THE OXFORD HANDBOOK OF

DIGITAL ETHICS

Edited by

CARISSA VÉLIZ

OXFORD
UNIVERSITY PRESS

CHAPTER 15

ETHICAL DIMENSIONS OF PERSUASIVE TECHNOLOGY

JAMES WILLIAMS

Introduction

Persuasion is as old as humanity. It is one of the most prevalent and consequential forms of interaction. Its pathways of influence span the conscious and unconscious, rational and nonrational, verbal and nonverbal. Its nature and ethics have been debated since antiquity. However, as networked digital technologies have rapidly transformed our capacities for managing information and attention, they have amplified and extended the means of persuasion in unprecedented ways. These technological advancements have occurred in tandem with advances in the understanding of human psychology and decision making, knowledge of which has been increasingly brought to bear on their design and deployment. The proliferation and power of these new forms of technological persuasion have brought fresh attention to age-old ethical questions about persuasion and raised other ones which are newly salient.

The ethics of persuasive technology intersects with many other areas of applied ethics, yet it also raises distinctive ethical considerations in its own right. Issues of persuasion may coexist with ethical issues that stem from the nature of a technology, for example, its reliance on automation, machine learning, or robotics. Similarly, the ethics of persuasive technology may overlap with ethical issues that emerge from the particular application domain of a technology's usage, such as in health, education, or advertising. Yet, persuasive technologies also present unique ethical considerations of their own which arise from the persuasion *per se*—considerations that broadly pertain to issues of autonomy. While the ethics of persuasive technologies remains an emerging area of study, scholarly attention to the issues it presents has advanced briskly in recent years. Research in the area has broadly been concerned with clarifying the ethical issues persuasive technologies raise and bringing ethical thinking to bear on their design.

In this chapter, I provide an overview of the high-level issues in this area by addressing three questions. First, what counts as a 'persuasive' technology? Second, what are the main ethical issues raised by persuasive technologies? Third, what major questions loom on the horizon for the ethics of persuasive technology?

What is a 'persuasive' technology?

A toy tiger encourages a two-year-old to exercise. A mobile app hooks a teenager into loops of constant status-checking. An advertising platform's algorithmic logic nudges a father to buy a new lawnmower (for more on digital nudging, see Ienca and Vayena, this volume). A home-care robot reminds an elderly woman to take her evening medicine. A self-driving taxi offers its occupant a discount for stopping at a nearby fast-food restaurant to get a milkshake.

There is a particular class of digital technology, only recently emergent in our lives, that as yet has no consensus term. It overlaps with but is not coextensive with design that has been described in terms of 'nudging', 'gamification', 'behaviour change', 'dark patterns', 'smart' technology, and many other descriptors. It has also become common to describe these technologies as 'persuasive' technologies. While this term is not wholly satisfactory, it is sufficient to enable a high-level ethical discussion of this domain (and is much preferable to hastily concocting another ill-considered neologism).

What does it mean to describe a technology as 'persuasive'? At its most basic level, the label is a metaphor drawn from the domain of interpersonal communication which emