paper, there is a sense in which this complaint is narrower than the Marxist objection. It does not condemn *all* forms of wage labor. On the other hand, our complaint is broader in the sense that it raises a concern that could apply to forms of labor that could persist even if the workers collectively owned the means of production (e.g., if one's co-op uses a distracting, anxiety-inducing messaging platform). See Aylsworth and Castro (2022).

[47] It brings to mind Othello's request: "Leave me but a little to myself."

We would like to extend Anderson's analysis to the ways that employers require employees to use technology. As we mentioned earlier in our discussion of external constraints in Chap. 3, many employers require employees to remain in constant contact through digital channels (email, text message, WhatsApp, Slack, etc.), and this inevitably requires workers to have an unhealthy relationship with their mobile devices. They find themselves responding to messages after work hours, checking email when they are trying to sleep, and so on.

The external constraint is not the only drag on autonomy here. It is not merely that employers are requiring employees to use devices in ways that they do not want to use them. We are also concerned about the downstream effects that this has. Imagine having a job that requires you to take a pill every day. The pill makes you work harder for eight hours, but it has some unwelcome side effects. When you sit down to read a novel in your free time, you find yourself struggling to read. You also find that the pill causes feelings of alertness in the evenings, and this interferes with your sleep.

Many workers would rightly object to such an arrangement. It is not just that the employer is requiring you to take a pill (an external constraint), it is that the pill is weakening your capacities more generally. In an effort to squeeze more productivity out of you, the employer is making it harder for you to read and get enough sleep. They are making you less capable of setting and pursuing your own ends. This is a violation of the duty to respect others as ends in themselves and refrain from using them as a mere means. Obviously, requiring workers to respond to emails after working hours isn't nearly as egregious as denying workers bathroom breaks to such an extent that they end up in diapers. We are not defending such a false equivalence. Nevertheless, both situations may be seen as instances of domination and can thus be understood as expressing disrespect for the employees' humanity.[48]

In the previous section, we explained what Kant meant by the duty to refrain from "giving scandal." The idea was to avoid leading others into temptation, preventing them from acting in ways that are inconsistent

[48] Domination, which Pettit (2014) defines as being subject to the arbitrary will of another, should be understood as a spectrum. Almost no one is entirely free from domination, but some workers are more dominated than others. The workers' freedom of exit is of the utmost importance. Highly skilled workers, who have abilities that are in short supply and high demand, have much greater freedom of exit than Anderson's poultry workers.

with their moral obligations. If we recognize that other people are obligated to cultivate their capacities (a component of their duty to respect humanity in their own person), then we would be wronging them if we did something that made it more difficult for them to fulfill that obligation.

This is not identical to the case described above where parents refused to foster the autonomy of their child. That was an instance of someone failing to fulfill a positive obligation. Parents have a positive obligation to cultivate their children's capacities. Employers are under no such obligation. But they are, nevertheless, required to refrain from *undermining* the autonomy of their employees. And by requiring employees to have a problematic relationship with their mobile devices, employers are violating this duty.

There are several ways to avoid such work arrangements. Workplaces could establish firmer boundaries and explain that employees do not need to respond to emails on weekends or in the evening. The problem, of course, is that companies who enact such a lax policy might worry that they are giving an edge to their competitors who require their employees to be in communication 24/7. But, as Cal Newport argues in *A World Without Email*, there are reasons to believe that companies are not achieving maximal efficiency by requiring workers to be in constant communication. He writes, "When employees are miserable they perform worse" (2021, 38). He points to the high costs of burnout and turnover to argue that companies may fare *better* if they allowed employees to have some boundaries when it comes to technology. So this might not be such a significant worry after all.

Nevertheless, it may be difficult to make these changes at an individual level. Even when it comes to internal triggers, many of us struggle to establish healthy boundaries when it comes to work emails. There is also social pressure, which does not necessarily come from one's supervisors or managers. For instance, we might worry about our teaching evaluations if we know that our colleagues are responding to student emails over the weekend or late at night. For this reason, we must think of some of these issues as collective action problems—ones where it takes more than an individual effort to address the concern. It is just like James Williams said, you cannot solve a pollution problem by wearing a gas mask.[49] We will address the collective perspective in Chap. 7, and we will discuss policies

[49] Hari 2022, 101.

like the French labor law in Chap. 6. But first, we should discuss other instances of the duty to respect the autonomy of others.

5.4.2 Implications for Developers and Deployers

Throughout the book, we have discussed how tech companies and developers have leveraged the psychology of addiction in order to get us hooked on their products. As Eyal (2014) explains, their aim is to make us so habituated to their use that we unlock our phones without giving it a moment's thought. They do this by taking advantage of various psychological vulnerabilities: the desire for social approval, feelings of urgency, fear of missing out, etc. We saw in Chap. 3 how this often leads to deleterious consequences for those who end up using mobile devices in excess.

Within the framework of Kantian ethics, it is easy to explain why this action is morally wrong. When tech companies or developers act this way, they are pursuing their own ends (maximizing profit by means of maximizing engagement) with little or no regard for our capacity to set and pursue our own ends. We are treated as mere means. When he was at Google, Tristan Harris repeatedly explained his concerns to his co-workers, and he ultimately resigned when he realized that he would not be able to make significant changes. Companies like Google (whose products like YouTube depend crucially on maximally capturing our attention) are held hostage by a business model that is incompatible with concerns for the autonomy of their users.

Harris gave his first presentation on this topic to Google in 2013. It was titled, "A Call to Minimize Distraction and Respect Users' Attention." He raised many of the worries that we have discussed in this book, and he takes careful note of the psychological vulnerabilities that are exploited by tech companies. In one slide, he acknowledges how the conflict of interest presents an obstacle to making industry-wide changes:

> The problem is … successful products compete by exploiting these vulnerabilities, so they can't remove them without sacrificing their success and growth, creating an arms race that causes companies to find more reasons to steal people's time. A tragedy of the commons that destroys our common silence and ability to think. (Harris 2013)

In our discussion of beneficence, we mentioned the parity between perfect and imperfect duties. Following Wood (1999), we claimed that you have

an imperfect duty to promote certain ends (e.g., the happiness of others), and you have a perfect duty to refrain from setting ends that are antithetical to the ends you are required to adopt. When tech companies adopt the end of pursuing profit by maximizing engagement, they fail to respect the autonomy of the people who use their products. They make it harder for us to be fully autonomous. We are more distracted and anxious; we get less sleep; we find it harder to focus; etc. We struggle to fulfill our duty to cultivate humanity and perfect our rational capacities.

What's more, Kantian ethics has the resources to explain why such actions would be wrong even if they failed to achieve their aim. Because Kant's ethics does not fundamentally ground the wrongness of actions in their consequences, it can condemn the disrespect of an action regardless of its failure to cause actual harm. By contrast, some people may hold the view that Rosa wrongs Hanlon only if Rosa's action makes Hanlon worse off in some respect than he otherwise would have been (if the action had not been performed).[50] But this principle is subject to some well-known counterexamples.[51] For instance, Woodward (1986) presents an example of a Black man who tries to buy an airplane ticket but the airline refuses to sell him one because it has a policy of racial discrimination. The plane goes on to crash and everyone who was on board dies.[52] Obviously, we cannot say that the man is worse off than he otherwise would have been. If the airline had let him buy a ticket, he would have died. Nevertheless, it seems clear that he was wronged. Woodward suggests that the wrongness may consist in the fact he was treated unfairly or unjustly. Or, on Kantian grounds, we could say that this is a failure of respect.[53]

[50] This is roughly the formulation given by Gardner (2015).

[51] Gardner (2015) calls it the "counterfactually worse off condition." This principle is given attention in many discussions of the non-identity problem. See Woodward (1986), Harman (2004), Gardner (2015), and Purves (2019). Woodward provides several compelling counterexamples. For instance, he mentions Viktor Frankl, whose book, *Man's Search for Meaning*, chronicles the many ways his life was enriched by certain things he endured in the Holocaust. Nevertheless, it is obvious that the Nazis wronged Frankl. A revised version of this example is discussed by Harman (2004) and Gardner (2019).

[52] Woodward (1986) uses this as a counterexample to the "counterfactually worse off" account of wronging, but it is a modified version of an example from Adams (1972). Thanks to Molly Gardner for directing us to these papers.

[53] Of course, we do not want to imply that consequentialists lack the resources to explain the wrongness (or blameworthiness) of the airline's action. Consequentialists could say that the action was wrong (or blameworthy) on the grounds that it had negative expected utility.

So even if tech companies have not managed to undermine your autonomy in any way, they have still treated you with disrespect insofar as they developed and marketed products with total disregard for your capacity to set and pursue your own ends.[54] The duty to refrain from "giving scandal" requires you to avoid tempting others to transgress the moral law. You must not make it harder for them to fulfill their obligations. If you place temptations in their path, you have acted wrongly whether they succumb or not.

There are many things that tech companies can do to improve in this regard. Tristan Harris made four suggestions in his presentation:

We can design to reduce the volume and frequency of interruptions.
We can design to be **respectful** about *when* to notify users—let it wait, unless it's important.
We can design to **keep users focused**, by putting temptations further away when they're trying to accomplish goals.
We can batch up notifications & messages into **digests by default**, instead of one at a time.[55]

We must recognize, as Harris did, that this is an uphill battle. Tech companies realize that these suggestions are at odds with their aim of maximizing our screen time. This is akin to asking tobacco companies to switch to products that do not contain nicotine. In order to make significant progress, tech companies may have to abandon this business model altogether. For instance, they could move to models where users pay monthly fees instead of generating all of their revenue from advertising. There will surely be costs associated with such transitions, but that very well may be what morality demands.

5.5 Conclusion

In this chapter, we have defended the existence of a duty to be an attention ecologist—a duty to promote digital minimalism in others. We followed Kant in adopting the distinction between duties of love and duties

[54] And even if you are lucky enough to have escaped the autonomy-undermining forces of the attention economy at an individual level, we explain in Chap. 7 how it is likely that you have been harmed by the effects of the attention economy on collective autonomy.

[55] Harris (2013, emphasis in original).

of respect. The former can be understood as obligations to cultivate the virtue of digital minimalism in others. In particular, we discussed parents and teachers, who have special obligations to cultivate the autonomy of their children and students. This also applies to close friends who wish to promote the ends of their loved ones. Duties of respect, by contrast, are obligations to refrain from tempting others by luring them into the vice of technological heteronomy. We paid special attention to the duties of respect for employers and developers.

Of course, duties of respect could be applied more broadly. Parents and teachers are equally obligated to refrain from "giving scandal." Kantian ethics is committed to the idea that we must *always* treat others as ends in themselves and *never* as a mere means. When it comes to the use of technology, this means that we have positive duties to promote good habits in others and we have negative duties to refrain from tempting others into a compulsive relationship that undermines their autonomy.

As we explained in the introduction, these are all duties of virtue. They concern the moral obligations of individuals, and this excludes the possibility of intervention by the state. In the next chapter, we turn to duties of right as we explore the possibility of legislation that might improve this situation. We do not think that legislation should be seen as a panacea, and we should exercise caution when it comes to enacting such laws. But it is worth exploring the moral grounds of such laws, as Kant's political philosophy provides an interesting perspective on the importance of the state's role in preserving our freedom.

References

Adams, Robert M. 1972. Must God create the best? *The Philosophical Review* 81: 317–332.

Albertzart, Maike. 2019. A Kantian solution to the problem of imperceptible differences. European Journal of Philosophy 27, Nr. 4 (December): 837–851. https://doi.org/10.1111/ejop.12456

Anderson, Elizabeth. 2017. *Private government: How employers rule our lives (and why we don't talk about it)*. Princeton: Princeton University Press.

Atterton, Peter. 2007. A duty to be charitable? A Rigoristic Reading of Kant. *Kant Studien* 98: 135–155.

Austin, Michael. 2007. *Conceptions of parenthood: Ethics and the family*. New York: Routledge.

Aylsworth, Timothy, and Clinton Castro. 2022. On the duty to be an attention ecologist. *Philosophy and Technology* 35: 1–22.

Bacin, Stefano. 2021. My duties and the morality of others: Lying, truth and the good example in Fichte's normative perfectionism. In *Fichte's system of ethics: A critical guide*, ed. Stefano Bacin and Owen Ware. New York: Cambridge University Press.

Baer, Drake. 2017. What if instead of controlling your life, your phone helped you live it? *Medium*. https://medium.com/thrive-global/what-if-instead-of-controlling-your-life-your-phone-heped-you-live-it-f20a979c2767. Accessed 21 July 2023.

Barrow, Robin, and Ronald Woods. 2006. *An introduction to the philosophy of education*. New York: Routledge.

Bean, Travis. 2021. The unstoppable 'CoComelon' just broke its own Netflix record. *Forbes*. https://www.forbes.com/sites/travisbean/2021/06/25/the-unstoppable-cocomelon-just-broke-its-own-netflix-record/. Accessed 23 June 2023.

Bell, Vaughan. 2013. The unsexy truth about dopamine. *The Guardian*. https://www.theguardian.com/science/2013/feb/03/dopamine-the-unsexy-truth. Accessed 23 June 2023.

Besche-Truthe, Fabian. 2022. The global trajectories of compulsory education: Clustering sequences of policy development. In *Global pathways to education: Cultural spheres, networks, and international organizations*, ed. Kerstin Martens and Michael Windzio, 65–96. Global Dynamics of Social Policy. Cham: Springer International Publishing.

Blustein, Jeffrey. 1982. *Parents and children: The ethics of the family*. Oxford: Oxford University Press.

Bowles, Nellie. 2018a. A dark consensus about screens and kids begins to emerge in Silicon Valley. *The New York Times*. https://www.nytimes.com/2018/10/26/style/phones-children-silicon-valley.html. Accessed 22 June 2023.

———. 2018b. Silicon Valley nannies are phone police for kids. *The New York Times*. https://www.nytimes.com/2018/10/26/style/silicon-valley-nannies.html. Accessed 23 June 2023.

Brighouse, Harry. 1998. Civic education and liberal legitimacy. *Ethics* 108: 719–745.

———. 2000. *School choice and social justice*. New York: Oxford University Press.

———. 2006. *On education*. New York: Routledge.

Brighouse, Harry, and Adam Swift. 2014. *Family values: The ethics of parent-child relationships*. Princeton: Princeton University Press.

Callan, Eamonn. 1997. *Creating citizens: Political education and liberal democracy*. New York: Clarendon Press.

———. 2002. Autonomy, child-rearing, and good lives. In *The moral and political status of children*, ed. D. Archard and C.M. Macleod, 118–141. New York: Oxford University Press.

Cohen, Alix. 2016. The role of feelings in Kant's theory of moral education. *Journal of Philosophy of Education* 50: 511–523.
Denis, Lara. 2001. *Moral self-regard duties to oneself in Kant's moral theory*. New York: Routledge.
Eyal, Nir. 2014. *Hooked: How to build habit-forming products*. New York: Penguin.
Feinberg, Joel. 1980. The child's right to an open future. Reprinted in *Freedom and fulfillment: Philosophical essays*, ed. Joel Feinberg. Princeton: Princeton University Press.
Fernandez, Christian. 2010. Education and diversity: Two stories of a liberal dilemma. *Public Affairs Quarterly* 24: 279–296.
Fichte, Johann Gottlieb. 2005. *Fichte: The system of ethics*, ed. Daniel Breazeale and Guenter Zöller. Cambridge: Cambridge University Press.
Friedman, Milton. 1975. *Capitalism and freedom*. Chicago: Chicago University Press.
Frierson, Patrick. 2016. Making room for children's autonomy: Maria Montessori's case for seeing children's incapacity for autonomy as an external failing: making room for children's autonomy. *Journal of Philosophy of Education* 50: 332–350.
———. 2021. The moral philosophy of Maria Montessori. *Journal of the American Philosophical Association* 7: 133–154.
———. forthcoming. Kant on education. In *The Oxford handbook of Kant*, ed. Anil Gomes and Andrew Stephenson. Oxford: Oxford University Press.
Galston, William A. 1995. Two concepts of liberalism. *Ethics* 105: 516–534.
Gardner, Molly. 2015. A harm-based solution to the non-identity problem. *Ergo, an Open Access Journal of Philosophy* 2.
———. 2019. When good things happen to harmed people. *Ethical Theory and Moral Practice* 22: 893–908.
Gregor, Mary J. 1963. *Laws of freedom; A study of Kant's method of applying the categorical imperative in the Metaphysik Der Sitten*. Oxford: B. Blackwell.
Gutmann, Amy. 1987. *Democratic education*. Princeton: Princeton University Press.
———. 1995. Civic education and social diversity. *Ethics* 105: 557–579.
Hari, Jari. 2022. *Stolen focus: Why you can't pay attention – And how to think deeply again*. New York: Crown.
Harman, Elizabeth. 2004. Can we harm and benefit in creating? *Philosophical Perspectives* 18: 89–113.
Harris, Tristan. 2013. A call to minimize distraction and respect users' attention. https://digitalwellbeing.org/googles-internal-digital-wellbeing-presentation-transcript-and-slides/. Accessed 23 June 2023.
Herman, Barbara. 2007. *Moral literacy*. Cambridge: Harvard University Press.
Jamieson, Dale. 2007. When utilitarians should be virtue theorists. *Utilitas* 19 (2): 160–183.
Johnson, Robert. 2011. *Self-improvement: An essay in Kantian ethics*. New York: Oxford University Press.

Kant, Immanuel. 1997. *Lectures on ethics*. Trans. Peter. Heath, ed. Peter Heath and J.B. Schneewind. New York: Cambridge University Press.
———. 2005. Religion within the boundaries of mere reason. In *Religion and rational theology*. Trans. Allen Wood and George di Giovanni. New York: Cambridge University Press.
———. 2007a. *Anthropology, history, and education*, ed. Günter Zöller and Robert B. Louden. Cambridge University Press.
———. 2007b. *Critique of the power of judgment*. Trans. Paul Guyer and Eric Matthews. New York: Cambridge University Press.
———. 2008a. Critique of practical reason. In *Practical philosophy* Trans. Mary J. Gregor. New York: Cambridge University Press.
———. 2008b. *Groundwork of the metaphysics of morals*. In *Practical philosophy* Trans. Mary J. Gregor. New York: Cambridge University Press.
———. 2008c. *The metaphysics of morals*. In *Practical philosophy* Trans. Mary J. Gregor. New York: Cambridge University Press.
———. 2013. *Lectures on anthropology*. Tran. Robert B. Louden, Allen W. Wood, Robert R. Clewis, and G. Felicitas Munzel. Cambridge: Cambridge University Press.
Korsgaard, Christine. 1996. *Creating the kingdom of ends*. New York: Cambridge University Press.
Levinson, Meira. 1999. *The demands of liberal education*. New York: Oxford University Press.
Lissak, Gadi. 2018. Adverse physiological and psychological effects of screen time on children and adolescents: Literature review and case study. *Environmental Research* 164: 149–157.
Louden, Robert. 2000. *Kant's impure ethics: From rational beings to human beings*. New York: Oxford University Press.
Macedo, Stephen. 1990. *Liberal virtues*. New York: Oxford University Press.
Madigan, Sheri, Dillon Browne, Nicole Racine, Camille Mori, and Suzanne Tough. 2019. Association between screen time and children's performance on a developmental screening test. *JAMA Pediatrics* 173: 244–250.
Newport, Cal. 2021. *A world without email: Reimagining work in an age of communication overload*. New York: Portfolio.
Nozick, Robert. 1974. *Anarchy, state, and Utopia*. New York: Basic Books.
O'Neill, Onora. 1975. *Acting on principle*. New York: Cambridge University Press.
———. 1980. The moral perplexities of famine relief. In *Matters of life and death*, ed. Tom Regan, 260–298. Philadelphia: Temple University Press.
O'Neill, Onora, and William Ruddick. 1979. *Having children: Philosophical and legal reflections on parenthood: Essays*. New York: Oxford University Press.
Pappas, Stephanie. 2020. *What do we really know about kids and screens?* American Psychological Association. https://www.apa.org/monitor/2020/04/cover-kids-screens. Accessed 23 June 2023.

Pettit, Philip. 2014. *Just freedom: A moral compass for a complex world*. New York: W. W. Norton & Company.

Ponti, Michelle. 2023. Screen time and preschool children: Promoting health and development in a digital world. *Paediatrics & Child Health* 28: 184–192.

Puls, Heiko. 2016. Kant's justification of parental duties. *Kantian Review* 21: 53–75.

Purves, Duncan. 2019. Harming as making worse off. *Philosophical Studies* 176: 2629–2656.

Roth, Klas, and Paul Formosa. 2019. Kant on education and evil—Perfecting human beings with an innate propensity to radical evil. *Educational Philosophy and Theory* 51: 1304–1307.

Sandel, Michael. 1984. The procedural republic and the unencumbered self. *Political Theory* 12: 81–96.

Schouten, Gina. 2018. Political liberalism and autonomy education: Are citizenship-based arguments enough? *Philosophical Studies* 175: 1071–1093.

Segal, David. 2022. A kid's show juggernaut that leaves nothing to chance. *The New York Times*. https://www.nytimes.com/2022/05/05/arts/television/cocomelon-moonbug-entertainment.html. Accessed 23 June 2023.

Sensen, Oliver. 2022. Universal law and poverty relief. *Ethical Theory and Moral Practice* 26: 177–190.

Sherman, Nancy. 1987. Aristotle on friendship and the shared life. *Philosophy and Phenomenological Research* 47: 589–613.

Smeyers, Paul. 2012. Chains of dependency: On the disenchantment and the illusion of being free at last (part 1). *Journal of Philosophy of Education* 46: 177–191.

Sohn, Sei Yon, Philippa Rees, Bethany Wildridge, Nicola J. Kalk, and Ben Carter. 2019. Prevalence of problematic smartphone usage and associated mental health outcomes amongst children and young people: a systematic review, meta-analysis and GRADE of the evidence. *BMC Psychiatry* 19: 356.

Sziron, Monika and Elisabeth Hildt. 2018. Digital media, the right to an open future, and children 0–5. *Frontiers in Psychology*. https://www.frontiersin.org/articles/10.3389/fpsyg.2018.02137/full. Accessed 23 June 2023.

Trott, Mike, Robin Driscoll, Enrico Iraldo, and Shahina Pardhan. 2022. Changes and correlates of screen time in adults and children during the COVID-19 pandemic: A systematic review and meta-analysis. *EClinicalMedicine* 48: 101452.

United States. Public Health Service. Office of the Surgeon General. 2023. *Social media and youth mental health. The U.S. Surgeon General's advisory*. Rockville: U.S. Department of Health and Human Services, Public Health Service.

Vallentyne, Peter. 2002. Equality and the duties of procreators. In *The moral and political status of children*. New York: Oxford University Press.

Vrousalis, Nicholas. 2013. Exploitation, vulnerability, and social domination: Exploitation, vulnerability, and social domination. *Philosophy & Public Affairs* 41: 131–157.

———. 2020. Socialism unrevised: A reply to Roemer on Marx, exploitation, solidarity, worker control. *Philosophy & Public Affairs* 49: 78–109.
———. 2023. *Exploitation as domination: What makes capitalism unjust*. Oxford: Oxford University Press.
Warren, Paul. 2015. In defense of the Marxian theory of exploitation: Thoughts on Roemer, Cohen, and others. *Social Theory and Practice* 41: 286–308.
Wenzel, Jennifer M., Noah A. Rauscher, Joseph F. Cheer, and Erik B. Oleson. 2015. A role for phasic dopamine release within the nucleus accumbens in encoding aversion: A review of the neurochemical literature. *ACS Chemical Neuroscience* 6: 16–26.
Wisconsin v. Yoder. 1972. 406 U.S. 205.
Wise, Roy A., and Mykel A. Robble. 2020. Dopamine and addiction. *Annual Review of Psychology* 71: 79–106.
Wolff, Jonathan. 1991. *Robert Nozick: Property, justice and the minimal state*. Stanford: Stanford University Press.
Wood, Allen. 1999. *Kant's ethical thought*. New York: Oxford University Press.
———. 2004. *Kant*. New York: Wiley-Blackwell.
———. 2009. Duties to oneself, duties of respect to others. In *The Blackwell guide to Kant's ethics*, ed. Thomas E. Hill Jr., 229–252. Oxford: Wiley-Blackwell.
———. 2016. *Fichte's ethical thought*. New York: Oxford University Press.
———. 2017. Marx and Kant on capitalist exploitation. *Kantian Review* 22: 641–659.
———. 2020. *Marx*. New York: Taylor & Francis.
Woodward, James. 1986. The non-identity problem. *Ethics* 96: 804–831.

CHAPTER 6

The Duty to Promote Digital Minimalism in Others II: Duties of Right

Therefore, if a certain use of freedom is itself a hindrance to freedom in accordance with universal laws (i.e., wrong), coercion that is opposed to this (as a hindering of a hindrance to *freedom) is consistent with freedom in accordance with universal laws, that is, it is right.*
—Kant, *Metaphysics of Morals* 6:231

I used to work at Facebook and joined because I think Facebook has the potential to bring out the best in us. But I am here today because I believe that Facebook's products harm children, stoke division, weaken our democracy and much more. The company's leadership knows ways to make Facebook and Instagram safer and won't make the necessary changes because they have put their immense profits before people.
Congressional action is needed.
They cannot solve this crisis without your help.
—Frances Haugen, Facebook whistleblower (This quote comes from Frances Haugen's testimony to the US Senate's Sub-Committee on Consumer Protection, Product Safety, and Data Security. She gave this statement on October 5, 2021 after blowing the whistle on *Facebook* by releasing a trove of documents about the harmful effects of Facebook's efforts to maximize engagement. She revealed how *Facebook* was repeatedly confronted with the negative consequences of their actions, but she claims that whenever they were faced with a choice between the public good and maximizing profits, *Facebook* always chose the latter.)

T. Aylsworth, C. Castro, *Kantian Ethics and the Attention Economy*,
https://doi.org/10.1007/978-3-031-45638-1_6

6.1 Introduction

Thus far, all of the obligations we have described in the book are what Kant refers to as "duties of virtue." That is to say, they are moral obligations that individuals have to perform certain actions (or refrain from performing them)[1] and to do so for the right reasons. One of the first things that students typically discover about Kant's ethics is his emphasis on the idea of moral worth—acting from the motive of duty. The *Groundwork* is, by far, Kant's most well-known work of moral philosophy, and it begins with a discussion of this concept.

He famously argues not only that we should *do* what is morally right, but we should do it *because* it is morally right. If you do what the moral law requires, but you do it only because it is in your self-interest, then you rob your action of moral worth. Kant gives the example of the shopkeeper who charges a fair price only because it is a prudent business practice, and he contrasts this with the idea of "acting from the motive of duty" (*G* 4:397). He reiterates the point with another example of a philanthropist whose beneficence is motivated by the warm feelings he gets from helping others. Such an action conforms with duty and is therefore deserving of "praise and encouragement," but Kant says that it "has nevertheless no true moral worth" (*G* 4:398). By contrast, he thinks we can clearly see moral worth in the action of the philanthropist who helps others without the help of any inclinations whatsoever. He is motivated by pure reason alone; he does it simply because it is the right thing to do.

Many commentators like Allen Wood (1999) and Christine Korsgaard (1996) have rightly pointed out how people are often misled by Kant's discussion of moral worth.[2] He is not, by any means, condemning actions motivated by sympathy or telling us to rid ourselves of charitable dispositions like compassion. On the contrary, he tells us in the *Metaphysics of*

[1] Kant distinguishes between actions that are obligatory, those that are forbidden, and those that are merely permitted. He writes, "[A] categorical imperative is a law that either commands or prohibits, depending upon whether it represents as a duty the commission or omission of an action. An action that is neither commanded nor prohibited is merely *permitted*" (*MS* 6:223).

[2] As Wood (2018) puts it, "[R]eaders form a common image of Kant as the representative of a kind of inhuman moralism, an attitude of moral rigidity and hostility to all the more tender human emotions. This image is conspicuously at odds with things Kant actually says in other writings, such as the *Metaphysics of Morals*" (iv).

Morals and his lectures on ethics that we are morally *required* to cultivate love and sympathy for others (*MS* 6:401, 6:546 VE 27:66).[3]

For our discussion in this chapter, the important thing to notice about such "duties of love" is that they would be impossible to enforce or coerce. Try as you might, you can never compel someone to feel a particular feeling or adopt an intention. You are in control of your own intentions, but you cannot direct the maxims of other agents. This means that it is impossible (both in theory and in practice) to enforce duties of virtue.

Duties of right, on the other hand, are ones that *can* be coerced. They involve acts that you can compel someone else to perform. Kant sometimes talks about our ability to control the *external* dimension of someone's behavior (e.g., what they do with their body) as opposed to our inability to affect their incentives or maxims (the *internal* dimension). Coercion applies only to external duties.

Kant shows how both of these are often implicated in the same situation: "So it is an external duty to keep a promise made in a contract; but the command to do this merely because it is a duty, without regard for any other incentive, belongs to *internal* lawgiving alone" (*MS* 6:220). If someone is contractually obligated to repay a loan, we can certainly compel them to fulfill this obligation (a duty of right) by means of external coercion. Kant also thinks that they ought to keep their promise from the motive of duty. But there is nothing we can do to force them to comply with *that* duty.

He summarizes the distinction in precisely those terms: "What essentially distinguishes a duty of virtue from a duty of right is that external constraint to the latter kind of duty is morally possible, whereas the former is based only on free self-constraint" (*MS* 6:383). Of course, once we start talking about the possibility of legitimate coercion, we have entered the territory of political philosophy. Kant thinks of the state (and political philosophy in general) in terms of the possibility of legitimate (i.e., rightful) coercion. He begins the Doctrine of Right with a discussion of private

[3] In *Groundwork* I, Kant focuses on "acting from the motive of duty" in order to draw out the categorical imperative through reflection on this idea. As David Velleman puts it, Kant's aim is "to derive the content of our obligations from the very concept of an obligation" (2006, 16). By reflecting on the purity of the motive of duty (the absence of a self-interested incentive), Kant comes to the conclusion that the maxim must be suitable as a universal law.

right (property, contracts, and status) and he then goes on to develop his conception of the state (public right).[4]

In this chapter, we are interested in Kant's political philosophy insofar as it offers some distinctive resources for thinking about the possibility of regulating the attention economy. We want to reiterate, however, that it is important to refrain from thinking of regulation as the answer to all of these problems. We should be cautious when it comes to state interventions, as there is a considerable risk of doing more harm than good.

By exercising the state's coercive power, we inevitably place a constraint on people's freedom. And, as we will see shortly, Kant thinks that the preservation of freedom is the one and only legitimate foundation for the state. This means that the state should use its power exclusively for the sake of *promoting* freedom. At first glance, this may seem like a contradiction. How can the state promote freedom by restricting freedom? In the next section, we will explain how Kant resolves this tension. Once we have established the basic framework of Kant's political philosophy, we will then put his view to work in order to defend the legitimacy of policies that would regulate the attention economy in various ways.

6.2 Duties of Right Versus Duties of Virtue

It would be reasonable to expect Kant's political philosophy to be a straightforward extension or application of his moral theory. After all, this is how political theory is often done. Utilitarians like Bentham argued that political philosophy should simply be an extension of his moral rule: the principle of utility. He thought that our moral obligations involve maximizing overall happiness, and he advocated for political principles that he believed would accomplish that aim. But this is not how Kant approaches political philosophy. Many readers of the *Metaphysics of Morals* have been disappointed to discover that Kant does not simply apply the categorical imperative to politics.[5] Instead, he uses an entirely different foundational principle.

This is the "Universal Principle of Right." And although it resembles the categorical imperative insofar as it requires us to respect the freedom

[4] He concludes with a brief discussion of "cosmopolitan right" which deals with peaceful relations between states. Kant was well ahead of his time when he advocated for things like freedom of movement and a "league of nations" in "Toward Perpetual Peace."

[5] See, for example, Brown (1962). Cf. Ripstein (2009, 2).

of others in accordance with a universal law, it is certainly not identical to it:

Universal Principle of Right: "Any action is right if it can coexist with everyone's freedom in accordance with a universal law, or if on its maxim the freedom of choice of each can coexist with everyone's freedom in accordance with a universal law." (*MS* 6:230)

As we explained earlier, Kant insists that the state should not be in the business of promoting happiness. His reasoning for this claim is similar to his argument against grounding ethics in the concept of happiness. He believes that human nature is too variable. People have different conceptions of what will make them happy, and he thinks it would be grievously immoral for someone to force their conception of happiness on you. This is especially true of the state. We saw in Chap. 2 how Kant refers to that kind of paternalistic government as "the greatest despotism thinkable" (TP 8:290–91). You should be free to pursue your own version of happiness, no matter what that may be (provided it does not involve wronging others).

What Kant takes to be truly universal is the freedom of rational agents. As beings endowed with rationality, all human beings are free to set and pursue their own ends.[6] For precisely this reason, in his political

[6] Some readers might be troubled by the metaphysical implications of this claim. Many of Kant's critics (even those who are sympathetic to his moral theory) worry that Kant made a disastrous mistake by grounding his ethical theory in his concept of transcendental freedom. In the *Critique of Pure Reason*, Kant explains how this concept of freedom can be defended only by means of transcendental idealism's distinction between appearances and things in themselves. Although objects of appearance (phenomena) are strictly determined by causal laws (and thus unfree), things in themselves (noumena) are not subject to causal necessity and thus it is possible to think of them as spontaneous first causes. In his discussion of the third antinomy, Kant makes it clear that the intelligibility of concepts like praise and blame depend crucially on this distinction. See, for instance, his discussion of the malicious lie at A554/B582.

Unfortunately, many commentators think that Kant's view is incoherent. As O'Neill puts it: "Kant is revered for his unswerving defense of human freedom and respect for persons, and for his insistence that reason can guide action. He is also reviled for giving a metaphysically preposterous account of the basis of freedom" (1989, ix). It would be outside the scope of our project here to explain why we think this issue is less worrisome than critics have argued. In short, the idea is that it is possible to maintain Kant's conclusions about the value of rational agency and autonomy without committing to any metaphysically extravagant claims. For further discussion, see Aylsworth (2020).

philosophy, Kant argues that freedom is the one and only innate right, and he builds his entire political theory on the foundation of freedom. He says, "Freedom (independence from being constrained by another's choice), insofar as it can coexist with the freedom of every other in accordance with universal law, is the only original right belonging to every man by virtue of his humanity" (*MS* 6:237).

Given the significant overlap between his moral and political principles, some readers have found it odd that Kant did not simply make the categorical imperative the bedrock of his political theory. But there are good reasons for Kant's decision to articulate a distinct principle. As we explained in the introduction, the central distinction between duties of right and duties of virtue is the enforceability of the former. The categorical imperative may function perfectly well as a moral principle that tells us which actions are right or wrong, but it says nothing whatsoever about coercion.[7] Kant's moral law may tell us that a particular action is wrong, but it does not tell us whether or not it would be appropriate for us to coerce the agent in question.

For instance, imagine you need help moving some heavy furniture. You ask your friend if you can count on her when the truck arrives. She says yes even though she has no intention of helping. The categorical imperative can explain why her action is wrong, but we cannot conclude from this condemnation that it would be appropriate for the state to force her to help you move. If, on the other hand, she comes to your house on moving day and steals something that belongs to you, then it would be perfectly appropriate for the state to force her to return the stolen property. Both actions involve moral wrongs. One of them justifies coercion, and the other does not. Thus, we need a principle that provides some guidance when it comes to questions about the legitimacy of coercion.

This is where the principle of universal right comes into play. According to the principle, the deciding factor is the freedom of choice. Someone may be rightly coerced whenever they act in such a way that they constrain your freedom—your one innate right. Perhaps the most obvious example of this would be someone restricting your freedom of movement (e.g., by physically restraining you). It is very easy to see how this constrains your

[7] Arthur Ripstein (2009) provides an illuminating discussion of the difference between the two principles and the rationale behind the distinction. As he puts it, "The possibility of coercion is not contained in the Categorical Imperative as it is formulated in either the *Groundwork* or the *Critique of Practical Reason*" (366).

freedom. By interfering with your person, the wrongdoer is making it impossible for you to pursue your own ends. This principle would extend to any assault on your person. For similar reasons, Kant thinks the state can legitimately enforce your property rights. If someone were to steal your car, this would also constrain your ability to set and pursue your own ends.[8]

Like Locke, Kant foregrounds property rights in his political theory. But Kant departs from Locke, who sees property initially as a relation between a person and a thing. Instead, Kant recognizes that property is fundamentally a relation between persons.[9] To have a property right is to have a claim that someone else refrain from using (or damaging) an object that is rightfully yours. Kant thinks that the possession of personal property is essential to the exercise of our freedom. Given the kinds of creatures we are, we are prone to pursuing ends that require us to make use of more than just our bodies. For instance, we could not write this book without using our computers.

Since Kant recognizes that property is, at bottom, a relation between persons, he believes that property rights cannot exist in any substantive way in the "state of nature." In the absence of a government, a person's claim to property can only be "provisional." Without the coercive power of state institutions, there is nothing to guarantee the kinds of robust property ownership that Kant thinks are essential to the exercise of our freedom.

It is for this very reason that Kant argues we are morally required to leave the state of nature and establish a "civil" or "rightful" condition. To remain in the state of nature is to be in a state of permanent insecurity regarding our fundamental right to freedom. In the absence of laws, you may not be able to "wrong" anyone in the legal sense, but Kant claims that everyone acts wrongly by refusing to establish a state that can safeguard their basic right to freedom. He says "But in general they do wrong in the highest degree by willing to be and to remain in a condition that is not rightful, that is, in which no one is assured of what is his against violence" (*MS* 6:307–8).

[8] In order to move from persons to property, Kant adds an additional postulate about how the use of external objects precludes other people using them in the same way. He writes, "It is therefore an a priori presupposition of practical reason to regard and treat any object of my choice as something which could objectively be mine or yours" (*MS* 6:250).

[9] Kant ridicules what he calls the "guardian spirit" theory of property rights. See *MS* 6:260. Cf. Ripstein (2009, 22).

Unlike Hobbes, who argues that we should leave the state of nature in order to guarantee our basic security, Kant thinks we should leave the state of nature to protect our freedom. Of course, Kant shares Hobbes's concern with the absence of security in the state of nature. The key difference, however, is their understanding of the tradeoff that one accepts when establishing a civil condition. Hobbes sees the state of nature as a condition of total freedom; every human being has complete liberty (in virtue of natural right) to do whatever it takes to survive.[10] But Hobbes also thinks that this is a state of constant war and insecurity. By establishing a sovereign, he believes that we accept constraints on our freedom in exchange for the robust protection of our security.

Kant has a very different understanding of the exchange. Freedom, as Kant understands it, cannot possibly exist in the state of nature. Freedom requires the capacity to exercise rights that can be guaranteed only by means of institutions with coercive power. Thus, Kant does not see the state as a restriction on our freedom; he sees it as the only conceivable way for a group of human beings to enjoy their innate right to freedom.[11]

This does, however, set up a tension. On the one hand, Kant fundamentally thinks of the state in terms of coercion, and he recognizes coercion as a restriction of freedom. On the other hand, the state exists solely for the purpose of safeguarding freedom.[12] These two claims may seem inconsistent at first glance. But Kant resolves this tension by explaining that the state should only hinder the freedom of those who are hindering the freedom of others: "Therefore, if a certain use of freedom is itself a hindrance to freedom in accordance with universal laws (i.e., wrong), coercion that is opposed to this (as a *hindering of a hindrance to freedom*) is consistent with freedom in accordance with universal laws, that is, it is right" (*MS* 6:231).[13] This means that the state is actually *promoting*

[10] Similarly, Locke refers to the state of nature as a "state of perfect freedom" and a "state of liberty" (1988, Chap. 2).

[11] Once again, the influence of Rousseau is evident. Kant concurs with Rousseau about the idea that large groups of humans can achieve freedom only by means of constraints. Like Rousseau, Kant also argues that we should see the state as legislating not from the perspective of any particular citizen but from the notion of a general will.

[12] In the introduction to the Doctrine of Right, Kant puts it like this: "Now whatever is wrong is a hindrance to freedom in accordance with universal laws. But coercion is a hindrance or resistance to freedom" (*MS* 6:231).

[13] See Ripstein (2009) for an excellent explanation and defense of Kant's political philosophy. Ripstein writes, "Coercion is objectionable where it is a hindrance to a person's right to freedom, but legitimate when it takes the form of hindering a hindrance to freedom. To stop you from interfering with another person upholds the other's freedom" (55).

freedom when it uses its coercive power. When the state prevents someone from stealing your car, it is promoting *your* freedom by hindering the *thief's* hindrance of your freedom.

This conclusion should not be terribly controversial. After all, some applications of Kant's political theory may resemble the harm principle that Mill defends in *On Liberty*.[14] Mill's principle certainly yields the same result when it comes to the examples mentioned above (someone stealing your property or restricting your movement). But Kant's principle of equal freedom has some advantages over Mill's harm principle, which may lead to both false positives and false negatives. As Arthur Ripstein argues, there are some actions that are clearly wrong even though they do no harm, and there are actions that are harmful but should be permitted by the state.[15]

With its emphasis on the centrality of freedom,[16] Kant's political philosophy provides an excellent foundation for our argument in the next section as we defend the possibility of regulations concerning the technologies we have discussed throughout the book. Rather than defending such policies on the grounds that they promote well-being, we will argue that the status quo is incompatible with maximal human freedom in accordance with universal laws. By giving free rein to employers and tech companies, the state is allowing them to hinder the freedom of others. Policies that constrain their actions can thus be seen as promoting freedom by hindering a hindrance.

[14] We mentioned Mill's principle in Chap. 2: "The only purpose for which power can be rightly exercised over any member of a civilized community against his will, is to prevent harm to others" (1988, 9).

[15] See Ripstein (2006). His example of the former is that of a "harmless trespass." He describes a situation in which he breaks into your home and takes a nap on your bed. But he takes incredible care to make sure that the action does not harm you in any way (e.g., he cleans everything, uses his own sheets, etc.). Even if the action does no harm, it is clearly wrong. For cases of the latter, he gives examples of what Feinberg (1988) calls "fair contests." You build a product but then a competitor makes a better one. You close your hotel, and this harms the owner of the neighboring restaurant. And so on. These actions cause harm to others, but they are not wrong.

[16] There is something very promising about a political theory grounded in freedom. As Pettit (2014) points out, it has been difficult to build consensus around ideas of social justice. Intuitions about just social arrangements often conflict. Some may agree with Nozick and argue that justice requires minimal constraints and the smallest state possible. Others prefer Rawls and believe that justice requires a redistributive state in order to promote fair equality of opportunity. Pettit suggests that we might make more progress by grounding our political theory in freedom (Pettit 2014, xviii).

6.3 The Argument to Be an Attention Ecologist from Duties of Right

There are several important conclusions to take away from the previous section. First, we are dealing with a political theory that takes freedom as its primary concern. Kant makes freedom the foundational principle of his entire political philosophy. While other moral theories could defend regulations on the grounds that they promote various interests or maximize happiness, these justifications are not available to Kantians. The state exists to promote and preserve our freedom. Second, the state must exercise coercive power by compelling people to comply with duties of right. Whenever someone wrongfully constrains the freedom of another person, the state is justified in using its power to prevent such actions. Finally, Kant resolves the apparent tension by arguing that state coercion is not incompatible with freedom. On the contrary, human beings cannot possibly exist in a rightful condition with one another unless the state is there to safeguard our freedom and defend our rights.[17]

With those conclusions in hand, we are in a better position to put forward the argument for regulation. And it is important to recall some of the empirical details from previous chapters. In particular, we must remember the ways that certain agents (or groups of agents) constrain the freedom of others by weakening our capacity to set and pursue our own ends. For instance, we will revisit our discussion of employers, schools, and software developers. Each of those examples presents us with situations where someone's freedom is externally constrained by the actions of another agent. Thus, they are all candidates for regulation. Here is the general argument for the duties concerning regulation:

$P1_d$ The state should promote freedom.
$P2_d$ If $P1_d$, then the state should promote freedom by using legislation to restrict hindrances to freedom.
$P3_d$ Employers, developers, and other agents hinder our freedom with respect to our use of technology.

[17] As Allen Wood puts it, "You become free from constraint by me only if there is a power that coercively prevents me from constraining you, a power that coercively protects your rightful freedom. Freedom for all depends, therefore, on the existence of a coercive power to which all of us are subject" (Wood 2017).

C1$_d$ The state should promote freedom by using legislation to restrict the ability of employers, developers, and other agents to hinder our freedom with respect to our use of technology.

In the previous section, we developed a better understanding of why Kant is committed to the first two premises. But more should be said about violations of the principle of right. What exactly does it mean to hinder the freedom of others?

Under the broadest possible interpretation, this claim about freedom is clearly implausible. Some critics have gone so far as to claim that it is *impossible* to justify the state in terms of universal freedom, as certain freedoms necessarily come at the expense of others. For instance, G.A. Cohen (and several others) argue that one person's property rights restrict the ability of other people to make use of the land.[18] Given that Kant begins the Doctrine of Right with a defense of property rights, this is surely not what he meant.

Kant's principle of equal freedom does not require us to build a society in which every citizen is unconstrained in every respect. That kind of freedom is indeed incoherent. After all, by simply standing in a particular place, one person constrains the physical location of another. As Ripstein (2009) points out, the equal freedom principle must avoid defining freedom in terms of our capacities to pursue *particular* interests. Kant's idea is to build a society in which no person's *general* ability to set and pursue ends is constrained by another person. Once we take someone's interests into account, it is obvious that other people will constrain their ability to pursue that particular end.

If you think of freedom in terms of your ability to own Van Gogh's *Starry Night*, then you will see yourself as lacking this freedom. This would also mean that the Museum of Modern Art (which owns the painting), has a freedom that you lack. But this is simply a mistaken conception of freedom. Kant takes the principle of equal freedom to mean that each person is equally capable of setting and pursuing her own ends. It does not mean that every person has equal capacities relative to some particular interest. For Kant, freedom means having the ability to set and pursue the end of art collection if you so choose. You can collect whatever art you are

[18] In his defense of Kant's equal freedom principle, Ripstein (2009) mentions Cohen's objection, and he gives a list of its historical antecedents (Samuel Taylor Coleridge, Henry Sidgwick, etc.). See Ripstein (2009, 32). Cf. Cohen (1981).

able to buy on the market. That is a freedom that each person can enjoy equally (even if there are inequalities in wealth that allow some people to own more art than others). The principle of equal freedom would indeed be incoherent if it meant that each person has the right to own a Van Gogh.

Freedom means that your will is not constrained by the will of other people. But, once again, there is a risk of overstating the requirements of freedom, particularly with the issue at hand. Employees are constrained by the will of their employers. They are told what hours to work, what tasks to perform, etc. But Kant does not condemn wage labor as a wrongful restriction of freedom. On the contrary, he puts wage contracts on his list of permissible contracts in his theory of private right (*MS* 6:285). In the previous section, we saw how someone could wrong you by touching your person or property without your consent. If, however, you had given someone permission to use your car, then they would not have wronged you at all. Kant applies the same idea to wage labor as one person consents to letting someone else direct their productive powers in exchange for a specified amount of money.[19] Kant thinks that this arrangement does not undermine the freedom of the worker because the worker freely consented to the contract, and she could withdraw from this labor contract and begin working for another employer.[20]

So we should not see all wage labor contracts as wrongful restrictions of freedom. But some contracts are indeed prohibited by Kant's doctrine of right. Most notably, Kant thinks that contracts of serfdom and slavery are incompatible with a rightful condition. For Kant, a contract is an agreement between two persons that unites their will. It is predicated on the idea that each person has rights and duties toward one another. This is why it would be impossible to enter into a slavery contract. Contracts are possible only if both parties have rights. To become someone's slave is to

[19] He writes, "A contract of letting of work on hire (*locatio operae*), that is, granting another the use of my powers for a specified price (*merces*). By this contract the worker is hired help (*mercennarius*)" (*MS* 6:285).

[20] Those who see wage labor as exploitative might point out that Kant is being shortsighted here. Although the workers are free to work for *this* or *that* employer, they have no choice but to work for *some* employer. Allen Wood (2017) argues that this makes room for what Kant refers to as "general injustice" (652). This rules out one way of condemning labor contracts. That is, it would be a mistake to say that *this* employer is wronging *this* employee in terms of constraining her freedom. Instead, we could say that the class of employers (i.e., capital) wrongs the working class by constraining their freedom. In making this claim, Wood goes beyond what Kant says, but he uses Kantian resources to explain what he thinks is wrong with exploitative wage labor.

give them more than just your labor for some specified period; it is to give them your *person* and forfeit all of your rights. This is incompatible with your innate right to freedom. The right to freedom is not something that you can give away (nor is it something that can be taken from you). Slavery does more than constrain your will with regard to a particular end; it undermines your ability to set and pursue ends more broadly.

This goes back to the distinction from earlier. Freedom should not be understood in terms of any *particular* end but in terms of the general capacity to set and pursue ends. A typical labor contract will constrain what you do at a certain time, but it does not undermine your capacities in the general sense. On the other hand, if the labor contract requires you to do something that is at odds with your capacity to be autonomous, then it may indeed violate a duty of right and be subject to constraint.

We do not, by any means, want to suggest that labor contracts are equivalent to slavery. The point here is simply that Kant's doctrine of right explains what is wrong with contracts that undermine someone's future freedom. Slavery is the most extreme form of that. By entering into bondage, someone forfeits all of their future freedom. They give up their entire personhood, and Kant thinks that such a forfeiture can never be permitted by a political system grounded in our innate right to freedom. In general, compromising someone's capacity to set and pursue their own ends is something that is at odds with Kant's principle of right.

As we saw in Chap. 3, there are reasons to believe the problematic use of mobile devices has just such an effect on us. It not only affects what we do with particular moments of time, it also affects our capacities and our authenticity more broadly. Thus, when employers require employees to remain in constant communication by means of mobile devices (even after hours), they may be contributing to the deterioration of the employee's autonomy. In his explanation of the principle of right, Ripstein (2009) explains how it also entails a definition of wrong: "uses of a person's powers to set and pursue ends in ways that are not consistent with the ability of others to use their powers" (2009, 363). If employers require employees to use their mobile devices in ways that undermine their autonomy, then this would fall within the domain of the doctrine of right. Thus, there are not only duties of virtue at stake (as we argued in the previous chapter), there are also duties of right.

This means that policymakers would be justified in using the state's coercive power to regulate the behavior of employers. Once again, we will avoid making specific policy recommendations. Our job as moral

philosophers is to explore the ethical framework through which different policies could be justified. We believe that policy experts are in a better position to work out the precise details of laws that would be entailed by these moral commitments.

For instance, in Chap. 3, we mentioned the French labor law that requires employers to negotiate after hours email policies. This has come to be known as the "right to disconnect." Policies like this are in line with what we are suggesting here. Legislation of this kind can be seen as hindering the employer's ability to hinder the freedom of their employees. But we do not have a strong view about the details. For example, the French law applies only to companies that have 50 or more employees. In Canada, the rule applies to companies with more than 25 employees. We leave it to policymakers to determine the size and nature of such exceptions. Perhaps there are good reasons to carve out exceptions for small businesses or for certain industries. But we believe that the status quo justifies having the state intervene in order to protect workers' autonomy from being undermined by overly demanding tech policies.

The same could be said for public education. As we explained in the previous chapter, there are serious concerns about exposing young children to screens, and it is important to make informed choices about the content that is used in educational settings. For instance, there is evidence to support the claim that there are advantages to reading on paper rather than reading on screens.[21] Schools should be mindful of this research when using digital technologies to promote literacy. They should also be cautious about using "gamification" to increase student engagement. There is a danger of giving too much weight to certain metrics (e.g., hours spent with learning materials) at the expense of other values. Kant's political philosophy gives us some reasons to be concerned about the ways that we are using digital technology in the classroom. Careful, thoughtful deployment of technology can enhance the autonomy of students, but there is also the danger of undermining students' capacities through excessive reliance on technology.

Of course, the state has control of public education, so there is already a regulatory framework in place that could be used to implement more thoughtful technology policies. But, in the US, regulatory control of

[21] See, for example, Clinton (2019). It often depends heavily on the presentation of the text on the screen. Comprehension and retention are significantly affected by the presence of other disruptive elements (e.g., notifications) on the screen. See Furenes et al. (2021).

schools is highly decentralized. Educational policies are issued at nearly every level of government (federal, state, and local school district). On the one hand, this will make it difficult to standardize tech policies. On the other hand, this gives individual citizens greater control, as they can discuss these issues at their local school board meetings. However the policies are implemented, the idea behind them would be to keep an eye on the autonomy of students when using technologies (especially mobile devices) in the classroom.

This year (2023), our home state of Florida enacted a law that bans the use of social media in public schools. Students are now unable to access any social media platforms through the school's internet. Of course, students can turn off their Wi-Fi and continue to access social media through their phones' mobile data, but this does create a minor roadblock to accessing social media at school. What's more, teachers are now allowed to regulate the use of phones in their classrooms. It is especially encouraging to see that the law requires students in grades 6 through 12 to receive lessons about the "negative effects of social media on mental health, including addiction; the distribution of misinformation on social media; [and] how social media manipulates behavior…" (FL HB 379). We think that educational policies of this kind are a step in the right direction. By teaching students about the dangers of social media at an early age, we can hope that children learn to be more careful and deliberate about their engagement with the attention economy.[22]

But the discussion of children's autonomy should extend beyond public education as well. States also regulate daycare facilities. They limit the size of classes; daycares are subject to state oversight, licensing, and more. Some states (including Florida) also have rules about the use of screen time in daycare facilities. Florida does not permit children under two to be exposed to any "electronic media," and it limits media time for older children to two hours per day.[23] This may not go far enough, however.

[22] Of course, there are also reasons to be skeptical about the efficacy of these efforts. The inefficacy of the D.A.R.E. program comes to mind. Research continues to show that the D.A.R.E. program did not bring about any noticeable reduction in the use of tobacco, alcohol, or drugs. See West and O'Neal (2004). There is an important difference between these programs, however. Education about social media does not require students to abstain entirely. It simply educates them about the dangers of social media in the hopes that they can avoid its most harmful aspects. But, as we have explained, it is not easy to encourage moderate use of things that are known to be addictive.

[23] See Fla. Admin. Code r. 65C-22.001(7).

Florida's policy gives a rather coarse-grained definition of "electronic media" as it encompasses "television, videos, movies, or computer games," and we ought to have more specific definitions.

As we learned in the previous chapter, not all content has the same effect on the developing brains of small children. Policies governing screen time in schools and daycares should be updated in order to take account of the empirical findings about the effects of the attention economy on the brain. Given the current verbiage of the Florida daycare policy, it would be permissible to give a two-year-old child *two hours of tablet time*. But we should do more to protect the autonomy of small children from the harms associated with this kind of use. It is our duty as guardians of the child's "right to an open future." This obligation extends beyond parents and teachers, it is something that the state ought to do to protect and foster the autonomy of children from the many things that threaten to undermine it.

Finally, we would like to consider the possibility of regulating the tech industry itself. We began by discussing employers, schools, and daycares, but each of these are instances of people deploying technologies that already exist. It may be worth considering what could be done upstream of these implementations. What, if anything, should the state do to regulate the companies that produce, disseminate, and market the technologies that capture our attention?

Once more, we think it is important to exercise caution. Freedom of speech is vitally important, and there are great dangers associated with allowing governments to moderate the content of the attention economy. Exceptions can be made (and have been made) for speech that incites violence or that violates some other law (e.g., child pornography). But outside of those narrow exceptions, there are very good reasons for prohibiting the state from restricting the exercise of free speech in any way. In "Theory and Practice," Kant says that "freedom of the pen" is "the sole palladium of the people's rights."[24] Without free speech, people cannot possibly defend their rights.

As we will explain in greater detail in the next chapter, social media platforms have become notorious for spreading all kinds of false and inflammatory content. This happens because people are prone to diving into the rabbit hole of algorithmically curated recommendations, and the

[24] The word "palladium" is used to refer to something that offers protection. It is a reference to the statue of Pallas (Athena), which was thought to protect the city of Troy.

platforms tend to direct users toward increasingly radical ideas. This is part of why so many people (possibly as many as 30 million Americans) believe the earth is flat.[25] It also helps explain the rise of vaccine hesitancy, as an alarming number of people have come to believe all kinds of false claims about vaccines.[26] Once someone dips their toes into the waters of vaccine skepticism on social media, they are quickly pulled into the vortex of misinformation. It begins with false claims linking vaccines to autism (a claim that has been repeatedly discredited),[27] and then it moves to even more ludicrous suggestions about vaccines containing microchips.

As harmful as these false beliefs might be, there are still very good reasons for preventing the state from interfering with people's right to create and spread this kind of misinformation. Of course, the social media platforms themselves are not bound by the First Amendment in the same way. The state is strictly prohibited from restricting free speech, but platforms have every legal and moral right to set their own terms of service and to see to it that users comply with those terms. They own the platforms that are hosting the content, so they are permitted to take down content as they see fit.

But they are financially disincentivized from doing so. The conspiracy rabbit hole is incredibly lucrative for social media companies, as they are known to keep users engaged for hours on end.[28] As of now, platforms like YouTube permit people to post videos about vaccines or about the flat earth conspiracy, but they simply add a disclaimer linking to articles about vaccines and science. In other cases, they go some way toward taking

[25] This figure certainly sounds too big to us, but it comes from a nationwide survey of over 1000 Americans. They found that 10% of Americans believe the world is flat. They also found that 12% believe the moon landing was faked and 9% believe that vaccines contain microchips. See Hamilton (2022). Conspiracy theories of this kind have floated around for decades, but what makes the flat earth theory worthy of our focus here is its rise can be attributed almost entirely to YouTube. At a flat earth convention, 39 out of 40 respondents said that they discovered the flat earth theory through YouTube recommendations. The single outlier found it through his son, who found it through YouTube. See Mirsky (2020).

[26] There is good evidence to support the connection between vaccine hesitancy and social media. See Puri et al. (2020), Wilson and Wiysonge (2020), and Clark et al. (2022). It is also discussed extensively in Fisher (2022).

[27] There are far too many studies to cite here. See, for example, Hviid et al. (2019). The initial study, which purported to establish the link between the MMR vaccine and autism, was totally discredited. It was redacted by the *Lancet* and the physician responsible was stripped of his license to practice medicine. See Omer (2020).

[28] See Fisher (2022).

down content that violates their rules (whatever those rules might be).[29] But we should not give this power to the state. It is one thing to give companies the right to take down content hosted on their own servers, but it is quite another to give this power to the government.

So we will not defend such policies here. We do not think it would be productive (or morally defensible) to advocate for state moderation of online content. But that is not to say that the state should do nothing to regulate the tech industry. Once again, the state has a particularly strong interest in protecting children from the effects of mobile devices and social media. Even if we think that fully capable adult humans should be free to make their own choices with respect to the attention economy, we may not feel the same way about young children. The state has already made similar rules about other activities that are harmful and addictive. Children are not permitted to gamble, smoke, or drink alcohol. The state respects the choices of autonomous adults who choose to do these things, but it does not allow children to do the same.

Florida is not the only state that has taken action to limit children's access to social media. Utah has taken even more dramatic steps in this direction. While Florida's recent law pertains only to social media and mobile devices in school, the Utah law applies to children's use of social media *in general*. Social media companies are now required to get permission from parents before allowing a child to create an account. Parents are also given access to that account, and children are prohibited from using social media between the hours of 10:30 pm and 6:30 am. What's more, Utah's law bans social media companies from using "practices, designs, or features" that are known to promote addiction in children. We are in favor of policies like these as well, and Kant's political philosophy provides a helpful justificatory framework for them. These policies hinder the capacity of tech companies to hinder the freedom of children.

It may seem striking that so many of our policy recommendations involve children. But this restriction of our focus is intentional. There are several reasons for giving special legislative attention to children. First, as we explained in previous chapters, children are uniquely vulnerable to the harms of the attention economy. Their capacities have not been fully formed. Their brains are particularly susceptible to being rewired by the

[29] Though they often do not dedicate nearly enough resources to this. See Fisher (2022) for discussion. For an illuminating and incisive philosophical discussion of the ethics and epistemology of content moderation, see Frost-Arnold (2023).

surges of dopamine they receive from media content. In Chap. 3, we explained the importance of the frontal lobe's executive function when it comes to capacities that matter for autonomy. Executive function refers to a set of capacities including self-control, following directions, and staying focused. This skill first emerges in children between the ages of three and five, but it continues to develop in adolescence.[30] This is the first reason why it is so important to focus on protecting the autonomy of children. They are especially vulnerable.

Second, we should be extremely cautious about interfering with the free choices of fully autonomous adults. Kant's political philosophy (and liberal political philosophy more generally) rarely permits us to interfere with citizens' choices (provided that they are not harming others), even if we believe that they are failing to do what is in their own self-interest. Morally, paternalism is inconsistent with respect for someone's humanity, and politically, it is wrong to coerce someone in order to promote their own well-being. Kant's framework does allow the state to intervene when one person is undermining someone else's freedom (as we argued in the case of employment), but, absent interference with the freedom of other citizens, the state should take a hands-off approach to interfering with the free choices of autonomous adults. But children are not fully autonomous, so they are not wronged by policies that restrict their options, especially when those restrictions are put in place to promote their future autonomy.

In an interview with the *New York Times,* Governor Spencer Cox, who signed Utah's bill, draws on precisely this distinction. The government can do certain things to protect children even though it cannot do the same thing for adults:

> We've done our homework on this one. We've spent time with parents and children, all across the state, and there is a general consensus and acknowledgment that social media and access to these devices is causing harm. Significant harm.
>
> If you look at the increased rates of depression, anxiety, self-harm since about 2012, across the board but especially with young women, we have just seen exponential increases in those mental health concerns. Again, the research is telling us over and over and over again that it is not just correlated, but it's being caused, at least in part, by the social media platforms…
>
> Again, we have a longstanding tradition in our country of drawing lines around ages for brain development when it comes to certain activities. We

[30] See Best and Miller (2010).

> don't let kids smoke or drink or drive a car before certain ages, because we know the danger and the damage that is being done there, and the science will back that up.
>
> If I could wave the magic wand and have all adults spend less time on these devices, social media platforms, I would love to be able to do that. But that isn't something I could do. It's not something I'm comfortable doing, and it's not something that sits nicely within the general legal tradition of our country. (Coaston 2023)

The comparison to cigarettes is particularly instructive. Not only does the government prevent children from purchasing cigarettes, tobacco companies are not permitted to target their advertisements to children.[31] For decades, tobacco companies pushed back against public health efforts that aimed to convince people that cigarettes are harmful and addictive. The current conversation about tech companies should sound eerily familiar.

In 1970, Richard Nixon signed a bill banning TV advertisements for cigarettes. Tobacco companies were not too troubled by the law, however. The FCC's fairness doctrine required TV stations to give equal time to public service announcements about the dangers of smoking.[32] Those statements were effectively bringing about a reduction in smoking, so the tobacco industry simply moved to other media to market its products.

By the 1990s, dozens of states were suing the tobacco companies. The attorneys general of more than 40 states brought suits against them, and one of their contentions was that tobacco companies were marketing cigarettes to children. Joe Camel was an especially stark example of this practice. Eventually, the tobacco companies settled the lawsuits and they agreed to a set of restrictions that prevent them from targeting children with their advertisements. They also agreed to pay billions of dollars (in perpetuity) to compensate states for the added medical costs that result from smoking. Some of the settlement money is also being used to fund anti-tobacco initiatives and anti-smoking public service announcements.

Once again, there are striking parallels with the current situation facing tech companies. Not only has the public become increasingly aware of the fact that mobile devices and social media are harmful and addictive, some groups have begun to bring lawsuits against tech companies. In January 2023, Seattle Public Schools filed a lawsuit against TikTok, Instagram,

[31] See Castro and Pham (2020) for a similar comparison.
[32] See Whiteside (1970).

YouTube, Facebook, and Snapchat, alleging that social media has done considerable harm to the mental health of their students. As of now, more than 40 other school districts across the country have joined Seattle by filing suits against social media companies. Families have brought separate lawsuits making similar allegations. The Seattle lawsuit alleges that tech companies "have successfully exploited the vulnerable brains of youth, hooking tens of millions of students across the country into positive feedback loops of excessive use and abuse."[33] It also claims that "Youths are particularly susceptible to Defendants' manipulative conduct because their brains are not fully developed, and they consequently lack the same emotional maturity, impulse control, and psychological resiliency as other more mature users" (Ibid.).

It is difficult to say how such lawsuits will fare. Similarly, the political prospects of legislation that regulates the tech industry may be dim. Regardless of how the aim is accomplished (whether by judicial decisions or legislation) there is a strong justification for state intervention. By giving tech companies the unfettered ability to market their addictive products to children, we are allowing them to hinder their freedom. We open the door to rampant manipulation in ways that are inconsistent with the principle of universal freedom. Every human has a right to equal freedom in accordance with universal laws.

6.4 Conclusion

In this chapter, we showed how Kant's political philosophy could be used to justify certain state interventions that would regulate the attention economy. According to Kant's doctrine of right, the state should use its coercive power only for the sake of promoting freedom. When it restricts certain activities (such as the marketing of addictive products to children), we should understand these policies as a "hindrance of a hindrance to freedom." By requiring employers to negotiate tech policies with employees, we promote the freedom of workers to be free from an unhealthy relationship with their mobile devices—one that would undermine their autonomy more generally.

Government regulation is not a panacea. As we will see in the next chapter, there are collective social problems that require the cooperation of millions of people. We cannot fix everything by passing a few laws. But

[33] See Seattle School District No. 1 (2023).

if we believe that the state has an obligation to protect the freedom of its citizens, then we must recognize the implications this has for the tech industry and the products that undermine our autonomy. Kant thinks of freedom as the one innate right of all rational beings. To neglect the duties of right that stem from this innate right is to imperil our most precious capacities. We should never allow a human being, endowed with an inalienable right to freedom, to become a mere tool of someone else's will. Not only do we have moral duties to refrain from treating people in such a way, we have political obligations to create the social conditions that prevent such mistreatment from occurring.

References

Aylsworth, Timothy. 2020. Bolstering the keystone: Kant on the incomprehensibility of freedom. *Archiv für Geschichte der Philosophie* 102: 261–298.

Best, John R., and Patricia H. Miller. 2010. A developmental perspective on executive function. *Child Development* 81: 1641–1660.

Brown, Stuart M. 1962. Has Kant a philosophy of law? *Philosophical Review* 71: 33–48.

Castro, Clinton, and Adam K Pham. 2020. Is the attention economy noxious? *Philosophers' Imprint* 16: 1–13.

Clark, Shannon E., Megan C. Bledsoe, and Christopher J. Harrison. 2022. The role of social media in promoting vaccine hesitancy. *Current Opinion in Pediatrics* 34: 156–162.

Clinton, Virginia. 2019. Reading from paper compared to screens: A systematic review and meta-analysis. *Journal of Research in Reading* 42: 288–325.

Coaston, Jane. 2023. The Republican governor of Utah wants to spare kids from their phones. *The New York Times*. https://www.nytimes.com/2023/07/11/opinion/spencer-cox-social-media-utah.html. Accessed 17 July 2023.

Cohen, G.A. 1981. Freedom, justice and capitalism. *New Left Review* I/126: 3–16.

Feinberg, Joel. 1988. *Offense to others*. Reprint ed. New York: Oxford University Press.

Fisher, Max. 2022. *The chaos machine: The inside story of how social media rewired our minds and our world*. New York: Little, Brown and Company.

Furenes, May Irene, Natalia Kucirkova, and Adriana G. Bus. 2021. A comparison of children's reading on paper versus screen: A meta-analysis. *Review of Educational Research* 91: 483–517.

Frost-Arnold, Karen. 2023. *Who Should We Be Online?: A Social Epistemology for the Internet*. Oxford, New York: Oxford University Press.

Hamilton, Lawrence. 2022. Conspiracy vs science: A survey of U.S. public beliefs. University of New Hampshire: Carsey School of Public Policy. https://carsey.

unh.edu/publication/conspiracy-vs-science-a-survey-of-us-public-beliefs. Accessed 17 July 2023.
Hviid, Anders, Jørgen Vinsløv Hansen, Morten Frisch, and Mads Melbye. 2019. Measles, mumps, rubella vaccination and autism. *Annals of Internal Medicine* 170: 513–520.
Kant, Immanuel. 1997. *Lectures on ethics*. Trans. Peter Heath, ed. Peter Heath and J. B. Schneewind. New York: Cambridge University Press.
———. 1998. *Critique of pure reason*. Cambridge: Cambridge University Press.
———. 2008a. Groundwork of the metaphysics of morals. In *Practical philosophy*. Trans. Mary J. Gregor. New York: Cambridge University Press.
———. 2008b. The metaphysics of morals. In *Practical philosophy*. Trans. Mary J. Gregor. New York: Cambridge University Press.
———. 2008c. On the common saying: That may be correct in theory, but it is of no use in practice. In *Practical philosophy*. Trans. Mary J. Gregor. New York: Cambridge University Press.
Korsgaard, Christine. 1996. *Creating the kingdom of ends*. New York: Cambridge University Press.
Locke, John. 1988. *Two treatises of government*. Cambridge: Cambridge University Press.
Mill, John Stuart. 1988. *On liberty*. Indianapolis: Hackett Publishing Company.
Mirsky, Steve. 2020. Flat earthers: What they believe and why. *Scientific American*. https://www.scientificamerican.com/podcast/episode/flat-earthers-what-they-believe-and-why/. Accessed 17 July 2023.
O'Neill, Onora. 1989. *Constructions of reason*. Cambridge: Cambridge University Press.
Omer, Saad. 2020. The discredited doctor hailed by the anti-vaccine movement. *Nature*. https://www.nature.com/articles/d41586-020-02989-9. Accessed 17 July 2023.
Pettit, Philip. 2014. *Just freedom: A moral compass for a complex world*. New York: W. W. Norton & Company.
Puri, Neha, Eric A. Coomes, Hourmazd Haghbayan, and Keith Gunaratne. 2020. Social media and vaccine hesitancy: New updates for the era of COVID-19 and globalized infectious diseases. *Human Vaccines & Immunotherapeutics* 16: 2586–2593.
Ripstein, Arthur. 2006. Beyond the harm principle. *Philosophy & Public Affairs* 34: 215–245.
———. 2009. *Force and freedom: Kant's legal and political philosophy*. Harvard University Press.
Seattle School Districts No 1 v. Meta Platforms et al. 2023. https://storage.courtlistener.com/recap/gov.uscourts.wawd.317950/gov.uscourts.wawd.317950.1.0.pdf. Accessed 17 July 2023.
Velleman, David. 2006. *Self to self*. Cambridge: Cambridge University Press.

West, Steven L., and Keri K. O'Neal. 2004. Project D.A.R.E. Outcome effectiveness revisited. *American Journal of Public Health* 94: 1027–1029.
Whiteside, Thomas. 1970. Cutting down. *The New Yorker*. https://www.newyorker.com/magazine/1970/12/19/the-fight-to-ban-smoking-ads. Accessed 17 July 2023.
Wilson, Steven Lloyd, and Charles Wiysonge. 2020. Social media and vaccine hesitancy. *BMJ Global Health* 5: e004206.
Wood, Allen. 1999. *Kant's ethical thought*. New York: Cambridge University Press.
———. 2017. Marx and Kant on capitalist exploitation. *Kantian Review* 22: 641–659.
———. 2018. Introduction. *Groundwork for the metaphysics of morals*. Trans. Allen Wood. New Haven: Yale University Press.

CHAPTER 7

The Duty to Promote Digital Minimalism in Group Agents

[Human beings'] propensity to enter into society … is combined with a thoroughgoing resistance that constantly threatens to break up this society. The predisposition for this obviously lies in human nature … Now it is this resistance that awakens all the powers of the human being … to obtain for himself a rank among his fellows, whom he cannot stand*, but also cannot* leave alone.
—Kant, Idea 8:20–21

Our algorithms exploit the human brain's attraction to divisiveness.
—Facebook, internal memo to senior executives (Quoted in Orlowski *2020*.)

7.1 Introduction

So far, we have canvassed the moral reasons we have to restructure our relationship with technology in virtue of the effects it has on us as individuals. But if we restrict our focus to the ways that technology can harm us as individuals, we overlook some morally significant effects for groups. In this chapter, we argue that addictive technology weakens our capacity to act autonomously *as a group*. We defend this claim by arguing that certain features of the attention economy (e.g., that it contributes to polarization[1]) threaten to undermine the legitimacy of political institutions.

[1] See Rathje et al. (2021).

T. Aylsworth, C. Castro, *Kantian Ethics and the Attention Economy*,
https://doi.org/10.1007/978-3-031-45638-1_7

We begin by explaining what is distinctive about group-level harms. In short, these are harms that cannot be fully explained without reference to a group agent. There are many cases where a harm to a group is both constituted by and reducible to harms to individuals. For instance, a group of pensioners might be harmed by embezzlement because the pension fund balance is reduced. This is not a genuine group-level harm, since there is no harm over and above harms to the individuals (e.g., each pensioner receives a smaller disbursement). Group-level harms, on the other hand, involve harms to structured groups. Insofar as a group collectively pursues goals and acts on the basis of shared intentions,[2] they are vulnerable to being harmed in ways that undermine the group's capacity to achieve those ends.

We draw on this account of group-level harms to demonstrate how the legitimacy of democratic institutions is threatened by the attention economy. We argue that legitimacy is partly a function of citizens' trust in institutions, as well as those institutions' competencies in the relevant domain, sensitivity to citizens' needs, and ability to signal competence and sensitivity to citizens' needs.[3] But the corrosive effects of fake news, polarization, echo chambers, and a motley of other features of the attention economy place a drag on all of these factors.[4] Thus, the attention economy not only harms us individually. It harms us collectively as well.

7.2 Group Autonomy and Group Harms

To better understand the kind of group-level harms at issue here, it is helpful to contrast them with related (but importantly distinct) harms. For instance, there are situations where a harm to an individual is thought to be a harm to the group of which that individual is a member. This is commonly pointed out in cases of racial injustice and genocide. When an individual is targeted for violence because of their racial identity, it is often said that this act harms the racial group *as a group*. The Nazis' anti-Semitism is an obvious example. Acts of violence against individual Jews on Kristallnacht harmed not only those in Germany and Austria, who were most immediately affected; it harmed European Jews as a group insofar as

[2] We discuss various accounts of group agents in Sect. 7.2. For the most part, our account of group agents is similar to List and Pettit (2011).

[3] Our argument in this section draws on Purves and Davis (2022).

[4] See Rubel et al. (2021) and Castro and Pham (2020).

it put them all in danger. This notion of group harms could be seen as the underpinning of Martin Luther King's famous claim that "Injustice anywhere is a threat to justice everywhere" or the workers' slogan: "An injury to one is an injury to all."[5] A similar concept is invoked in justifications of international criminal law when actions such as genocide are characterized as "crimes against humanity."[6] This issue also arises in bioethics, as practices like genetics research are frequently discussed in terms of risks to groups or communities.[7]

Although group affiliations play an important role in examining the moral implications of things like racial injustice and genocide, they do not harm a group *qua* group. They are what Dan Hausman (2007) calls "group-mediated harms." In situations of this kind, the harms are suffered by individuals, even though the individuals are targeted because of their membership in a group. Consider the case of a person whose application for a federally insured mortgage is rejected because he is Black. In a scenario like this, the harm is mediated by his group membership (it happens to him in virtue of his race), but it is the *individual* who was wronged, not the group. His interest in securing this particular loan is not one that Black Americans hold collectively. It is not a purely collective interest that was disrespected, but an individual one.

By contrast, there are many cases where a group has shared interests that it pursues collectively. Groups that pursue interests through shared

[5] The famous quote about injustice comes from Martin Luther King Jr.'s "Letter from Birmingham Jail." The "injury to all" slogan is used by the Industrial Workers of the World and has been attributed to David C. Coates. See Haywood (1929, 186).

[6] The idea here is that the harm of an action like genocide extends beyond national borders because it harms all of humanity. This idea is often deployed in the justification of humanitarian interventions and international criminal punishments. Larry May argues that an act qualifies as a crime against humanity when it targets people for their group affiliation (e.g., ethnicity, religion, etc.) rather than some property that is unique to the individual (87). May claims that such acts demonstrate a "callous disregard for the individuality of the person" and that this constitutes "an assault on what is common to all humans and hence to all of humanity" (84). See May (2000). In addition to May's philosophical justification, the concept of "crimes against humanity" has been codified into international law. The 1998 Rome Statute, which established the International Criminal Court (ICC), requires that all parties to the Statute recognize "that such grave crimes threaten the peace, security and well-being of the world." It outlines a list of such crimes in Article 7.

Others have criticized this justification. Andrew Altman and Christopher Wellman argue that "harm to humanity is a convenient but ultimately unpersuasive fiction." Altman and Wellman (2004, 42).

[7] See Davis (2000) and Weijer et al. (1999).

agency are susceptible to being harmed in a distinctive way. Corporations, nation states, and sports teams have interests of this kind. For instance, if their star player is hit by a bus, Manchester United will be less likely to win the European Cup.[8] Not only is the player harmed as an individual, the club is harmed as a group. Winning the cup is a shared interest that no player on the team would be capable of pursuing on his own. In cases of group-mediated harms, when we say that the group is harmed, this simply amounts to the claim that individual members of the group have become more susceptible to a certain kind of harm (e.g., individual Jews were subject to violence during the Holocaust, individual Black Americans were discriminated against, etc.). Group-level harms, by contrast, must involve groups that have collective interests, aims, or intentions.

Appiah (2011) draws a similar distinction when discussing group rights.[9] In his view, "collective rights" are ones that are exercised by groups, whereas "membership rights" are rights that individuals have in virtue of their membership in a group. He says, for example, that a democratic state has a collective right to self-determination. A right of this kind can be held only by a group, since it is not a right that any individual citizen is capable of exercising on her own (2011, 268).[10] By contrast, a person's right to vote in the American presidential election is a membership right that she has in virtue of her US citizenship. The right is mediated by her membership in the group, but it is not a right that is held by a group.

[8] Not every harm to a member of the team is necessarily a harm to the team itself. It depends on the player's contribution to the team's collective pursuits. Hausman writes, "An injury to a star player, a crucial executive, or an important leader is also an injury to the team, corporation, or tribe. This is not because any injury to any member of a structured group is automatically an injury to the group. If some of the least able players on a baseball team could be replaced with equally good players from the team's farm club, then an injury to them would not be an injury to the team." Hausman (2007, 357).

[9] In order to avoid the controversy surrounding group rights, we have chosen to discuss the issue here in terms of group harms. We can have obligations to groups even if groups do not have rights. Appiah defends the notion of group rights in the article, but he is responding to those who are critical of group rights. The standard objection to group rights is that group agents are not the kinds of entities that can bear rights and that all putative cases of group rights ultimately dissolve into individual rights. See, for example, Narveson (1991). Appiah is responding most directly to James Sterba's skepticism: "Moral entitlements are not held by groups … Rights are possessed by persons. As when persons are entitled to be made whole for some injury earlier done to them, the duty owed is … to them as individuals" (Sterba 2009, 57–58).

[10] See also Appiah (2005a, b).

Given the controversial status of group rights in the philosophical literature, we tend to frame our argument here in terms of group harms. The argument of this chapter does not depend on the existence of group rights. In our view, we have obligations to respect the *interests* of certain groups even if groups do not have *rights*. To push the claim further, we believe such obligations would not be undermined even by the claim that groups have no moral status whatsoever—a rejection that is stronger than denying group rights. It is not uncommon for philosophers to defend thoroughgoing individualism about moral status; they argue that our moral concerns should be limited exclusively to individuals.[11] Even on this view, it is still possible to make sense of the claim that we have prima facie obligations to respect the interests of certain groups. The interests of groups would have moral significance whenever the group's ability to pursue its collective interests is something that matters to individuals.[12] Even if Manchester United has no intrinsic moral status as a group, undermining the club's capacity to win the European cup would run afoul of the interests of many individuals (the players, the coaches, the fans, etc.). Of course, not all instances of causing harm should count as moral wrongs. If Liverpool FC were to sign a contract with Manchester's star player, this would harm Manchester United (or, for the individualist, we could say that it harms those to whom the club matters), but it would not necessarily be a moral wrong. In order for a harm of this kind to constitute a moral wrong, we must have a sufficient reason to respect the interests that are at stake.

When the Acme Corporation creates a product that is superior to and cheaper than Biffco's equivalent product, there is no question that this harms Biffco. But, other things being equal, this does not seem to be a moral wrong. Even though Acme's action goes against the collective interests of Biffco (and against the interests of individual shareholders, employees, etc.), there is no reason to think that Acme has an obligation to respect

[11] There are many examples of this kind of individualism. See the discussion of Sterba and Narveson above. Similarly, Thomas Pogge claims that "the ultimate units of concern are human beings, or persons—rather than, say, family lines, tribes, ethnic, cultural, or religious communities, nations, or states" (Pogge 1992, 48).

[12] As noted in earlier chapters, according to Kant, beneficence is an imperfect duty to promote the ends of others as if they were our own. But there is an important qualification; we are obligated to promote only their morally permissible ends. When we extend this idea of beneficence to groups, we can mitigate the worry that we are under an obligation to promote the ends or respect the interests of groups that are pursuing morally impermissible goals.

those interests. The moral evaluation seems quite different, however, when we consider a case of one state interfering with another state's capacity for self-government. If Arendelle were to undermine a democratic election in Blefuscu, this would rightly be seen as a moral wrong. Arendelle and its citizens have a moral obligation to respect Blefuscu's capacity for self-determination.[13] We do not necessarily have obligations to respect the interests of sports teams and corporations (though we might in certain cases),[14] but we almost always have obligations to respect a democratic state's interest in self-determination.[15]

We will develop our account of democratic legitimacy more fully in Sect. 7.3, but before moving on, we should provide one final clarification about the nature of group-level harms. Thus far, we have specified only a single condition of what makes a group susceptible to group harms: a

[13] Although this claim might be intuitively obvious to most readers, providing a full justification of it would be a difficult philosophical task and one that lies outside the scope of this book. Perhaps the simplest route would be to point to certain historical realities. It is widely accepted that states have this right. Formal recognitions of a nation's right to sovereignty are at least as old as the Treaty of Westphalia (1648). The right to self-determination was enshrined in Article 1 of the United Nations Charter, which states that one of the purposes of the UN is "To develop friendly relations among nations based on respect for the principle of equal rights and self-determination of peoples, and to take other appropriate measures to strengthen universal peace." Self-determination was also acknowledged as a "right" in the Atlantic Charter signed by the US and the UK.

Skeptics might argue, however, that an existing social contract is not a sufficient foundation for a moral right. Another approach would be to ground a state's right to self-determination in the rights of individuals. For instance, one could argue for a Lockean view according to which the democratic legitimacy of the state is necessary to protect the natural rights of its citizens. For a more thorough defense of the right to self-determination as a concept of international law, see Margalit and Raz (1990).

[14] For instance, the CEO of a corporation may have an obligation to promote the company's interests in virtue of commitments she has made. The manager of the football club has a similar obligation.

[15] The justifiability of humanitarian interventions is controversial. See Walzer (1977), ch 6. It can be argued, in extreme cases, that a state loses its claim to sovereignty when it fails to provide for the basic security of its citizens. In such a case, the foreign power who intervenes might argue that they are not interfering with the affairs of a sovereign state. Such interventions ought to be rare, however. The history of colonialism is one important reason why international law promotes reluctance to intervene. Tasioulas and Verdirame write: "By embracing this notion [of self-determination] international law gave recognition to the moral and political value of self-government, accepting that people prefer to be ruled by their own bad rulers rather than foreigners, including those foreigners with some claim to greater competence" (2022).

collectively held interest. More needs to be said, however, about what is distinctive about collective interests. Someone might object that the putative cases of group-mediated harms, which we discussed above, actually *do* involve shared interests. It could be argued that Black Americans shared an interest in putting an end to housing discrimination; European Jews shared an interest in the downfall of the Nazi Party, etc. What distinguishes such groups from "structured"[16] groups, however, is that racial and ethnic groups as such lack the structure to pursue their shared interests collectively. They do not act *as a group*. In addition to having shared interests, corporations, states, and sports teams are organized in such a way that they can pursue their interests collectively. In short, they are group agents.

There are several competing accounts of group agency, and our argument in this essay is not necessarily committed to a particular view.[17] At a minimum, agents must have the capacity to represent states of affairs in their environment, motivational states, and the ability to intervene in the environment in accordance with their motivations (List and Pettit 2011, 20). In order for a group to qualify as an agent, we must be able to attribute some form of representational states, motivations, and capacities to the group. Some critics worry that this requires us to posit something ontologically extravagant like a "group mind," but most accounts in the literature are, in some sense, reductionist.[18] We can attribute representational states and motivations to groups insofar as these are defined functionally; groups are capable of jointly engaging in rational deliberation, for instance. There is nothing mysterious about this kind of group deliberation; it happens all the time in department meetings and boardrooms. The department's conclusion that it would be good to hire a new faculty member is a representational state that can be attributed to the group. The

[16] Hausman (2007) uses the term "structured" groups to refer to groups that are susceptible to harm. For reasons that we explain below, we prefer to frame the discussion in terms of "group agents."

[17] See List and Pettit (2011), Ludwig (2016), and Tuomela (2013).

[18] John Searle says that talk of group minds is "at best mysterious and at worst incoherent. Most empirically minded philosophers think that such phenomena must reduce to individual intentionality" (1990, 404). All three of the accounts mentioned above are ontologically reductionist. They do not posit an independently existing "group mind." Group agents are nothing more than sets of individuals acting in a particular way. List and Pettit argue that they can reject eliminativism about group agents without rejecting methodological individualism—the claim that all explanations of the social world ultimately boil down to facts about individuals (2011, 3–6).

same can be done with intentions. It seems perfectly sensible to say that the orchestra intentionally played Beethoven's 9th Symphony. But the intention to play the symphony cannot be attributed to any particular individual.[19] A flutist in the orchestra is not capable of playing Beethoven's 9th Symphony by herself, though she is certainly capable of doing her part.[20] Talking about groups in this way (i.e., in terms of shared beliefs, desires, and intentions) makes it possible to attribute states to them that are characteristic of rational agency. And to the extent that groups are indeed capable of rational agency, they can be said to have the capacity for autonomy. Understood in this way, one way to harm a group *qua* group would be to violate or undermine its autonomy. The racial and ethnic groups mentioned above, which are susceptible only to group-mediated harms, are not group agents (though advocacy organizations such as the NAACP and the Shoah Foundation are).

Insofar as groups are capable of acting as group agents, we can think of them as either possessing or lacking autonomy. Groups can have or lack the capacity to set and pursue their own ends. The rational agency of groups can be exercised well or it can function poorly. Structured groups, such as sports teams and corporations, typically have an interest in functioning well. Undermining the capacities of a group agent can therefore be understood as a harm to the group itself. You could harm a group agent by undermining its autonomy (its capacity to set and pursue its own ends) in ways that are perfectly analogous to ways that you could harm an individual agent. You could undermine the group agent's capacities or its

[19] A claim of this kind is defended by Velleman (1997); he argues against those like Bratman and Searle who hold views of group agency that do not involve genuinely shared intentions. In their view, group agency is nothing more than individuals who intend to coordinate their actions and intentions with other individuals. Velleman's view is closer to that of Margaret Gilbert (1992), who talks about a "plural subject" brought about by a "pool of wills." Velleman's point, which we concur with, is that it would not make sense to talk about an individual intending to perform the symphony since it is not clear that an agent can intend to do something that she is not capable of doing herself. For Bratman, an individual has a "we-intention" whose content may be something like this: I intend that *we* play the symphony.

[20] In fully reductionist views, the collective intention is to be understood simply in terms of a set of individual intentions (see, e.g., Bratman 1992, 1993). Each individual merely intends to do her part and to "mesh" her sub-plans with other members of the group.

authenticity.[21] In some cases, harming a group in this way constitutes a moral wrong. As we suggested earlier, one of the clearest examples of this is democratic legitimacy.

There are at least two ways of understanding the nature of this moral wrong. Some might believe that the group has what Appiah calls a "collective right" to self-determination, and such an account would make the role of the group agent ineliminable. Others might reject the notion of group rights and argue that only individuals have been wronged. When a state's democratic legitimacy is undermined, this wrongs the individual citizens who have an interest in their state's legitimacy or in its ability to function properly. We prefer the view that some group agents have moral status and that some moral wrongs should be understood, at least partially, in terms of harms to groups. But we believe that the main conclusion of this section (i.e., that there is a moral obligation to respect the autonomy of certain groups) does not depend on this view. The argument is open to the reductionist view as well.

Even according to the individualistic account, the moral wrong cannot be fully understood without reference to the group agent. The only way to make sense of the claim that individuals were wronged is to point out that they had an interest in the group agent's autonomy. So it does not ultimately matter whether the group agent's autonomy has intrinsic or instrumental moral value; either way, there are cases where it is morally wrong to undermine a group agent's autonomy. The important conclusion here is that some groups have shared interests, which they pursue

[21] For example, Blefuscu's army is a group agent. And as we saw in Chap. 2, autonomous capacities require the ability to form coherent plans, revise those plans in light of new information, etc. If a spy from Arendelle were to disrupt communication between the units of Blefuscu's army, that would undermine the group's baseline capacities. It might be helpful to use Bratman's (1993) model of group agency to clarify this example. In Bratman's view, the members of a group agent can jointly intend an action only if the agents intend to mesh their sub-plans. Rosa and Ray can be said to share the intention to paint the house only if they intend to mesh their sub-plans (such as the sub-plans involving color choice). If Rosa is painting the house green, and Ray is painting it red, then we must reject the claim that they are painting the house together. Shared agency requires means-end coherence. If the spy successfully disrupts communication between units of Blefuscu's army, then she would undermine that group agent's autonomy by making it impossible for them to mesh sub-plans. Similarly, the spy could interfere with the Blefuscu army's authenticity. She might disseminate false information that leads the group to attack the wrong targets. This kind of manipulation undermines the group's autonomy in much the same way that Iago undermined Othello's autonomy by means of deceit.

collectively, and there are situations where it would be morally wrong to undermine the group's ability to pursue its ends. In the next section, we demonstrate what this would look like by considering the group agency of the state.

Kant was particularly concerned with the proper functioning of the state. As we saw in the previous chapter, Kant believed that it would be impossible for groups of human beings to enjoy freedom without creating a state to safeguard their rights. He also worries that the threat of violence would always loom in the absence of a state. In a footnote from *Toward Perpetual Peace*, he writes:

> Within each state it [human malevolence] is veiled by the coercion of civil laws, for the citizens' inclination to violence against one another is powerfully counteracted by a greater force, namely that of the government, and so not only does this give the whole a moral veneer (*causae non causae*) but also, by its checking the outbreak of unlawful inclinations, the development of the moral predisposition to immediate respect for right is actually greatly facilitated. (8:375–76)

As we will see in what follows, the attention economy poses a variety of threats to this stability. Kant's moral philosophy gives us compelling moral reasons to be concerned about the proper functioning of the state and the capacity of groups to set and pursue their own ends.

7.3 Trust, Trustworthiness, and Democratic Legitimacy

We'll now show how trust and trustworthiness play an important role in preserving a democratic state's legitimacy. Many of the state's aims can be accomplished only if those who are subject to its rule see it as trustworthy. This means that undermining public trust (and/or the state's trustworthiness) hinders the state's ability to fulfill its responsibilities. For instance, criminal justice systems depend crucially on citizens' willingness to serve on juries, testify as witnesses, cooperate with the police, etc. And citizens are far less willing to cooperate when they do not trust the institutions or

the individuals charged with carrying out these tasks.[22] We will argue that states are vulnerable to group-level harms insofar as their group agency is susceptible to being undermined by threats to public trust.

As we suggested above, respecting a democratic state's right to self-determination is one of the most compelling examples of an obligation to respect group autonomy.[23] In the absence of a compelling reason to intervene, it is typically seen as morally wrong to interfere with a state's ability to govern itself. This right to self-determination can be seen as having both an external dimension, which involves being free from external control (e.g., colonialism, puppet governments, etc.), and an internal dimension which concerns "the rights of people to pursue democratic governance domestically" (Tasioulas and Verdirame 2022).[24] When dealing with democratic states, whose rightful authority derives from the consent of the governed, this issue is often framed in terms of "legitimacy."

Philosophers tend to discuss legitimacy in normative terms; they ask questions about the justificatory grounds of a state's claim to exercise authority and use coercive power. But it is important not to lose sight of descriptive legitimacy, which concerns citizens' attitudes and beliefs about the institutions that govern them.[25] Simply put, a state is legitimate, in the

[22] See Purves and Davis (2022) for an interesting discussion of how algorithmic opacity threatens the legitimacy of criminal justice institutions by eroding the public's trust. For a compelling example of how declining public trust undermines institutions of criminal justice, see Fenton (2021). He discusses how years of corruption and violence had made it nearly impossible for the Baltimore police to get citizens to cooperate: "While the police department leadership begged citizens to cooperate, some of its elite officers were running roughshod on Black men in poor neighborhoods, creating a free-fire zone for anyone seeking to exploit them" (2021, 269).

[23] Many of the most complicated debates in the literature on self-determination involve questions of secession and the formation of new states. It is notoriously difficult to pin down exactly when a nation, people, or ethnic group has a legitimate claim to secede and form an independent state. See Crawford (2006) for an extensive discussion. Crawford begins by acknowledging the widely held view that the formation of new states is "a matter of fact, not of law" (2006, 4). For this reason, we tend to speak in terms of a *state's* right rather than a nation's right or a people's right.

[24] Tasioulas and Verdirame cite Franck (1992) for a defense of the importance of the internal dimension. Franck writes, "Since self-determination is the oldest aspect of the democratic entitlement, its pedigree is the best established. Self-determination postulates the right of a people organized in an established territory to determine its collective political destiny in a democratic fashion and is therefore at the core of the democratic entitlement" (1992, 52).

[25] Our argument here (as well as this way of framing legitimacy) is heavily indebted to the discussion in Purves and Davis (2022).

descriptive sense, just in case the citizens *believe* that the state is legitimate.[26] This kind of legitimacy is important because a state would be unable to function properly if citizens see the state as illegitimate and thus do not comply with its directives. This means that maintaining descriptive legitimacy requires some amount of public trust. Citizens will see their government as legitimate only if they trust state institutions.

7.3.1 Trust and Trustworthiness

Given the abundant philosophical literature on the nature of trust, it may seem difficult to determine what exactly is distinctive about trust and trustworthiness. But many of the accounts share certain basic features. Following Annette Baier (1986), it has become common to begin the discussion by recognizing that trust involves accepting one's vulnerability to another person.[27] If you trust your friend to watch your laptop while you go to the bathroom, you must accept the fact that this makes you vulnerable to certain risks. She could steal your laptop or deliberately pour coffee on it. If you trust her, you are *relying* on her to refrain from betraying you in those ways. What is more, if she were to betray you, it would be appropriate for you to feel disappointment or resentment; you could rightfully demand an apology. To some, the aptness of these reactive attitudes is one of the things that makes trust distinct from "mere reliance."[28]

[26] As Tom Tyler puts it, legitimacy is "the belief that authorities, institutions, and social arrangements are appropriate, proper, and just" (Tyler 2006, 376). Discussions of descriptive legitimacy are usually traced back to Weber. See Peter (2017) for an overview.

[27] See Baier (1986, 235).

[28] Pinning down exactly what distinguishes trust from reliance is the source of most of the controversy in the literature. Baier (1986) argues that trust involves relying on the trustee's goodwill toward the trustor. Jones (1996) also provides an account in which the trustee must be motivated by goodwill. Hardin (2002), on the other hand, offers a self-interest account. When you trust your tax preparer to do your taxes, you might not believe that he acts out of goodwill for you; you may think only that your interests converge insofar as it is profitable for him to maintain a relationship with you. Hawley (2014) criticizes motives-based accounts of trust, like Hardin's and Jones's, on the grounds that they cannot easily make sense of distrust. As we discuss below, Hawley defines trust as relying on someone to fulfill a commitment.

If you are relying on your car to get me to work and it breaks down, it would be inappropriate to feel betrayed by the car.[29]

Karen Jones (2012) expands the discussion of trust by analyzing the related concept of "trustworthiness," the property someone could have that would make it fitting for us to trust her. Trustworthiness, like trust, is usually restricted to a particular domain.[30] For instance, you trust your tax preparer to do your taxes, but you would not trust her to grade upper-division philosophy papers. As Katherine Hawley puts it: "Trust is a three-place relation, involving two people and a task" (2014, 1). The domain sensitivity of trustworthiness is what leads Jones to conclude that competence plays a key role in determining whether or not someone is trustworthy. Our students trust us to grade their philosophy papers only if they believe that we have competence in this domain (Purves and Davis 2022). According to Jones, if the trustee is to be regarded as "richly trustworthy" then she must also signal her competence to those who trust her.

Trust is typically understood as an interpersonal phenomenon, however, and there are some who are skeptical about applying concepts of trust and trustworthiness to groups or institutions.[31] Some of the

[29] See Hawley (2014). Most accounts of trust use the fittingness of "betrayal" to distinguish reliance from trust from mere reliance. Nguyen (2022) is one of the few exceptions to this consensus. He argues that we might feel betrayed by an object (like a smartphone that fails to give us a calendar reminder that we were relying on), but this kind of "betrayal" is not exactly identical to the kind of betrayal we experience when we are let down by an agent. Even if we feel betrayed by the artifact, we do not ascribe moral notions of blame toward it.

[30] See Jones (2012).

[31] See Budnik (2018). Budnik argues specifically that democratic governance does not require trust. Budnik takes it as given that trust is a strictly interpersonal phenomenon, so he is not talking about trust in institutions at all. Budnik focuses instead on placing trust in individual governmental officials. This means that the target of his opposition is not exactly the same as our claim in this essay. We are not suggesting that trust in institutions necessarily requires trusting particular governmental officials. Rather, we believe it requires trust in the group agents that are constituted by the institutions themselves.

Hawley (2017) is more open to the idea of describing groups as trustworthy, but she ultimately concludes that the standard distinction between trustworthiness and reliability does not apply to groups and that this makes it more appropriate to talk about groups as "reliable." She argues that the reactive attitudes that are appropriate in breakdowns of trust (anger, resentment, betrayal, etc.) apply to individual members of groups rather than to groups themselves.

Outside of philosophy, it is fairly common for those in the social sciences to talk about groups in terms of trust and trustworthiness. Political science titles like *The Oxford Handbook of Social and Political Trust* are commonplace (Uslaner 2018).

reluctance to attribute trustworthiness to groups stems from the fact that accounts of trustworthiness often involve attributing certain mental states to the trustee (goodwill, self-interest, adopting certain reasons, etc.) and this may seem implausible when applied to groups or institutions.[32] In *Cooperation Without Trust?* Karen Cook, Russell Hardin, and Margaret Levi argue that motivations and intentions cannot be ascribed to institutions or groups, and they claim that we cannot trust the individual members of these groups because we are not sufficiently familiar with their motives.[33] For these reasons, it might be preferable to adopt Hawley's view, since she defines trust as relying on someone to fulfill a commitment (Hawley 2014).[34] Even if we cannot attribute mental states to groups, it certainly seems reasonable to claim that groups and institutions make commitments.[35] For instance, the Preamble to the US Constitution expresses the government's commitment to "provide for the common defence" and "promote the general Welfare." To say that the citizens of the US "trust their government" simply amounts to the claim that they rely on the government to fulfill its commitments. To claim that the US government is trustworthy, we must believe that it would be *appropriate* for us to rely on it to fulfill its commitments.

But this is a fairly thin conception of trustworthiness. What are the conditions under which it would be appropriate for citizens to rely on their government to fulfill its commitments? Here, it would be useful to return to the criteria suggested by Jones (2012), even if they must be amended in order to apply to groups. Jones explained why it is important that trustees signal their ability and their willingness to act in the ways that we are counting on them to act. If an institution fails to signal its

[32] For instance, Jones's account requires that the trustee treat the fact that the trustor is counting on her as a reason to act accordingly. Some may be skeptical about the capacity of groups to treat such facts as reasons. On our view of group agency, which draws on List and Pettit's view, this is not so problematic. Groups are certainly capable of something that is functionally akin to rational deliberation. So it does not seem terribly problematic to talk about groups taking facts as reasons.

[33] See Cook et al. (2005). Cf. Purves and Davis (2022).

[34] Kirby et al. (2018) discuss trustworthiness as it applies to corporations. They adopt Hawley's account on the grounds that it can plausibly be applied to corporations and that it makes sense of the appropriateness of trust in terms of commitments. They then cite Gilbert's many defenses of the claim that groups make commitments. See Gilbert (1992, 2006, 2013). See Hawley (2017) to understand how she applies trustworthiness to groups and organizations.

[35] See Kirby et al. (2018). Cf. Gilbert (1992, 2006, 2013).

reliability, its trustworthiness would be undermined. For instance, imagine a country that kept its military a secret. The citizens know nothing about the armed forces or its arsenal; they do not even know that their country has a military. How could they possibly trust their government to protect them? It would not be appropriate for the citizens to rely on their government to provide for their defense. If they were threatened by foreign invasion, the citizens might stockpile their own weapons, train their own militias, or flee the country entirely. They would have no reason to trust their government because the government failed to signal that it was able and willing to defend them from invasion.

Jones (2012) also includes a criterion about the trustee being responsive to the right sort of reasons. In her view, the trustee must be responsive to the fact that others are counting on her and she must take this fact as a compelling reason to act as counted on (Jones 2012, 71). Someone who is typically reliable in fulfilling their commitments might be seen as untrustworthy if they are not responsive to the fact that someone is counting on them. Perhaps you want to know if you should trust your colleague to finish a project that you have committed to work on together. You know that she is competent to finish the work, and she has adequately signaled her competence to you. But you also know that her work is motivated entirely by her desire to get a promotion. Say that she goes up for promotion on Monday and you are counting on her to finish the project on Wednesday. The fact that you are counting on her to finish the work means nothing to her, however. She cares only about the promotion. In this case, you would conclude that she is not trustworthy; you should not rely on her to finish the project. If she gets the promotion on Monday, she will no longer be motivated to fulfill the commitment. This is precisely why Jones (2012) argues that trustworthiness requires the trustee to take the fact that you are counting on her as a compelling reason to act. The robustness of the trustee's motivation matters.

Once again, this complicates the account of trustworthiness when applied to groups. While individuals can respond to facts as reasons, it is less clear that groups have this capacity (although we are less troubled by this worry, since we believe groups can be seen as agents). Purves and Davis (2022) express this concern as follows: "It is at best unclear whether institutions possess such a capacity and hence whether they can take the fact of someone else's dependency as a direct and compelling reason to act" (2022, 7). Nevertheless, Purves and Davis want to apply the concept of trustworthiness to institutions, so they reframe Jones's responsiveness

criterion in such a way that it can be applied to groups. Instead of "taking a fact as a reason," they suggest that trustworthy institutions must be "non-accidentally responsive" to the fact someone is counting on them. Institutions may have all sorts of mechanisms that make them responsive in this way (democratic elections could play this role, for instance). This mitigates the worry about ascribing mental states to groups while preserving Jones's core idea that trustworthiness requires a certain kind of responsiveness to the dependence of trustors.

Like Kirby et al. (2018), we think that Hawley's "commitment" account may be the most appropriate view when talking about placing our trust in groups. This means that institutional trust should be thought of as relying on a group to fulfill its commitment. We would also like to supplement Hawley's view of trust with Jones's account of trustworthiness.[36] As Purves and Davis show, Jones's criteria can be fruitfully extended to institutions, mutatis mutandis. This yields the following definitions of trust and trustworthiness when applied to groups[37]:

Trust in groups: An individual *I* trusts a group *G* in a domain *D* if and only if *G* has made a commitment in *D* and *I* is relying on *G* to fulfill its commitment.

Trustworthiness of groups: A group *G* is trustworthy to an individual *I* in a domain *D* if and only if (1) *G* is competent to fulfill its commitments in *D*, (2) *G* is non-accidentally responsive to the fact that *I* is counting on G to fulfill its commitments, and (3) *G* provides adequate reason for *I* to believe that *G* is competent to fulfill its commitments and that *G* is non-accidentally responsive to the fact that *I* is counting on *G*.

[36] Hawley (2014) provides a rather thin account of trustworthiness. Hawley suggests that individuals can be seen as trustworthy in virtue of the simple fact that they reliably fulfill their commitments: "On the commitment account, trustworthiness requires us to ensure that our commitments do not outstrip our actions. This requires judiciousness in acquiring commitments as well as doggedness in fulfilling commitments already acquired, independent of others' expectations. Trustworthy people must sometimes disappoint up-front by refusing new commitments, rather than violate trust later on: this is the moral 'power of no'" (2014, 15).

[37] Our definition of trustworthiness is nearly identical to that of Purves and Davis (2022) except that we have added Hawley's language about commitments. It is helpful to think about institutional trustworthiness in terms of commitments because this makes the relevant domains explicit.

A trustworthy group must be competent in the relevant domain. And it must signal its competence and responsiveness to us in such a way that we are justified in believing that it will fulfill its commitments.

7.3.2 Democratic Legitimacy

Now that we have these definitions of trust and trustworthiness in hand, we are in a position to explain how they are connected to legitimacy. The state's legitimacy depends crucially on its ability to fulfill its commitments. This could be demonstrated easily enough through the lens of something as narrow as the Hobbesian view according to which the sovereign has a legitimate claim to authority as long as it provides for the basic security of its citizens. Imagine a king who refuses to enforce conscription and relies entirely on a volunteer army. He remains committed to securing the defense of his subjects, however. He offers a variety of incentives in order to attract recruits to the army: high pay, generous benefits, and so on. This works at first. But then rumors begin to circulate that the treasury has fallen on hard times and the soldiers won't receive full pay. This narrative becomes so widespread that the soldiers no longer trust the king; they do not rely on him to fulfill his commitment to pay them. They stop showing up for service, and the erosion of public trust means that there are no new recruits to take their place. The lack of trust also affects the king's trustworthiness. Recall that the king is trustworthy only if he is competent to fulfill his commitment. Now that he has no army, he is unable to guarantee the basic security of his subjects. The erosion of trust undermined his trustworthiness because it made him unable to fulfill his commitments. This undermines his legitimacy as sovereign. The social contract that granted him authority was contingent on his ability to provide for the security of his subjects.

Purves and Davis (2022) highlight the importance of this feedback loop between trust and trustworthiness in the context of criminal justice. The proper functioning of the criminal justice system depends on "compliance and help-seeking behavior" (19).[38] When citizens no longer trust the criminal justice system, it becomes less able to do its job (i.e., to meet its commitments). When it becomes less able to fulfill its commitments, it becomes less trustworthy and citizens have even less reason to rely on it:

[38] They cite the work of Tyler and Jackson (2014) to provide an account of descriptive legitimacy. Cf. Tyler (2006) and Tyler and Huo (2002).

> Notice that this can (and surely does) generate the following feedback loop: decreasing legitimacy within a given population causes lower levels of compliance and help-seeking behavior by those in that population. The inability to secure voluntary compliance means it is less capable of achieving desired outcomes and delivering on its distinctive mandate. This failure gives rise to a further decrease in legitimacy, as individuals find themselves less trusting of a less capable institution. And the cycle starts anew. Declining descriptive legitimacy therefore negatively affects the institution's ability to achieve the aims that define the institution, which in turn erodes trust and confidence. (Purves and Davis, 2022, 19–20)

In order for the government to be able to meet its commitments, it must be trustworthy. Preserving the government's competence and capacities also requires that citizens have some trust in the government.[39] This means that a trustworthy government must effectively signal both its competence and its responsiveness to the citizens.

7.3.3 How the Attention Economy Undermines Trust and Trustworthiness

The example given above evaluated legitimacy in Hobbesian terms, but the conditions of democratic legitimacy are more expansive. A legitimate

[39] There is a danger here of making it sound like less trust is always bad for democracy and more trust is always good. But this is obviously not the case. While a lack of trust undermines the government's capacities, an excess of trust leads to other problems. Hetherington (2008) points out, for instance, how support for military action positively correlates with political trust. Having too much trust in the government could result in citizens supporting unjust wars or using questionable tactics to win those wars. He writes: "A compelling case for the pernicious effects of high trust could be made today. Specifically, many believe that the Bush administration will not be remembered kindly by history because of its willingness to pursue extra-Constitutional means to battle terrorism, such as the use of wiretaps without first obtaining warrants; its reluctance to ban torture; and its desire to jail suspected nonmilitary enemy combatants without habeas corpus rights. Republicans' high levels of trust may be at the heart of their support of these initiatives and, consequently, the Bush administration's willingness to pursue them" (2008, 23).

Hetherington and Husser (2012) and other political scientists have shown that declining political trust has a variety of effects on the abilities of government to accomplish goals; it has a particularly negative effect on the government's ability to pursue liberal domestic policies. Klein and Robison put it even more broadly in terms of democratic health: "Trust in government is a crucial indicator of democratic health as trust enables governments to tackle difficult policy problems" (2020, 47).

democratic state must do more than simply provide basic security for its citizens. This makes democratic states more susceptible to being undermined by the erosion of public trust. For instance, citizens might not see their elected officials as legitimate if they believe that their elections are not free and fair. In this section, we show how certain noxious elements of the attention economy contribute to the erosion of citizens' trust and of the government's trustworthiness. One way this might happen is that the spread of fake news could disrupt the government's ability to signal its competence in some domain (e.g., misinformation about vaccines undermines public trust in health agencies and regulatory bodies). But this is certainly not the only threat. A growing body of evidence shows how social media contributes to polarization and radicalization, and this can lead to legislative gridlock by making it harder to achieve compromises.[40] We show how some of these negative effects are unintentional byproducts of how the algorithms are designed, but we also explain how and why social media has been weaponized by bad actors who are making a deliberate effort to undermine the efficacy of democratic states.

Because of their dependence on ad revenue, tech companies are incentivized to maximize engagement. This has led, naturally enough, to the development of algorithms that aim to direct users to content they might enjoy. This business model, when combined with certain features of human psychology, is what makes the attention economy such a powerful social force. Unfortunately, it is also implicated in some of the more pernicious aspects of our relationship with technology.

For instance, whether online or offline, people generally like to associate with others who are similar in some way. A great deal of research has shown how this tendency toward "homophily" in social media networks

[40] Hetherington and Rudolph (2017) argue that the breakdown of trust is an especially important part of the problem of polarization. They "explain how the polarization of political trust has contributed to ongoing political dysfunction in Washington" (2017, 579).

contributes to the existence of echo chambers.[41] When like-minded individuals cluster in online communities, especially ones with positive and negative reinforcement mechanisms (likes, upvotes, shares etc.), certain views are promoted, while other perspectives are either ignored or actively discredited. While much social media advertises itself as a platform for connecting, unfriending in response to political posts plays a large role in pruning our online networks.[42]

Here it is important to address some skepticism around what we have called echo chambers. Some have argued that claims about echo chambers are false or overblown, citing, for example, the evidence that political polarization is often highest among older populations, a group that uses social media the least (Boxell et al. 2017). In response to this specific argument, we follow Van Bavel et al. (2021) in thinking that the balance of evidence suggests otherwise. Boxell et al. (2017) was based on observational data; however, other studies—such as Allcott et al. (2020) and Asimovic et al. (2021)—have shown that there does, indeed, seem to be a causal effect on polarization borne of social media use. For instance, Allcott et al. (2020) had some users but not others deactivate their account leading up to the 2018 US election. Those who deactivated were, at the end of four weeks, less polarized than participants in the study who did not deactivate their accounts. Similarly, Asimovic et al. (2021) had some users but not others from Bosnia and Herzegovina deactivate their Facebook accounts during Genocide Remembrance week, and those who did deactivate had lower feelings of ethnic outgroup animosity than those who did not (cf. Van Bavel 2021).

Further, we think that there is a complication in research about echo chambers. When the focus is on which news stories we see, there seems to

[41] See, for example, Cinelli et al. (2021). They write: "Social media may limit the exposure to diverse perspectives and favor the formation of groups of like-minded users framing and reinforcing a shared narrative, that is, echo chambers… Our results show that the aggregation of users in homophilic clusters dominate online interactions on Facebook and Twitter" (1). Similarly, Finkel et al. (2020): "Social-media technology employs popularity-based algorithms that tailor content to maximize user engagement, increasing sectarianism within homogeneous networks (SM), in part because of the contagious power of content that elicits sectarian fear or indignation" (Finkel et al., 2020, 534).

Talisse (2019) argues that much of the recent polarization is the result of widespread ideological segregation, and this extends well beyond social media. People increasingly tend to associate only with those who share their political views. And he argues that politics has permeated spheres of social life where it used to be absent.

[42] See Sasahara et al. (2021) and Goyanes et al. (2021).

be little effect (cf. Haidt and Bail n.d.). However, when we focus on networks, we get a different answer. It is undeniable that there is partisan social sorting on social media sites.[43] This is facilitated by, among other things, the ability to "unfriend" as discussed above. Zynep Tufekci notes that the fact that our networks are sorted can undo the effect that we might think seeing different news stories should have, as they tell us *how* to read the news stories:

> [W]hen we encounter opposing views in the age and context of social media, it's not like reading them in a newspaper while sitting alone. It's like hearing them from the opposing team while sitting with our fellow fans in a football stadium. Online, we're connected with our communities, and we seek approval from our like-minded peers. We bond with our team by yelling at the fans of the other one…. Belonging is stronger than facts. (Tufekci 2018)

There is empirical support for what is being said here. As Bail et al. (2018) have shown, being exposed via social media to information from the other side can *increase* belief polarization and Tufekci gives a plausible explanation as to why.

This is not to say that the selective exposure to news stories that social media and other digital tools enable is not a concern when it comes to polarization. Indeed, social media users are disproportionately exposed to like-minded political information because they tend to have relationships with like-minded users.[44] Confirmation bias also contributes to the rise of these informational environments as well. People usually prefer to see content that confirms their views, and news feed algorithms can pick up on this fingerprint to reinforce their views by feeding them congenial information.[45] To make matters worse, once people come to inhabit an echo chamber, it is also fairly natural for them to adopt more extreme positions over time.[46] This makes them more likely to believe content that is false, inflammatory, or conspiratorial.[47] All of these phenomena (echo

[43] See, e.g., Cinelli et al. (2021), Barberá (2015), Hong and Kim (2016), Mosleh et al. (2021), and Halberstam and Knight (2016).

[44] See Halberstam and Knight (2016).

[45] See Cho et al. (2020).

[46] See Sunstein (1999).

[47] See, e.g., Brady et al. (2021), Rathje et al. (2021), and Vosoughi et al. (2018). They respectively show that posts that use emotional language increase their diffusion, that posts about political outgroups are twice as likely to be shared than those about ingroups, and that falsehoods diffuse "significantly farther, faster, deeper, and more broadly than the truth in all categories of information" (Vosoughi et al. 2018, 1146).

chambers, polarization, and fake news) have a deleterious effect on our capacity for democratic governance.

We explain each of these issues (and their interconnectedness) in greater detail below. We will also consider the ways that those who have an interest in undermining democratic governments have seized upon these vulnerabilities, effectively weaponizing the tools of the attention economy. Western democracies have become acutely aware of the ways that certain foreign powers (Russia, most notably) are using social media to foment polarization and spread misinformation.[48] This provides additional evidence for the claim that social media poses a threat to democracy.

To begin, it would be helpful to provide a fuller account of what polarization is and how the attention economy contributes to it.[49] This is a natural place to start, given the many ways that polarization is connected to the mechanisms mentioned above (echo chambers, fake news, inflammatory content, etc.). The first thing to note is that polarization is not a singular phenomenon; there are several different senses in which American society is becoming increasingly polarized.[50] Talisse (2019) distinguishes between "political polarization," which concerns things like the distance of political parties, from "belief polarization," which refers to the way that an individual's belief becomes more extreme over time (106).

Within these two broad categories, there are even more fine-grained distinctions. One way for a belief to become more "extreme" is to increase the agent's confidence or credence in the belief. For instance, someone might go from *thinking* that Biden's spending is the cause of inflation to *feeling certain* that Biden's spending is the cause of inflation. Or someone might shift from being *somewhat opposed* to the bill to being *strongly opposed*. Other instances of belief polarization might shift the content of the belief entirely. Someone might go from thinking that *vaccines cause*

[48] In the UK, the Intelligence and Security Committee of Parliament released its report on Russian interference in 2020. The US Senate Intelligence Committee reached similar conclusions in the reports they released in 2019 and 2020. Both reports detailed the ways that Russia has been using social media to undermine democracy.

[49] We will generally avoid saying that social media "causes" polarization, because social media is certainly not the only cause of polarization. The causal story is multifaceted, and the trend toward polarization began long before the rise of social media.

[50] Although our comments here might apply more broadly, we will generally discuss polarization in the US. Similar things are happening in the UK and Europe, but most of the examples we give will pertain to the US.

autism to thinking *vaccines contain mind-control chips*. Both of these anti-vaccination beliefs are false, but their content is quite different.

Talisse also distinguishes between different kinds of political polarization. For most people, this brings to mind the idea that the distance between political parties has shifted. Either Democrats have moved further left, Republicans have moved to the right, or both. Partisan polarization, by contrast, involves the "ideological uniformity" of the party (Talisse 2019, 98–99). This reinforces the importance of being part of the ingroup, as those who stray from the party's ideological commitments are treated with derision (e.g., calling someone a "RINO," a Republican In Name Only). Finally, there is "affective polarization," which is characterized by high levels of trust within the partisan group, as well as "distrust and antipathy toward the members of opposing groups" (99).

When ideologically uniform groups stick together, deliberate on issues, and discuss their attitudes toward the outgroup, the more they become susceptible to all forms of polarization. Talisse cites a considerable body of evidence that shows how easily this happens. For example, Sunstein et al. (2000) conducted a study of mock jurors in which groups who were disposed toward large punitive damages deliberated together. After deliberating as a group, many individuals agreed to numbers that were substantially higher than their pre-deliberation judgment. This kind of belief polarization happens all the time. Whenever like-minded individuals confer with one another, they are prone to becoming more confident in their beliefs and adopting more extreme forms of those beliefs.

As we explained above, some of the underlying mechanisms here come from features of our psychology. People want to be seen positively by members of the ingroup; they want to have their beliefs confirmed, etc. Social media combines these tendencies with algorithms that show people content they will probably like, which often means showing them things that align with their ideology.[51] It also gives users tools that make it very easy to measure whether or not one's ingroup approves of something (likes, upvotes, retweets, etc.).[52] In some cases, belief corroboration among peers has positive epistemic value. For instance, you might be uncertain whether or not Brad Pitt was in a movie you saw, so you ask Gabriel and he tells you that Pitt was in the movie. You then ask Valeria who also confirms that Pitt was in the movie. In each case, your

[51] See, for example, Halberstam and Knight (2016).
[52] See, for example, Goyanes et al. (2021).

confidence that Pitt was in the movie goes up, and this seems reasonable. As Nguyen (2020) points out, the same cannot be said of confirmation that happens in an echo chamber or filter bubble. He cites Wittgenstein's example of confirming the belief that *p* by looking at identical newspapers that all say *p* (2020, 144). The belief is not legitimately corroborated here because the papers are mere copies. In an echo chamber or filter bubble, people are prone to the same kind of bootstrapped corroboration. This might happen in cases where everyone in the bubble is getting their information from the same source. And even when they do not have the same source, if the informational ecosystem is the product of selective forces that omit certain perspectives, then confirmations from within the bubble should be discounted.

But when combined with affective polarization, this is the opposite of what happens inside an echo chamber. Because affective polarization involves increased trust of the ingroup and increased distrust of the outgroup, corroboration within the group counts for far more than external confirmation. This is why Nguyen (2020) argues that echo chambers are more dangerous than filter bubbles. An informational bubble can be popped simply by exposing someone to the omitted perspectives. But those who are trapped in an echo chamber are predisposed to reject any contrary evidence, and they are likely to distrust any sources that come from outside the chamber.[53]

[53] As Nguyen (2020) explains, filter bubbles are defined as information contexts in which certain views or ideas are omitted. Those who inhabit bubbles of this kind are simply not exposed to alternative perspectives. Social media can produce such informational environments by means of algorithmic filtering or by users' self-sorting. Nguyen cites those like Cass Sunstein (2009) and Eli Pariser (2011) who attribute much of the recent polarization and extremism to filter bubbles. Nguyen points out, however, that bubbles can easily be popped. One simply has to expose people to the omitted perspectives.

Echo chambers, on the other hand, do not simply ignore alternative views; opposing viewpoints (those held by the outgroup) are actively discredited. This is what makes echo chambers more dangerous and harder to eradicate. When someone is in an echo chamber, exposing them to alternative views does *not* weaken their conviction. On the contrary, it tends to bolster it.

The existence of echo chambers helps explain the empirical findings of Bail et al. (2018). They conducted an experiment in which people were exposed to opposing political perspectives for a month. Far from weakening subjects' confidence, this increased the strength of their convictions. They found that the effect was more pronounced on the right than it was on the left. Republicans became "substantially more conservative" after being exposed to opposing viewpoints on Twitter for a month (2018, 9216).

In addition to increasing confidence, these informational environments also contribute to radicalizing members of the group, as they are frequently exposed to content that is exaggerated, hyperpartisan, or simply false.[54] But the spread of fake news and conspiracy theories is not limited to echo chambers. As noted previously, Alfano et al. (2021) were able to show how watching certain categories of content on YouTube prompted the algorithm to start suggesting videos about conspiracies and other problematic content, "[sending] viewers down a conspiratorial rabbit hole" (Alfano et al. 2021, 840). Another study showed how incredibly common it is to be exposed to fake news. In just one month leading up to the 2016 US presidential election, the average American encountered between one to three fake news stories.[55] We have also learned that false information spreads faster than true information, "especially when the topic is politics" (Vosoughi et al. 2018).

What is more, the spread of fake news is harmful even when the content is not believed. Merten Reglitz (2022) explains why this is the case:

> [O]nline fake news threatens democratic values and processes by playing a crucial role in reducing the perceived legitimacy of democratic institutions. This decrease in perceived legitimacy is the outcome of the primary effect that fake news has on citizens: even if its content is not believed, fake news can be a major cause of a loss of citizens' epistemic trust in each other's political views and judgment. Such a loss of trust in each other is problematic for democratic institutions since these rely for their acceptance and functioning on citizens seeing them as morally justified. (Reglitz 2022, 164)

The spread of fake news undermines our ability to trust one another and to engage in collective deliberation. People are less willing to engage with others when they are under the impression that the other person's beliefs are the product of fake news. Our capacity to engage in public reason is dependent on our ability to recognize that other people's reasons have

[54] Ross et al. (2021) show how there is an excessive focus on "fake news" which involves flagrantly false content. They argue that fake new is less widespread than "hyperpartisan" news, which depicts real events but with a strong partisan bias. See also Brady et al. (2021), Rathje et al. (2021), Alfano et al. (2021), and Ribeiro et al. (2020). They present compelling evidence that inflammatory content is more successful at being spread across social media.

[55] Lazer et al. (2018) point out how this is probably a conservative estimate, since the study that produced this number tracked a limited set of fake news stories.

normative force for us as well.[56] The existence of fake news, whether it is believed widely or not, threatens our ability to reason with each other.[57]

Misinformation is not the only kind of polarizing content that spreads quickly on social media, however. Content that sparks outrage is far more likely to spread than content that does not.[58] Once again, the effect is especially pronounced when it comes to politics, and one of the most popular forms of this content is that which depicts political opponents in a way that sparks moral outrage.[59] Nguyen and Williams (2020) argue that seeking gratification through this kind of moral outrage is problematic for several reasons. One reason is that we become incentivized to seek satisfying representations of political opponents rather than truthful ones. This is yet another way that people might corroborate their beliefs with evidence that should be discounted.[60] They write: "This invites a problematic form of circularity—where one picks one's sources based on agreement with one's antecedent beliefs, and then goes on to use those sources to buttress one's antecedent beliefs" (Nguyen and Williams 2020, 162). In addition to making individuals worse off epistemically, when we are

[56] As Korsgaard puts it: "[I]f personal interaction is to be possible, we must reason together, and this means that I must treat your reasons…as reasons, that is, as considerations that have normative force for me as well as you, and therefore as public reasons" (2009, 192).

[57] Although we do not have the space here to explore the question of whether or not fake news is widely believed, there is at least some reason that belief in fake news has had a considerable impact. For instance, there is no credible evidence of widespread voter fraud in the 2020 presidential election, but polls show that as many as 75% of Republicans believe that Trump has a legitimate claim to the presidency because "real cases of fraud changed the results." See Montanaro (2021). It is clear that some people believe fake news. And it is certainly the case that people spread fake news and share it widely. Vosoughi et al. (2018) show that although bots play a role, it is mostly humans who spread fake news: "false news spreads more than the truth because humans, not robots, are more likely to spread it" (2018, 1146).

[58] See Brady et al. (2021), Rathje et al. (2021), Pew Research Center (2017), León and Trilling (2021), Corbu et al. (2020), Frimer et al. (2022), and Wang and Inbar (2022).

[59] See Brady et al. (2021).

[60] Nguyen and Williams write: "But when one is engaged with moral outrage porn, one is seeking out representations of moral outrage for the sake of the resulting gratification, and so one is incentivized to preselect those representations with which one agrees. This invites a problematic form of circularity—where one picks one's sources based on agreement with one's antecedent beliefs, and then goes on to use those sources to buttress one's antecedent beliefs" (2020, 162).

consumed with outrage and incivility, we weaken our ability to govern ourselves democratically.[61]

Some of the most extreme instances of this can be seen by considering recent events in Myanmar, a country whose rapid adoption of social media coincided with explosive violence against an ethnic minority, the Rohingya. The ongoing crisis in Myanmar has been recognized by the U.N. as a genocide in which social media—Facebook, in particular—played a "determining role" (Miles 2018). As we will soon recount, similar events have happened the world over.

The seeds of anti-Rohingya sentiment in Myanmar are at least 100 years old. In the early 1900s, the British colonial rulers of Myanmar—then Burma—imported large numbers of Muslim subjects as part of a "divide and rule" scheme in the majority Buddhist country (Fisher 2022). After the British left in 1948 and a newly independent Burma sought to establish itself, suspicions of the Muslim minority lingered.[62] This sentiment coalesced into a long-standing, government-supported persecution of Rohingya Muslims, who were denied citizenship and as recently as 2014 were not included in the census.[63]

In 2016 these simmering tensions exploded into what the U.N. has recognized as a "textbook example of ethnic cleansing." In his reporting on the atrocity, Max Fisher describes the stomach-turning violence in grim detail:

> The soldiers, sent to exterminate the impoverished minority that many of Myanmar's leaders and citizens had come to see as an intolerable enemy within, would arrive at a village, then begin by setting rooftops afire. They lobbed grenades through hut doorways and sent rockets slamming into the walls of longhouses. They fired into the backs of peasants fleeing across the surrounding fields. As the houses burned, the men of the village would be

[61] Frimer et al. (2022) conducted a study that looked at civility of politicians on Twitter. It found a considerable increase in incivility in the last decade, and they highlight the ways that this might pose a danger to democracy: "[P]olitical incivility can undermine respect for alternative viewpoints, erode public trust in the political process, and incite other forms of uncivil, undemocratic behavior" (2022, 1).

Jonathan Haidt, a well-known critic of the effect of social media on democracy, summarizes this point nicely in an interview: "So long as we are all immersed in a constant stream of unbelievable outrages perpetrated by the other side, I don't see how we can ever trust each other and work together again" (Illing 2018).

[62] See Fisher (2022).

[63] See BBC (2020).

> arrayed in a line and shot to death. Families streamed by the hundred thousand toward the border. The soldiers attacked these too. They hid land mines in the refugees' paths. Survivors who made it to relative safety in Bangladesh detailed horror after horror to journalists and aid workers who picked their way through the overcrowded camps. (Fisher 2022, 158)

With the fraught history between the Rohingya and the government, it is sadly not surprising that the Rohingya were persecuted in 2016 and still are to this day. But the timing and explosion of violence still call out for an explanation, with many—including the U.N.—pointing their finger at social media.[64]

To tell this side of the story, we need to go back in time once more. As late as 2011, Myanmar was heralded as "the last 'unphoned' country in the world," with less than 1% of the population having internet access (Blah 2018). Myanmar's information landscape changed dramatically in 2012 when, as part of an economic liberalization scheme, its telecommunications market was released from a long-standing state-owned monopoly.[65] As early as 2013, the recently "unphoned" country had its state-run newspaper saying that "a person without a Facebook identity is like a person without a home address" (McLaughlin 2018). By 2015, nearly 40% of the population was online, with nearly all internet being accessed via smartphone in a media environment where Facebook was "synonymous with the internet" (Ibid.). And even then, alarm bells were already ringing.

In 2014, the ultra-nationalist Buddhist monk Wirathu was able to use his Facebook megaphone to help turn a rumor—about a Muslim shop owner raping his Buddhist employee—into a deadly riot (Ibid.). Government officials desperate to stop the riots scrambled to get in touch with Facebook about controlling the hate speech on their platform that was fueling the violence. Eventually, the government—not knowing what else to do—decided to simply temporarily block access to Facebook, which almost immediately put an end to the mayhem (Ibid.). Those trying to reach Facebook before the site was blocked reported getting a response from the company only after the violence subsided; the companies' representatives were concerned over the site being unreachable and looking to get it back online (Ibid.).

[64] See Reuters (2018).
[65] See McLaughlin (2018).

Immediately after the riots, there was talk of improving content moderation. At the time of the riots, the community standards had not even been translated into Burmese (and still weren't as late as 14 months after the incident).[66] But problems with hate speech clearly weren't adequately addressed. Muslims, and Rohingya in particular, were still being referred to in derogatory terms and being attached to larger than life rumors to rationalize islamophobia. Examples from 2015 include viral memes telling Rohingya to "keep out" in idiomatic language typically reserved for dogs, and fake news "revealing" Burmese Muslims to be heavily armed. One popular meme—"liked" 8400 times—features a false picture of a stockpile of guns and ammunition (later revealed to be taken in Cairo in 2012). Another meme calling Rohingyas Bengalis (perpetuating the myth that they are illegal immigrants from Bangladesh) and cannibals (accompanied by cartoonishly fake pictures of human butcher shops) gained over 9000 "likes" and 40,000 "shares."[67]

By 2016, Facebook had many signs that it had an unresolved hate speech problem in Myanmar. But it pressed on with its expansion into this growing market by launching its "Free Basics" program, which allowed free access to data so long as it was accessed via the Facebook app (Fisher 2022). Soon after, 38% of people living in Myanmar were getting all or most of the news via the app (Ibid.).

By 2017, thousands of Rohingya, hundreds below the age of five, had been killed. The U.N. stated that social media, Facebook in particular, drove the violence:

> It was used to convey public messages but we know that the ultra-nationalist Buddhists have their own Facebooks and are really inciting a lot of violence and a lot of hatred against the Rohingya or other ethnic minorities … I'm afraid that Facebook has now turned into a beast, and not what it originally intended. (Miles 2018)

In 2018, the site reaffirmed its efforts to control hate speech (Stecklow 2018). Four months later, however, a Reuters report found that Myanmar's corner of Facebook was still awash in anti-Muslim propaganda. The report found over one thousand posts attacking Rohingya and other Muslims. One post mentioned in the report—still up in 2018, despite being posted

[66] Ibid.
[67] See C4ADS (2016).

in 2013—says that the Rohingya must be fought "the way Hitler did the Jews" (Ibid.). Another calls Rohingya "Non-human … dogs" and "Bengalis" (Ibid.). The report mentions that "In early 2015, there were only two people at Facebook who could speak Burmese reviewing problematic posts" and that in 2018, the company didn't have a single employee in Myanmar. It goes on to state that Facebook "continues to rely heavily on users reporting hate speech … because its systems struggle to interpret Burmese text" (Ibid.).

We should add that it isn't clear how many content moderators would be enough, though more would certainly be better. At the time that Reuters wrote their report, Myanmar had 18 million users. And, by Facebook's own admission, their algorithms—like those of many others geared toward promoting engagement—promote and reward divisive content at a very large scale.[68]

In response to all this, one might think that Myanmar is a special case. To show that it is not, let us consider a few other episodes.[69]

Sri Lanka—like Myanmar—is a country that has seen rapid internet adoption via *zero-rating* programs that, like Free Basics, involve generating a user base by exempting certain data from billing (such as data used in-app).[70] Like Myanmar, Sri Lanka has a Muslim ethnic minority that has a fraught relationship with the (Sinhalese, Buddhist) ethnic majority.

A few years into rapid adoption of the internet via free access to social media, familiar patterns began to emerge in Sri Lanka. In 2018, a video of a misunderstanding between a non-Sinhalese-speaking Muslim shop owner and an angry Sinhalese customer was recorded. The Sinhalese customer accused the shop owner of putting sterilization pills in his soup, this being one of the many Islamophobic rumors circulating on social media. The shop owner, trying to placate the customer and not fluent in Sinhalese, inadvertently admitted to the accusation. A video of the altercation went viral on Facebook. Eventually, riots erupted. The government could not contain them until they shut down access to social media. As in Myanmar, this ended the violence and also caught the attention of Facebook

[68] See Horwitz and Seetharaman (2020).

[69] For coverage of these episodes and others that we could not include for reasons of space, see Fisher (2022).

[70] See Fisher (2022, 169).

representatives who had ignored concerns over hate speech. The representatives wanted to know why they had lost traffic to their site.[71]

Sri Lanka and Myanmar are not the only places where these sorts of dynamics have played out. Similar events have unfolded in Ethiopia, for example, Frances Haugen, a whistleblower at Facebook, said, "What we saw in Myanmar and are now seeing in Ethiopia are only the opening chapters of a story so terrifying, no one wants to read the end of it." Haugen states—and we agree with her—that "engagement-based ranking" is "fanning ethnic violence" (Akinwotu 2021).

Her suspicion seems to have been corroborated by Karsten Müller and Carlo Schwarz, two researchers studying anti-refugee attacks and Facebook use in Germany. The team found that in towns where Facebook use was higher, so was anti-refugee violence. In fact, the relationship was disturbingly strong. The *New York Times* sharply summarizes one of their key findings as follows: "Wherever per-person Facebook use rose to one standard deviation above the national average, attacks on refugees increased by about 50 percent" (Taub and Fisher 2018).

Further, Müller and Schwarz were able to provide evidence of a causal relationship. German internet infrastructure is localized, enabling a study of connections between (localized) internet outages and anti-refugee violence. And there is a tight connection between the two. In areas of high Facebook usage, internet outages were associated with precipitous drops in anti-refugee violence (about 35%).[72] This effect, however, was *not* present in outages among communities where internet usage is high but Facebook usage is not, singling out the social network as a key variable in the drop in violence.

When we zoom out a little, the story of social media and democracy might seem so multifaceted that it may seem impossible to connect all of these disparate phenomena into a cohesive narrative. But this is why our argument in this chapter has been couched in terms of trust and trustworthiness. Framing these issues with an eye toward their effect on trust and trustworthiness makes it easier to see how the attention economy threatens democratic legitimacy.[73] Almost all of the issues described in this

[71] Ibid., 175.

[72] Ibid.

[73] Sabatini and Sarracino (2019) evaluated the effect of social media on three kinds of trust: institutional, trust in strangers, and trust in neighbors. They found that "all the forms of trust significantly decrease with participation in online networks" (229).

section connect to trust or trustworthiness in some way; in many cases, distrust is partially constitutive of the phenomenon in question. For instance, echo chambers involve attitudes of distrust toward outside sources, and affective polarization is characterized by a distrust of members of the outgroup and trust of the ingroup. And the spread of fake news and conspiracy theories undermines the government's ability to effectively signal its competence or its responsiveness to citizens (or to even be competent and responsive in the first case).

What is more, there are several feedback loops in the process. Not only is there the loop between trust and trustworthiness, which we discussed above, the different forms of polarization also feed into one another. It is not hard to see how belief polarization might contribute to affective polarization and vice versa. The more that someone distrusts the outgroup and trusts the ingroup (i.e., the more they are affectively polarized), the more likely they are to become belief polarized as well (their beliefs become more extreme and they become more confident in those beliefs). Both forms of polarization make it harder for us to engage in public reason. This is why Talisse comes to the conclusion that polarization strikes at the heart of some of democracy's core commitments: "This is the fundamental problem posed by polarization. Belief polarization directly attacks our capacities to properly enact democratic citizenship, dissolving our abilities to treat our fellow citizens as our political equals. Moreover, belief polarization is part of a larger dynamic by which partisan divisions expand and extremity intensifies, all within a structure of self-perpetuating social dysfunction" (Talisse 2019, 123).

Figure 7.1 is meant to demonstrate how all of these issues are connected to democratic legitimacy.

Although this diagram does not capture every component of the causal story linking social media to democracy, it does illustrate our understanding of the connection between some of the key issues.

Social media combines with certain features of human psychology in such a way that it exacerbates polarization. All of this coalesces to undermine both trust in the government and the government's trustworthiness.

As further evidence of this claim, we will conclude this section with a brief discussion of the ways that these vulnerabilities have been seized upon by those who want to undermine the efficacy of democratic governments. The fact that Russia's Internet Research Agency (IRA) has engaged in this kind of behavior would seem to bolster the claim that polarization

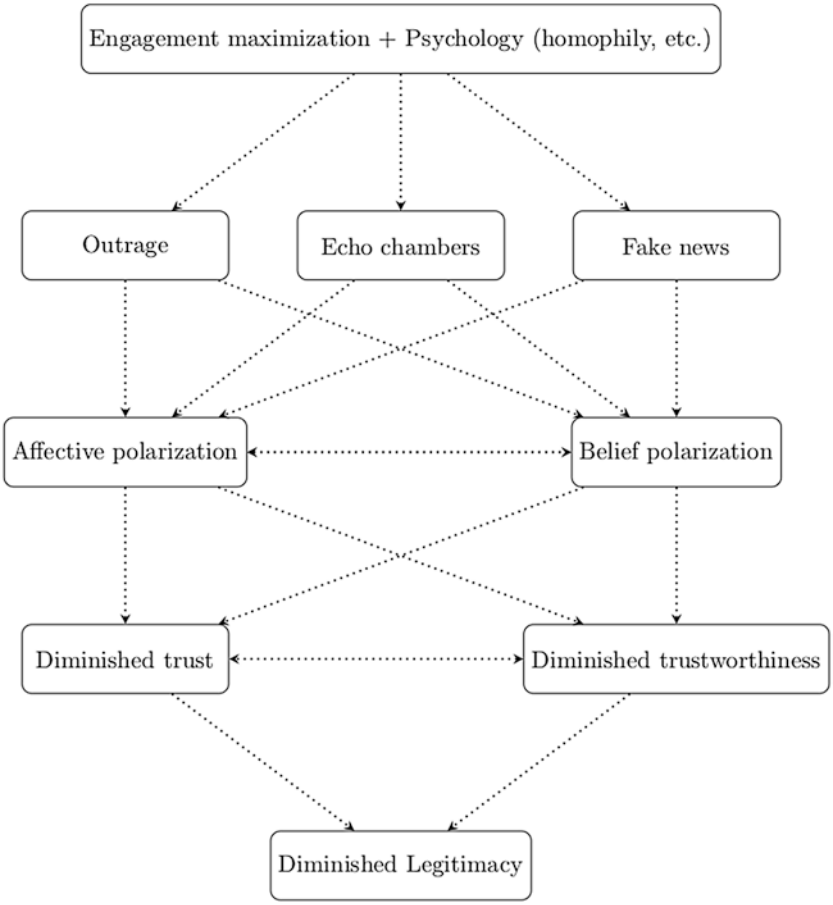

Fig. 7.1 How the attention economy affects democratic legitimacy

can be exacerbated through social media and that this weakens our democracy. Recent reports released by the governments of the UK and US have left little doubt about the IRA's playbook. They also paint a rather bleak picture about the efficacy of these efforts. Between 2015 and 2017, 30 million users of Facebook, Twitter, and Instagram shared, liked, or reacted to content that originated from the IRA.[74] But disseminating fake news is

[74] See Howard et al. (2019).

not their only tactic. The majority of their efforts were dedicated toward stoking polarization.[75] The US Senate report offered the following conclusion:

> The preponderance of the operational focus, as reflected repeatedly in content, account names, and audiences targeted, was on socially divisive issues—such as race, immigration, and Second Amendment rights—in an attempt to pit Americans against one another and against their government. The Committee found that IRA influence operatives consistently used hot-button, societal divisions in the United States as fodder for the content they published through social media in order to stoke anger, provoke outrage and protest, push Americans further away from one another, and foment distrust in government institutions. (U.S. Senate Committee on Intelligence 6)

This was consistent with the findings of the UK report, as they concluded that one of Russia's aims was the "general poisoning of the political narrative in the West by fomenting political extremism and 'wedge issues', and by the 'astroturfing' of Western public opinion; and general discrediting of the West" (Great Britain Intelligence and Security Committee 10).

Once again, the evidence of the efficacy of these efforts is disheartening. The IRA promoted and organized hundreds of rallies, with a particular emphasis on divisive issues (promoting protests for both Black Lives Matter and Blue Lives Matter, for example). One of the most striking successes took place in Houston, Texas, on May 21, 2016, when the IRA organized two events across the street from each other. On one side of the street, there was a "Stop the Islamization of Texas" rally which was organized by a group called "Heart of Texas." On the other side of the street, there was a rally called "Save Islamic Knowledge," organized by "United Muslims of America." Both groups turned out to be IRA accounts.[76] The entire event was orchestrated from St. Petersburg. To make matters worse, the Heart of Texas post encouraged rally goers to bring guns. Luckily, no one was hurt.

[75] See, for example Freelon and Lokot (2020): "State-sponsored disinformation agents have demonstrated success in infiltrating distinct online communities. Political content attracts far more engagement than non-political content *and appears crafted to exploit intergroup distrust and enmity*…Our results make it clear that group identity lies at the core of the IRA's attack strategy. Political audiences were addressed as liberals, conservatives, and Black people to *provoke anger against oppositional outgroups*" (2020, 2, emphasis added).

[76] See Timberg and Dwoskin (2018).

It is very clear *what* Russia's IRA is trying to do. They are using fake news, divisive issues, and inflammatory content in an attempt to polarize Western democracies like the US and UK. It is somewhat less clear *why* they are trying to do this. The UK report concludes that Russia's approach is "fundamentally nihilistic" as they believe "any actions it can take which damage the West are fundamentally good for Russia." Perhaps the idea is that weakening the US and the UK will make it easier for them to accomplish certain foreign policy aims. We can only speculate about the motivation for their activities. But we do not need to speculate about the activities themselves. And it is also clear what kind of effect these actions have. Senator Richard Burr chaired the Senate Intelligence Committee that authored the report on Russia's social media campaign. Rather fittingly, he identified the erosion of trust as one of their fundamental aims: "Russia is waging an information warfare campaign against the U.S. that didn't start and didn't end with the 2016 election. Their goal is broader: to sow societal discord and erode public confidence in the machinery of government. By flooding social media with false reports, conspiracy theories, and trolls, and by exploiting existing divisions, Russia is trying to breed distrust of our democratic institutions and our fellow Americans."[77]

Of course, problems like polarization and echo chambers cannot be attributed entirely to the actions of malicious groups like the IRA. They also cannot be attributed entirely to social media, as the evidence shows that these trends began much earlier. But, as we hope to have shown in this chapter, there is very good reason to believe that our relationship with the attention economy has poured gasoline on these fires. As Harinda Dissanayake, a presidential adviser in Sri Lanka, put it, "We don't completely blame Facebook. The germs are ours, but Facebook is the wind" (Dissanayake 2018). Putting out these fires will not be an easy task, but it is high time for us to start taking the problem seriously. Our capacity for democratic governance hangs in the balance.

7.4 Implications of the Group-Level Duties

Let us now take stock of what we have learned about group-level duties.

First, consider the obligation to be a digital minimalist. Our application of this duty to groups cannot apply as straightforwardly to groups as it does to people, since it is controversial whether groups have "humanity."

[77] US Senate Press Release (2019).

But, as we think that we have established, groups can be or fail to be autonomous. That is, they can succeed or fail to set and pursue their own ends. And—in the case of the groups we are concerned with in this chapter (i.e., democratically governed states)—there is something moral at stake in their autonomy. There are two ways to ground the moral weight of the group's autonomy. It could have intrinsic or final value (which would be a controversial standpoint for us to take, but not one that we necessarily want to reject). Or it could matter instrumentally, which would be much less controversial. Either way, these groups do have an obligation to safeguard their own autonomy and, as we have shown, their autonomy is threatened by the attention economy.

What, then, should groups do to protect themselves? We will have little new to say here, because many of the recommendations have been mentioned previously in this book, albeit under different headings. Groups should think about how their members are educated, for instance, as this, as we have shown, can curb some of the excesses of the attention economy (including excess polarization).[78] Groups can also think about how their members relate to the attention economy, in part by thinking about how that economy is structured and also, for example, how workers and children might be asked to interact with it. Some of these aims might be achieved through regulation (as we argued in Chap. 6), but others might be accomplished through lighter touch means, such as public service announcements.

We can also quickly address groups' obligations to one another to be attention ecologists. This can have both positive and negative sides to it. We have seen how the IRA has worked toward undermining our trust in each other through its active measures. This is clearly problematic, as it undermines the morally important autonomy of groups that is at least partly grounded in its subjects' legitimate interests in self-governance. Perhaps less obviously, it also follows from this that corporate entities such as Facebook have moral reasons to think about how its activities affect other agents. Such companies might not act on these obligations (unless they are compelled), but we take it that we have laid the groundwork for thinking that certain groups (such as states and nations) have justification for compelling corporate entities to refrain from undermining them. Finally, entities such as Facebook also have wide duties to promote the

[78] See Lees and Cikara (2021).

well functioning of nations and states, as these group agents—or their members—have an obligation to *promote* the autonomy of others.

Lastly, we would like to consider the interesting case of the duties that individuals might have to groups. We might have the intuition that individuals just as any other agent—group or otherwise—at least have the duty not to undermine group agents. However, we also likely have the thought that there isn't much we can actually do to promote or undermine such agents, at least not through ordinary activities. They are so large and we are so small.

There is some truth in this—no nation will ever crumble because of how either of us uses our phones—but it might not follow from the fact that we are relatively ineffective and that we have no duties here. There are many philosophical puzzles that bear resemblance to the problem we are discussing here. Climate change is a problem, but each of us, considered individually, won't make the problem any better or any worse. Likewise, factory farming is one of the major moral atrocities of our times. But none of us, individually, can stop it, whether or not we eat meat. Does this mean that we have no individual-level obligations to, say, drive less or become vegetarians?

This issue, which is sometimes referred to as the "problem of collective harm," has become a hot topic in contemporary ethics. It is particularly difficult for consequentialists to explain why individuals have moral obligations to perform certain actions (avoid eating meat, refrain from joyrides in gas-guzzling SUVs, vote in large-scale elections), when it appears that each contribution is causally and morally insignificant.[79] If moral rightness depends *entirely* on the consequences of your actions as an individual, why should you do any of those things?[80] The outcomes appear to

[79] Each of these have been discussed in depth in the collective harm literature. On vegetarianism, see Singer (1980), Norcross (2004), Kagan (2011), Harris and Galvin (2012), Budolfson (2019), and Aylsworth and Pham (2020). On climate change and gas-guzzling joyrides, see Sinnott-Armstrong (2005), Kingston and Sinnott-Armstrong (2018), and Broome (2012). On voting, see Brennan (2011) and Barnett (2020).

[80] This problem is particularly pressing for consequentialists. Shelly Kagan explains: "The problem, in effect, is this: consequentialism condemns my act only when my act makes a difference. But in the kind of cases we are imagining, my act makes no difference, and so cannot be condemned by consequentialism—even though it remains true that when enough such acts are performed the results are bad. Thus consequentialism fails to condemn my act. In cases of this sort, therefore, consequentialism seems to fail even by its own lights" (2011, 108). Kagan goes on to defend a version of the expected utility argument. But his view is challenged by Nefsky (2011) and Budolfson (2019).

be fixed no matter what you do. If we hold the rest of the world fixed and reduce your carbon emissions to zero, there is very little reason to think that this will have any effect on climate change. The same goes for factory farming or large-scale elections.

Given that Kant's ethics does not rely on consequences in the same way, we may be tempted to believe that the categorical imperative yields better results when applied to these situations. But the situation is not so simple. An undergraduate student of Kant's ethics might think that we can solve the problem simply by reflecting on what happens when everyone acts this way. They think of the "universalizability" test like this: We ask what would happen if everyone eats meat and drives gas-guzzling SUVs, and then we realize that such actions are immoral because the consequences are disastrous.

Unfortunately, that is not how the formula of universal law works. We cannot simply evaluate the disastrous consequences of everyone performing some action and rule it out. First, that is closer to rule utilitarianism than it is to Kant's ethics. The categorical imperative requires us to find a contradiction that results from the universal maxim; it is not enough to simply find bad results. Second, if we use the test this way, we could use it to derive a wide variety of faulty prescriptions. As Wood (1999) explains, the test would yield both false positives and false negatives. It would show permissible actions to be wrong and it would allow us to act on impermissible maxims.

The problem turns on the specificity of the circumstances designated in the maxim. This is a routine exercise in any introduction to ethics course. You could ask yourself whether or not you could universalize the maxim of getting coffee at the cafe on Bird Road at 67th Ave, Miami Florida at 8 am on Monday morning, and you would immediately realize that this yields a contradiction. If everyone in the world showed up at this cafe at this time, you would not be able to achieve your end. On the flip side, you could ask whether or not it is permissible to make a false promise on "Tuesday, August 21, to a person Hildreth Milton Flitcraft" Wood (1999, 102). This maxim is so specific that it is possible for you to universalize it. There are so few people who would find themselves in this situation that it does not yield a contradiction when willed to be a universal law.

When we return to the collective action problem, we see that the same issue arises.[81] As Andrew Chignell (2015) points out, there is a crucial difference between asking what would happen if *everyone* performed some action versus asking whether *anyone in relevantly similar circumstances* performed some action. He argues that this difference allows opportunistic carnivores to act on their maxim. By hypothesis, the opportunistic carnivore is someone who eats meat when they are reasonably sure that their action will not make a difference. Surely there is no contradiction in willing that anyone in that circumstance act on such a maxim.

As we saw earlier, the formula of universal law is successful when it shows us that we are making an exception of ourselves. We behave one way while expecting everyone else to behave differently when in similar circumstances. But this is harder to accomplish in cases of collective harm. As Christopher Kutz explains, "The CI [categorical imperative] test works when an individual's maxim can be realized only when it is exceptional, not when it, on the contrary, owes its success to the fact that others act in precisely the same way" (2000, 8). Typically the test works when you are performing an immoral action and you realize that your end can be achieved only if others refrain from performing this action. But in cases of collective harm, the opposite is true. Your action appears to be permissible precisely because you are doing what everyone else is doing. Your action is not exceptional at all.

But this does not mean that Kant's ethics has nothing to say about what we should do in these situations. In a recent paper, Maike Albertzart argues that Kant's ethics can respond to these issues by drawing on the duty of beneficence. As you may recall from previous chapters, this duty requires us to adopt the happiness of others as one of our ends, and Kant's theory of agency tells us that willing an end involves willing the necessary means to that end. Albertzart writes, "To adopt the happiness of others as an end implies willing the necessary means for achieving this end. Given the negative impact climate change is expected to have on human happiness, combating climate change qualifies as one of these necessary means" (2019, 844).

Albertzart asks us to imagine an agent who wants to fight climate change as a way of promoting the happiness of those who will be negatively affected by it. In order to will that end, she wills one of the necessary

[81] Shafer-Landau (1994) raises this issue for the universalizability test as it applies to vegetarianism.

means (viz. that everyone avoid unnecessary car trips). When she takes an unnecessary car trip, she acts in a way that is inconsistent. She cannot universalize the maxim of her action because it "contradicts her chosen means for combating climate change as part of the obligatory end of the happiness of others" (Ibid.).

We believe that Albertzart's solution is on the right track, but it seems to move a little too quickly when determining the necessary means to the end of fighting climate change. Strictly speaking, to fight climate change, we do not need *everyone* to refrain from unnecessary carbon-emitting activities. We only need a critical mass of people to do so. It is not true that we must require everyone to avoid unnecessary car trips. In order to will the end of mitigating the climate crisis, we need to will only that a sufficiently large number of people reduce their carbon emissions.

But this is where the formula of universal law demonstrates the problem with making an exception of oneself. Consider the case of overfishing—another collective harm problem. If one person ignores the fishing limit, this is unlikely to have any effect on the stability of the fish population. However, if a sufficiently large group of people take too many fish, the population will cross a critical threshold and collapse. The fish will not be able to repopulate the lake. In order to sustain the ecosystem, we do not have to will that *everyone* refrain from taking more than the limit. Instead, we must will that a sufficiently large set of anglers refrain from doing this. Imagine that you are fishing on the lake and you have willed the end of a sustainable fish population. When you decide to exceed your limit, you make an exception of yourself. You want a critical number of people to refrain from doing precisely what you are doing. You realize that the success of the end you have willed depends on other people not performing this action.

That is the Kantian response to the problem of collective harm. First, it is routed through the duty of beneficence. We can think of an end that we are required to adopt, and then we develop an understanding of the necessary means to that end. Second, we come to see how defecting when we expect others to cooperate involves making an exception of ourselves and is thus inconsistent with the formula of universal law.

When applied to the problems in this chapter, we can see how useful this would be. Given the massive scope of polarization, moral outrage, and fake news, we may be tempted to think that each person plays an insignificantly small role. The action of a single individual is too small to make a difference. But now we see the error of thinking exclusively about the

consequences of an individual action. Instead, we should think about how we are committed to certain social ends. We want healthy democratic governance; we know that this involves trust, trustworthiness, and social cooperation. In order to evacuate efficiently, we need people to follow directions when the National Hurricane Center tells them to evacuate or stay put. If we want to control pandemics, then we need them to trust regulatory bodies who are telling us that a vaccine is safe and effective.

If we will those ends, we must also will the necessary means to those ends. And that requires us to avoid spreading misinformation on the internet, to refrain from sharing inflammatory content that provokes moral outrage and distrust, and possibly to disengage from the attention economy in general. Given the empirical evidence on the radicalizing effect of the attention economy, willing the ends of a rightful civil condition requires us to will that a critical mass of people unplug from social media. There is a common thread uniting the genocide in Myanmar, the ethnic violence in Sri Lanka and Germany, and the distrust of health information about the Zika virus in Brazil.[82] Each of these were fueled, in part, by the fact that huge populations were radicalized by social media and the algorithms' aim of capturing as much attention as possible.

As individuals, we may not be able to put out these forest fires of collective harm. But that does not mean that we have no obligation to do our part. If we will the end, we must will the necessary means. And we must not make exceptions for ourselves. Our duty to be digital minimalists is not just something that we owe to ourselves. It is something that we owe to the groups to which we belong. In some cases, the autonomy of those groups matters. We want to live in functional, democratic states, and we want those institutions to succeed at setting and pursuing their own ends. If we succumb to the temptations of the attention economy, we will go along with everyone else in pouring gasoline on the fires of polarization, outrage, and distrust. We will undermine the necessary means to our chosen ends.

7.5 Conclusion

In this chapter, we have argued that the attention economy threatens to harm us not only as individuals but also as a group. Structured groups, like group agents, are susceptible to a distinctive form of harm insofar as they

[82] See Fisher (2022).

have shared interests that they pursue collectively. Democratic states are a paradigmatic case of a group agent that is liable to this kind of harm. And given the nature of democratic legitimacy, trust and trustworthiness are key components of the state's agency. Without trust, the government is unable to achieve its aims or fulfill its commitments. Various features of the attention economy threaten to undermine trust in the government and the government's trustworthiness. Western democracies (like the US and the UK) caught on to this problem only after serious damage had been done. We would do well to remain vigilant about this threat in the future. That is why it is so important to understand what is at stake and why it is of vital moral significance.

References

Akinwotu, Emmanuel. 2021. Facebook's role in Myanmar and Ethiopia under new scrutiny. *The Guardian*. https://www.theguardian.com/technology/2021/oct/07/facebooks-role-in-myanmar-and-ethiopia-under-new-scrutiny. Accessed 17 July 2023.

Albertzart, Maike. 2019. A Kantian solution to the problem of imperceptible differences. *European Journal of Philosophy* 27, Nr. 4 (December): 837–851. https://doi.org/10.1111/ejop.12456

Alfano, Mark, Amir Ebrahimi Fard, J. Adam Carter, Peter Clutton, and Colin Klein. 2021. Technologically scaffolded atypical cognition: The case of YouTube's recommender system. *Synthese* 199: 835–858.

Allcott, Hunt, Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow. 2020. The welfare effects of social media. *American Economic Review* 110: 629–676.

Altman, Andrew, and Christopher Heath Wellman. 2004. A defense of international criminal law. *Ethics* 115: 35–67.

Appiah, Anthony. 2005a. *The ethics of identity*. Princeton: Princeton University Press.

———. 2005b. *Thinking it through*. Oxford: Oxford University Press.

———. 2011. 'Group rights' and racial affirmative action. *The Journal of Ethics* 15: 265–280.

Asimovic, Nejla, Jonathan Nagler, Richard Bonneau, and Joshua A. Tucker. 2021. Testing the effects of Facebook usage in an ethnically polarized setting. *Proceedings of the National Academy of Sciences* 118: e2022819118.

Aylsworth, Tim, and Adam Pham. 2020. Consequentialism, collective action, and causal impotence. *Ethics, Policy & Environment* 23: 336–349.

Baier, Annette. 1986. Trust and antitrust. *Ethics* 96: 231–260.

Bail, Christopher A., Lisa P. Argyle, Taylor W. Brown, John P. Bumpus, M.B. Haohan Chen, Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky. 2018. Exposure to opposing views on

social media can increase political polarization. *Proceedings of the National Academy of Sciences* 115: 9216–9221.
Barberá, Pablo. 2015. Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data. *Political Analysis* 23: 76–91.
Barnett, Zach. 2020. Why you should vote to change the outcome. *Philosophy & Public Affairs* 48: 315–447.
Bavel, Van, J. Jay, Steve Rathje, Elizabeth Harris, Claire Robertson, and Anni Sternisko. 2021. How social media shapes polarization. *Trends in Cognitive Sciences* 25: 913–916.
BBC. 2020. Myanmar Rohingya: What you need to know about the crisis. *BBC*. https://www.bbc.com/news/world-asia-41566561
Blah, Dan. 2018. Access and openness: Myanmar 2012. *Open Technology Fund*. https://www.opentech.fund/news/access-and-openness-myanmar-2012/
Boxell, Levi, Matthew Gentzkow, and Jesse M. Shapiro. 2017. Greater Internet use is not associated with faster growth in political polarization among US demographic groups. *Proceedings of the National Academy of Sciences* 114: 10612–10617.
Brady, William J., Killian McLoughlin, Tuan N. Doan, and Molly J. Crockett. 2021. How social learning amplifies moral outrage expression in online social networks. *Science Advances* 7: eabe5641.
Bratman, Michael. 1992. Shared cooperative activity. *The Philosophical Review* 101: 327–341.
———. 1993. Shared intention. *Ethics* 104: 97–113.
Brennan, Jason. 2011. *The ethics of voting*. Princeton: Princeton University Press.
Broome, John. 2012. *Climate matters: Ethics in a warming world*. New York: W. W. Norton & Company.
Budnik, Christian. 2018. Trust, reliance, and democracy. *International Journal of Philosophical Studies* 26: 221–239.
Budolfson, Mark Bryant. 2019. The inefficacy objection to consequentialism and the problem with the expected consequences response. *Philosophical Studies* 176: 1711–1724.
C4ADS. 2016. Sticks and Stones: Hate speech narratives and facilitators in Myanmar. https://c4ads.org/reports/sticks-and-stones/. Accessed 17 July 2023.
Castro, Clinton, and Adam Pham. 2020. Is the attention economy noxious? *Philosophers' Imprint* 20 (17): 1–13.
Chignell, Andrew, 2015. Can We Really Vote with Our Forks? Opportunism and the Threshold Chicken. In: *Philosophy Comes to Dinner*, Edited by Andrew Chignell, Matthew C. Halteman, und Terence Cuneo, 182–202. Routledge.

Cho, Jaeho, Saifuddin Ahmed, Martin Hilbert, Billy Liu, and Jonathan Luu. 2020. Do search algorithms endanger democracy? An experimental investigation of algorithm effects on political polarization. *Journal of Broadcasting & Electronic Media* 64: 150–172.

Cinelli, Matteo, Gianmarco De Francisci Morales, Alessandro Galeazzi, Walter Quattrociocchi, and Michele Starnini. 2021. The echo chamber effect on social media. *Proceedings of the National Academy of Sciences* 118: e2023301118.

Cook, Karen S., Russell Hardin, and Margaret Levi. 2005. *Cooperation without trust?* New York: Russell Sage Foundation.

Corbu, Nicoleta, Alina Bârgăoanu, Raluca Buturoiu, and Oana Ștefăniță. 2020. Does fake news lead to more engaging effects on social media? Evidence from Romania. *Communications* 45: 694–717.

Crawford, James. 2006. *The creation of states in international law*. Oxford: Clarendon Press.

Davis, Dena S. 2000. Groups, communities, and contested identities in genetic research. *Hastings Center Report* 30: 38–45.

Dissanayake, Harindra. 2018. Quotation of the day: Where Facebook rumors fuel thirst for revenge. *The New York Times*. https://www.nytimes.com/2018/04/21/todayspaper/quotation-of-the-day-where-facebook-rumors-fuel-thirst-for-revenge.html. Accessed 17 July 2023.

Fenton, Justin. 2021. *We own this city: A true story of crime, cops, and corruption*. New York: Random House.

Finkel, Eli J., Christopher A. Bail, Mina Cikara, Peter H. Ditto, Shanto Iyengar, Samara Klar, Lilliana Mason, Mary McGrath, Brendan Nyhan, David Rand, Linda Skitka, Joshua Tucker, Jay Van Bavel, Cynthia Wang, and James Druckman. 2020. Political sectarianism in America. *Science* 370: 533–536.

Fisher, Max. 2022. *The chaos machine: The inside story of how social media rewired our minds and our world*. New York: Little, Brown and Company.

Franck, Thomas. 1992. The emerging right to democratic governance. *American Journal of International Law* 86: 46–91.

Freelon, Deen, and Tetyana Lokot. 2020. Russian disinformation campaigns on Twitter target political communities across the spectrum. Collaboration between opposed political groups might be the most effective way to counter it. *Harvard Kennedy School Misinformation Review* https://misinforeview.hks.harvard.edu/article/russian-disinformation-campaigns-on-twitter/. Accessed 24 July 2023.

Frimer, Jeremy A., Harinder Aujla, Matthew Feinberg, Linda J. Skitka, Karl Aquino, Johannes C. Eichstaedt, and Robb Willer. 2022. Incivility is rising among American politicians on Twitter. *Social Psychological and Personality Science* 194855062210838.

Gilbert, Margaret. 1992. *On social facts*. Princeton: Princeton University Press.

———. 2006. *A theory of political obligation: Membership, commitment, and the bonds of society*. Oxford: Oxford University Press.

———. 2013. *Joint commitment: How we make the social world*. Oxford: Oxford University Press.

Goyanes, Manuel, Porismita Borah, Homero Gil, and de Zúñiga. 2021. Social media filtering and democracy: Effects of social media news use and uncivil political discussions on social media unfriending. *Computers in Human Behavior* 120: 106759.

Great Britain and Intelligence and Security Committee. 2020. *Russia: Presented to Parliament pursuant to section 3 of the Justice and Security Act 2013*. https://isc.independent.gov.uk/wp-content/uploads/2021/03/CCS207_CCS0221966010-001_Russia-Report-v02-Web_Accessible.pdf. Accessed 24 July 2023.

Haidt, Jonathan, and Chri Bail. n.d. Social media and political dysfunction: A collaborative review. https://jonathanhaidt.com/social-media/. Accessed 17 July 2023.

Halberstam, Yosh, and Brian Knight. 2016. Homophily, group size, and the diffusion of political information in social networks: Evidence from Twitter. *Journal of Public Economics* 143: 73–88.

Hardin, Russell. 2002. *Trust and trustworthiness*. New York, NY: Russell Sage Foundation.

Harris, John Richard, and Richard Galvin. 2012. 'Pass the cocoamone, please': Causal impotence, opportunistic vegetarianism and act-utilitarianism. *Ethics, Policy & Environment* 15: 368–383.

Hausman, Daniel M. 2007. Group risks, risks to groups, and group engagement in genetics research. *Kennedy Institute of Ethics Journal* 17 (4): 351–369.

Hawley, Katherine. 2014. Trust, distrust and commitment. *Noûs* 48: 1–20.

———. 2017. Trustworthy groups and organizations. In *The philosophy of trust*, ed. Paul Faulkner and Thomas Simpson, 52–70. Oxford: Oxford University Press.

Haywood, William Dudley. 1929. *The autobiography of Big Bill Haywood*. New York: International Publishers.

Hetherington, Marc J. 2008. Turned off or turned on? The effects of polarization on political participation, engagement, and representation. In *Red and blue nation*, volume 2: 1–33. Stanford: Brookings Institution Press and Hoover Institution.

Hetherington, Marc J., and Jason A. Husser. 2012. How trust matters: The changing political relevance of political trust. *American Journal of Political Science* 56: 312–325.

Hetherington, Marc J., and Thomas J. Rudolph. 2017. Political trust and polarization. In *The Oxford handbook of social and political trust*, ed. Eric M. Uslaner. Oxford: Oxford University Press.

Hong, Sounman, and Sun Hyoung Kim. 2016. Political polarization on Twitter: Implications for the use of social media in digital governments. *Government Information Quarterly* 33: 777–782.

Horwitz, Jeff, and Deepa Seetharaman. 2020. Facebook executives shut down efforts to make the site less divisive. *The Wall Street Journal*. https://www.wsj.com/articles/facebook-knows-it-encourages-division-top-executives-nixed-solutions-11590507499. Accessed 17 July 2017.

Howard, Philip, Bharath Ganesh, Dimitra Liotsiou, John Kelly, and Camille François. 2019. The IRA, social media and political polarization in the United States, 2012–2018. U.S. Senate Documents, October. https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1004&context=senatedocs

Illing, Sean. 2018. Why social media is terrible for multiethnic democracies. *Vox*. November 15, 2016. https://www.vox.com/policy-and-politics/2016/11/15/13593670/donald-trump-social-media-culture-politics

Jones, Karen. 1996. Trust as an Affective Attitude. *Ethics* 107(1): 4–25. https://doi.org/10.1086/233694

———. 2012. Trustworthiness. *Ethics* 123: 61–85.

Kagan, Shelly. 2011. Do I make a difference? *Philosophy & Public Affairs* 39: 105–141.

Kant, Immanuel. 2007. *Anthropology, history, and education*, eds. Günter Zöller and Robert B. Louden. Cambridge University Press.

Kingston, Ewan, and Walter Sinnott-Armstrong. 2018. What's wrong with joyguzzling? *Ethical Theory and Moral Practice* 21: 169–186.

Kirby, Nikolas, Andrew Kirton, and Aisling Crean. 2018. Do corporations have a duty to be trustworthy? *Journal of the British Academy* 6: 75–129.

Klein, Elad and Joshua Robison. 2020. Like, Post, and Distrust? How Social Media Use Affects Trust in Government. *Political Communication* 37: 46–64. https://doi.org/10.1080/10584609.2019.1661891

Korsgaard, Christine. 2009. *Self-constitution: Agency, identity, and integrity*. Oxford: Oxford University Press.

Kutz, Christopher, 2000. *Complicity: Ethics and Law for a collective age*. Cambridge University Press.

Lazer, David M.J., Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger, Brendan Nyham, Gordon Pennycook, David Rothschild, Michael Schudson, Steven Sloman, Cass Sunstein, Emily Thorson, Duncan Watts, and Jonathan Zittrain. 2018. The science of fake news. *Science* 359: 1094–1096.

Lees, Jeffrey, and Mina Cikara. 2021. Understanding and combating misperceived polarization. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 376: 20200143.

León, Ernesto de, and Damian Trilling. 2021. A sadness bias in political news sharing? The role of discrete emotions in the engagement and dissemination of political news on Facebook. *Social Media + Society* 7: 20563051211059710.

List, Christian, and Philip Pettit. 2011. *Group agency: The possibility, design, and status of corporate agents*. Oxford: Oxford University Press.

Ludwig, Kirk. 2016. *From individual to plural agency: Collective action: Volume 1*. Oxford: Oxford University Press.

Margalit, Avishai, and Joseph Raz. 1990. National self-determination. *The Journal of Philosophy* 87: 439–461.

May, Larry. 2000. *Crimes against humanity: A normative account*. Cambridge: Cambridge University Press.

McLaughlin, Timothy. 2018. How Facebook's rise fueled chaos and confusion in Myanmar. *Wired*. https://www.wired.com/story/how-facebooks-rise-fueled-chaos-and-confusion-in-myanmar/. Accessed 17 July 2023.

Miles, Tom. 2018. U.N. investigators cite Facebook role in Myanmar crisis. *Reuters*. https://www.reuters.com/article/us-myanmar-rohingya-facebook-idUKKCN1GO2PN. Accessed 17 July 2023.

Montanaro, Domenico. 2021. Most Americans trust elections are fair, but sharp divides exist, a new poll finds. *NPR*. https://www.npr.org/2021/11/01/1050291610/most-americans-trust-elections-are-fair-but-sharp-divides-exist-a-new-poll-finds. Accessed 17 Aug 2022.

Mosleh, Mohsen, Cameron Martel, Dean Eckles, and David G. Rand. 2021. Shared partisanship dramatically increases social tie formation in a Twitter field experiment. *Proceedings of the National Academy of Sciences* 118: e2022761118.

Narveson, Jan. 1991. Collective rights? *Canadian Journal of Law and Jurisprudence* 4: 329–345.

Nefsky, Julia. 2011. Consequentialism and the Problem of Collective Harm: A Reply to Kagan: Consequentialism and the Problem of Collective Harm. *Philosophy & Public Affairs* 39 (4): 364–395. https://doi.org/10.1111/j.1088-4963.2012.01209.x

Nguyen, C. Thi. 2020. Echo chambers and epistemic bubbles. *Episteme* 17: 141–161.

———. 2022. Trust as an unquestioning attitude. *Oxford Studies in Epistemology* 7: 214–44.

Nguyen, C. Thi, and Bekka Williams. 2020. Moral outrage porn. *Journal of Ethics and Social Philosophy* 18: 147–172.

Norcross, Alastair. 2004. Puppies, pigs, and people: Eating meat and marginal cases. *Philosophical Perspectives* 18: 229–245.

Orlowski, Jeff. 2020. We need to rethink social media before it's too late. We've accepted a Faustian bargain. *The Guardian*. https://www.theguardian.com/commentisfree/2020/sep/27/social-dilemma-media-facebook-twitter-society. Accessed 17 July 2023.

Pariser, Eli. 2011. *The filter bubble: What the Internet is hiding from you*. London: Penguin UK.

Peter, Fabienne. 2017. Political legitimacy. In *The Stanford encyclopedia of philosophy*, ed. Edward N. Zalta. https://plato.stanford.edu/archives/sum2017/entries/legitimacy/. Accessed 17 Aug 2022.

Pew Research Center. 2017. Critical posts get more likes, comments, and shares than other posts. https://www.pewresearch.org/politics/2017/02/23/partisan-conflict-and-congressional-outreach/pdl-02-23-17_antipathy-new-00-02/. Accessed 17 July 2023.

Pogge, Thomas W. 1992. Cosmopolitanism and sovereignty. *Ethics* 103: 48–75.

Purves, Duncan, and Jeremy Davis. 2022. Public trust, institutional legitimacy, and the use of algorithms in criminal justice. *Public Affairs Quarterly* 36: 136–162.

Rathje, Steve, Jay J. Van Bavel, and Sander van der Linden. 2021. Out-group animosity drives engagement on social media. *Proceedings of the National Academy of Sciences* 118: e2024292118.

Reglitz, Merten. 2022. Fake news and democracy. *Journal of Ethics and Social Philosophy* 22: 162–187.

Reuters. 2018. Myanmar: UN blames Facebook for spreading hatred of Rohingya. *The Guardian*. https://www.theguardian.com/technology/2018/mar/13/myanmar-un-blames-facebook-for-spreading-hatred-of-rohingya. Accessed 17 July 2023.

Ribeiro, Manoel Horta, Raphael Ottoni, Robert West, Virgílio A. F. Almeida, and Wagner Meira. 2020. Auditing radicalization pathways on YouTube. In: *Proceedings of the 2020 conference on fairness, accountability, and transparency*, 131–141.

Ross, R. M., Rand, D. G., & Pennycook, G. (2021). Beyond "fake news": Analytic thinking and the detection of false and hyperpartisan news headlines. Judgment and Decision Making, 16(2), 484–504.

Rubel, Alan, Clinton Castro, and Adam Pham. 2021. *Algorithms and autonomy: The ethics of automated decision systems*. Cambridge: Cambridge University Press.

Sabatini, Fabio, and Francesco Sarracino. 2019. Online social networks and trust. *Social Indicators Research* 142: 229–260.

Sasahara, Kazutoshi, Wen Chen, Hao Peng, Giovanni Luca Ciampaglia, Alessandro Flammini, and Filippo Menczer. 2021. Social influence and unfollowing accelerate the emergence of echo chambers. *Journal of Computational Social Science* 4: 381–402.

Searle, John. 1990. Collective intentions and actions. In *Intentions in communication*, ed. P. Cohen, J. Morgan, and M. Pollack, 401–415. Cambridge: MIT Press.

Shafer-Landau, Russ. 1994. Vegetarianism, Causation and Ethical Theory. *Public Affairs Quarterly* 8 (1): 85–100.

Singer, Peter. 1980. Utilitarianism and vegetarianism. *Philosophy & Public Affairs* 9: 325–337.

Sinnott-Armstrong, W. 2005. It's not my fault: Global warming and individual moral obligations. In *Perspectives on climate change: Science, economics, politics, ethics*, ed. W. Sinnott-Armstrong and R. Howarth, 285–307. Amsterdam: Elsevier.

Stecklow, Steve. 2018. Hatebook: Inside Facebook's Myanmar operation. *Reuters*. https://www.reuters.com/investigates/special-report/myanmar-facebook-hate/. Accessed 17 July 2023.

Sterba, James P. 2009. *Affirmative action for the future*. Ithaca: Cornell University Press.

Sunstein, Cass R. 1999. *The law of group polarization*. SSRN Scholarly Paper. Rochester, New York.

———. 2009. *Going to extremes: How like minds unite and divide*. Oxford: Oxford University Press.

Sunstein, Cass, Daniel Kahneman, and David Schkade. 2000. Deliberating about dollars: The severity shift empirical study. *Columbia Law Review* 100: 1139–1175.

Talisse, Robert B. 2019. *Overdoing democracy: Why we must put politics in its place*. Oxford: Oxford University Press.

Tasioulas, John, and Guglielmo Verdirame. 2022. Philosophy of international law. In *The Stanford encyclopedia of philosophy*, ed. Edward N. Zalta. https://plato.stanford.edu/archives/sum2022/entries/international-law/. Accessed 24 July 2023.

Taub, Amanda, and Max Fisher. 2018. Facebook fueled anti-refugee attacks in Germany, new research suggests. *The New York Times*. https://www.nytimes.com/2018/08/21/world/europe/facebook-refugee-attacks-germany.html. Accessed 17 July 2023.

Timberg, Craig, and Elizabeth Dwoskin. 2018. Russians got tens of thousands of Americans to RSVP for their phony political events on Facebook. *The Washington Post*. Accessed 15 Aug 2022. https://www.washingtonpost.com/news/the-switch/wp/2018/01/25/russians-got-tens-of-thousands-of-americans-to-rsvp-for-their-phony-political-events-on-facebook/. Accessed 24 July 2023.

Tufekci, Zeynep. 2018. How social media took us from Tahrir Square to Donald Trump. *MIT Technology Review*. https://www.technologyreview.com/2018/08/14/240325/how-social-media-took-us-from-tahrir-square-to-donald-trump/. Accessed 17 July 2023.

Tuomela, Raimo. 2013. *Social ontology: Collective intentionality and group agents*. Oxford: Oxford University Press.

Tyler, Tom. 2006. *Why people obey the law*. Princeton: Princeton University Press.

Tyler, Tom, and Y. Huo. 2002. *Trust in the law: Encouraging public cooperation with the police and courts*. New York: Russell Sage Foundation.

Tyler, Tom, and J. Jackson. 2014. Popular legitimacy and the exercise of legal authority: Motivating compliance, cooperation, and engagement. *Psychology, Public Policy, and Law* 20: 78–95.

U.S. Senate Committee on Intelligence. 2020. Russian active measures campaigns and interference in the 2016 U.S. election. Volume 2: Russia's use of

social media with additional views. https://www.intelligence.senate.gov/publications/report-select-committee-intelligence-united-states-senate-russian-active-measures

U.S. Senate Press Release. 2019 Senate Intel Committee releases bipartisan report on Russia's use of social media. https://www.intelligence.senate.gov/press/senate-intel-committee-releases-bipartisan-report-russia%E2%80%99s-use-social-media. Accessed 17 Aug 2022.

Uslaner, Eric M. 2018. *The Oxford handbook of social and political trust*. Oxford University Press.

Velleman, J. David. 1997. How to share an intention. *Philosophy and Phenomenological Research* 57: 29–50.

Vosoughi, Soroush, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. *Science* 359: 1146–1151.

Walzer, Michael. 1977. *Just and unjust wars*. New York: Basic Books.

Wang, S.-Y.N., and Y. Inbar. 2022. Re-examining the spread of moralized rhetoric from political elites: Effects of valence and ideology. *Journal of Experimental Psychology: General* 151: 3292–3303.

Weijer, Charles, Gary Goldsand, and Ezekiel Emanuel. 1999. Protecting communities in research: Current guidelines and limits of extrapolation. *Nature Genetics* 23: 275–280.

Wood, Allen, 1999. *Kant's ethical thought*. New York: Oxford University Press.

CHAPTER 8

Conclusion

It is delightful to imagine that human nature will be developed better and better by means of education, and that the latter can be brought into a form appropriate for humanity. This opens to us the prospect of a future happier human species.
—Kant (Lectures on Pedagogy 9:444)

We need deep, systemic reform that will shift technology corporations to serving the public interest first and foremost. We have to think bigger about how much systemic change might be possible, and how to harness the collective will of the people … Rather than accepting a race to the bottom that downgrades and divides us, we can together create a technology landscape that enables a race to the top—one that supports our interconnection, civility, and deep brilliance. Change, I believe, is humanly possible.
—Tristan Harris (This quote comes from an essay Harris wrote for the MIT Technology Review. See Harris (*2021*).)

8.1 Polestar Cases Revisited

In the introduction, we issued a promissory note that we would revisit our three polestar cases in order to demonstrate the importance of the lessons we learned throughout the book. First, there was Esther, who struggled to accomplish her aim of reading because she kept falling prey to the temptation to get out her phone and check Instagram. Second, there was Monica, who felt a Pavlovian pull to check a Slack notification, and she found the

T. Aylsworth, C. Castro, *Kantian Ethics and the Attention Economy*,
https://doi.org/10.1007/978-3-031-45638-1_8

constant interruptions were taking a toll on her productivity and her peace of mind. Finally, there was Damon, who was alarmed to discover that his Facebook feed had devolved into a "battleground among partisan echo chambers."

The resources of this book should allow us to see that these are not disparate phenomena. On the contrary, each situation demonstrates different ways that our autonomy (both at the individual and group level) is being undermined by our relationship with the attention economy. Esther's situation exemplifies the importance of the duty we owe to ourselves to be digital minimalists. If Kant is right about the moral weight of humanity (the capacity to set and pursue our own ends), then we should recognize that we have a duty to ourselves to be more intentional about our relationship with mobile devices and social media.[1] These are powerful tools of behavior modification that can undermine our capacities and our authenticity.

This duty to ourselves is similar to what Carol Hay called the "obligation to resist one's own oppression" (2011). If we acquiesce in our own oppression, we demonstrate a failure to respect our own autonomy. As she puts it, "If Kant is right and our rational nature has ultimate value, then we ought to protect this nature by protecting all of it, including our capacity to act rationally. Oppression can harm rational capacities in a number of ways…" (23). In Chap. 3, we saw the many ways that the problematic use of mobile devices and social media can harm our rational capacities. This may not be tantamount to oppression, but the duty to resist it is grounded in the very same reasons.

The duty to resist technological heteronomy also resembles Thomas Hill's discussion of servility and self-respect (Hill 1973). Hill argues that "The objectionable feature of the servile person, as I have described him, is his tendency to disavow his own moral rights either because he misunderstands them or because he cares little for them" (97). He also points out how other people stand to gain from those who sheepishly disavow their own rights: "A submissive attitude encourages exploitation, and exploitation spreads misery in a variety of ways…When people refuse to press their rights, *there are usually others who profit*" (90, emphasis added).

[1] As Kant says in the *Metaphysics of Morals*, "Be an honorable human being (*honeste vive*). Rightful honor (*honestas iuridica*) consists in asserting one's worth as a human being in relation to others, a duty expressed by the saying 'Do not make yourself a mere means for others but be at the same time an end for them'" (MS 6:236).

In Chap. 5, we explained why the problem of the attention economy should not be understood entirely as one of individual responsibility. Those who seek to profit from capturing our attention have deployed a variety of manipulative tactics to undermine the autonomy of others. Esther notes the feeling of what Eyal (2014) calls "internal triggers." As Eyal explains, successful products get us to perform behaviors (such as unlocking our phones and opening apps) before we've even had a moment to consciously reflect on what we are doing. Monica, on the other hand, was disrupted by external triggers—notifications that reminded her to check in with the kind of communication platform that many employers require employees to use.

Once we appreciate the social nature of the problem, we should see that it goes far beyond individuals and the duties they owe to themselves. Parents have duties to safeguard their children from the dangers of addictive technology, and teachers have similar duties to their students. In both cases, the cultivation of autonomy is a constitutive aim of the activity. Successful parents and teachers are ones who protect the child's "right to an open future." All human beings have an innate right to set and pursue their own ends. But we jeopardize that right by allowing children and students to fall prey to the seemingly irresistible pull of the dopamine-driven attention economy.

For this very reason, we also argued that the state may be justified in using its coercive power to hinder some of these hindrances to freedom. As people have become increasingly aware of the scope of this problem, promising legislation has begun to emerge around the globe. France gave workers the "right to disconnect" by requiring employers to negotiate tech policies with employees. Utah requires tech companies to get parental permission before allowing children to open social media accounts. Schools and daycares are beginning to reconsider their policies about screen time. After recognizing that this is a social problem, we must start looking for collective solutions. We can all play our part in this saga of moral progress by being digital minimalists and attention ecologists.

Finally, we should now appreciate the importance of Damon's concern about his Facebook feed. What he saw was far more than a lamentable disagreement between his family and friends. It was a paradigmatic illustration of the breakdown of the kinds of social trust that are necessary for us to live in a rightful condition with one another. Once we have developed an appreciation of the ways that social media and mobile devices have fueled the fires of polarization and moral outrage, we must recognize

that our commitment to democracy requires us to act. We cannot sit idly by as these vital institutions crumble under the weight of forces we have created. From the genocide in Myanmar to the riots on Capitol Hill, we can no longer ignore the scope of the problem (Fisher 2022). The stakes are simply too high.

8.2 Some Reflections on Optimism

Before wrapping up, we would like to address an objection that we have encountered on many occasions. After we explain the disastrous effects of mobile phones and the attention economy in places like Myanmar and Sri Lanka, defenders of these technologies sometimes reply by arguing that social media can also be a tool for mobilizing social movements and bringing about positive change. They point to examples like the Arab Spring to argue that social media and mobile phones could be tools of liberation rather than oppression.

Our response to this kind of social media optimism is twofold. First, as we have explained many times throughout the book, we are not opposed to using technology in ways that are conducive to our individual and collective ends. It is perfectly consistent with our position to claim that mobile devices can facilitate our ability to communicate and organize. But, on the other hand, we are deeply skeptical about the prospects of relying on social media and the attention economy when it comes to bringing about positive social change.

For starters, even when social media enables people to organize quickly (getting people to show up for a spontaneous protest within 24 hours, for instance), it has proven to be an ineffective tool for organizing robust social movements.[2] The efficacy of protests and social movements is going

[2] See Tufekci (2018), Chenoweth (2020), and Fisher (2022). Chenoweth writes, "Recent movements have increasingly relied on digital organizing, via social media in particular. This creates both strengths and liabilities. On the one hand, digital organizing makes today's movements very good at assembling participants en masse on short notice. It allows people to communicate their grievances broadly, across audiences of thousands or even millions. It gives organizers outlets for mass communication that are not controlled by mainstream institutions or governments. But the resulting movements are less equipped to channel their numbers into effective organizations that can plan, negotiate, establish shared goals, build on past victories, and sustain their ability to disrupt a regime" (2020, 78). Max Fisher writes, "Even seven-plus years on from the Arab Spring and whatever lessons it might've held for online activists, the new social media democracy produced a lot of chaos but strangely few results" (2022, 250).

down just as the quantity of such events goes up.[3] Social media optimists who point to the Arab Spring fail to appreciate the work of Zeynep Tufekci, Erica Chenoweth, and other social scientists who have shown that social media "really advantages repression in the digital age much more than mobilization" (Fisher 2022, 251).[4] First, social media quickly becomes a tool that is used by repressive regimes to crush these movements.[5] It makes it much easier for these movements to be monitored, infiltrated, and disrupted. Second, there are reasons to think that the reliance on social media for organizing is making movements less effective precisely because it has replaced forms of organizing that were more effective at building resilient communities.[6]

If we embrace techno-optimism and put our faith in social media activism, then we do so at our own peril. As we have argued throughout the book, the companies who built these platforms are simply in the business of capturing as much of our attention as possible. This means that their ends are not always aligned with our own. The attention economy is very good at generating ad revenue, but we have seen that it is markedly worse at bringing about positive social change.

But this is not the only reaction that we have encountered when presenting this work. On the flip side of techno-optimism, we are also frequently confronted by pessimism about the prospects for improving the situation. Many people recognize that the attention economy poses a problem, but they are skeptical about possible solutions. We are sympathetic to their concerns. Disengaging from the attention economy and mobile devices is easier said than done. And in the case of young people, damage was done to their agency before they had any say in the matter.

[3] As Chenoweth (2016) puts it, "And in fact, nonviolent resistance has actually become less successful compared to earlier, pre-internet times. Whereas nearly 70 percent of civil resistance campaigns succeeded during the 1990s, only 30 percent have succeeded since 2010." Chenoweth goes on to explain that one reason for this is that "as political scientist Anita Gohdes has carefully documented, governments are simply better at manipulating social media than activists" (Ibid.).

[4] Although this reference cites Fisher's book, it comes from his interview with Erica Chenoweth.

[5] See Tufekci (2018), Chenoweth (2020) and (2016), Fisher (2022), and Guesmi (2021).

[6] As Fisher explains, "Before social media, activists had to mobilize through community outreach and organization-building. They met almost daily to drill, strategize, and confer. It was agonizing, years-long work. But it made the movement durable, built on real-world ties and chains of command … Without the underlying infrastructure, social media movements are less able to organize coherent demands, coordinate, or act strategically" (2022, 251).

Nevertheless, we should not abandon hope. In the *Critique of Pure Reason*, Kant sums up reason's interests in three questions: "What can I know? What should I do? What may I hope?" (A805/B833). The question about hope implicates both practical and theoretical reason. It is at the intersection of these faculties that Kant raises questions about happiness and our worthiness of it. As Andrew Chignell (2022) points out, the third question has been unjustly neglected.

Kant's views on hope show us how our attitude about our moral condition can shape our ability to do what is right. Without hope, we might ignore our duty and give in to temptation, following whatever path we believe will lead us to happiness. Although Kant's ethics grounds rightness in the maxims of our actions rather than their consequences, Kant was not insensitive to important facts about human psychology. Kant thought we might be discouraged to see wicked people prosper and to watch as the efforts of virtuous people come to naught and fail to make changes in the world. Because of this, Kant argues that we are morally justified in believing things that theoretical reason cannot possibly prove.[7]

Chignell (2020) extends Kant's thinking about practical belief to argue that we have moral-psychological reasons for believing that change is possible and that our actions might play a part in building a better world. He writes, "[I]f we morally ought to act a certain way, and we are threatened by resolve-sapping despair, then we are prima facie morally justified in seeking strategies that will sustain our hope and thus our commitment to the ought in question" (237). Luckily, we do not have to pin our hopes on our actions as individuals. We should also see the potential for improvement in future generations as we collectively improve people's understanding of this problem and ways to respond to it.

[7] In the case of the highest good, Kant is particularly concerned with God and the immortality of the soul. Without these two postulates, Kant thinks it would be impossible for complete happiness to be meted out in proportion to perfect virtue. See KpV 5:107ff. Kant also (rather confusingly) includes freedom as a postulate of the highest good even though he claims earlier in the book that the reality of freedom (unlike the reality of God and the immortal soul) is proven by our awareness of the moral law. See KpV 5:4.

This will require us to commit heavily to the aim of education.[8] The first step to responding to this issue is simply recognizing it as a problem. We must see what is at stake and why it matters morally. In some cases (particularly the social/political issues), people are beginning to wake up to the gravity of the issue. The conversation about the effect of the attention economy on democracy and social cohesion is becoming harder to avoid. The second step requires us to make progress with our educational efforts. This is work that must be done collectively.

Kant has been unfairly maligned as someone whose thinking is excessively individualistic. Much of Kant's work, especially his writings about anthropology, history, and education, emphasize the importance of the social conditions in which human beings are raised. Indeed, Kant hoped that all of human history was headed in a direction of moral progress—one that would culminate in a rightful civil condition and moral culture:

> [T]here was always left over a germ of enlightenment that developed further through each revolution and this prepared for a following stage of improvement … [T]here will be opened a consoling prospect into the future (which without a plan of nature one cannot hope for with any ground), in which the human species is represented in the remote distance as finally working itself upward toward the condition in which all germs nature has placed in it can be fully developed and its vocation here on earth can be fulfilled. (Idea 8:30)

By making strides in education, legislation, and social relations, we may improve the attentional ecosystem to such an extent that we will begin to engage with technology in ways that enhance our autonomy rather than undermine it. In the meantime, as we work toward building that world, we can improve our own lives and the lives of those around us by taking small steps. We can spend less time on our devices and more time on activities that fill our lives with meaning. We can break free of the chains

[8] Kant was very clear about the role of education in the perfection of humanity. In his lectures on ethics, there is a section titled "Of the final destiny of the human race." Kant claims that "The final destiny of the human race is moral perfection, so far as it is accomplished through human freedom" (Collins 27:470). He then explains that we cannot expect to receive this perfection as a gift from God. Instead, we were given free will and we are thus presented with the task of morally perfecting ourselves. Kant is similarly skeptical about the prospect of relying solely on the state. So he concludes that education is the only viable path: "How, then, are we to seek this perfection, and from whence is it to be hoped for? From nowhere else but education" (Ibid.).

we have chosen to put in our pockets and gaze at for hours on end. Instead of being bound by a dopamine-driven cycle of endless media content, we may choose to bind ourselves to the only true product of our rational will: the moral law.

References

Chenoweth, Erica. 2016. How social media helps dictators. *Foreign Policy*. https://foreignpolicy.com/2016/11/16/how-social-media-helps-dictators/. Accessed 18 July 2023.

———. 2020. The future of nonviolent resistance. *Journal of Democracy* 31: 69–84.

Chignell, Andrew. 2020. Hope and despair at the Kantian chicken factory: Moral arguments about making a difference. In *Kant and animals*, ed. John J. Callanan and Lucy Allais. Oxford: Oxford University Press.

———. 2022. Hopeful pessimism: The Kantian mind at the end of all things. In *Faith, hope, and love: The theological virtues and their opposites*, ed. Troy DuJardin and M. David Eckel, 125–144. Cham: Springer International Publishing.

Eyal, Nir. 2014. *Hooked: How to build habit-forming products*. New York: Penguin.

Fisher, Max. 2022. *The chaos machine: The inside story of how social media rewired our minds and our world*. New York: Little, Brown and Company.

Guesmi, Haythem. 2021. The social media myth about the Arab Spring. *Al Jazeera*. https://www.aljazeera.com/opinions/2021/1/27/the-social-media-myth-about-the-arab-spring. Accessed 18 July 2023.

Harris, Tristan. 2021. Big tech's attention economy can be reformed. Here's how. *MIT* Technology Review. https://www.technologyreview.com/2021/01/10/1015934/facebook-twitter-youtube-big-tech-attention-economy-reform/. Accessed 14 July 2023.

Hay, Carol. 2011. The obligation to resist oppression. *Journal of Social Philosophy* 42: 21–45.

Hill, Thomas. 1973. Servility and self-respect. *The Monist* 57: 87–104.

Kant, Immanuel. 1998. *Critique of pure reason*. Cambridge: Cambridge University Press.

———. 2007. Idea for a universal history with a cosmopolitan aim. In *Anthropology, history, and education*, ed. Günter Zöller and Robert B. Louden. Cambridge University Press.

———. 2008. Critique of practical reason. In *Practical philosophy*. Trans. Mary J. Gregor. New York: Cambridge University Press.

Tufekci, Zeynep. 2018. Twitter and tear gas: The power and fragility of networked protest. Reprint ed. Yale University Press.

Index[1]

[1] Note: Page numbers followed by 'n' refer to notes.

T. Aylsworth, C. Castro, *Kantian Ethics and the Attention Economy*,
https://doi.org/10.1007/978-3-031-45638-1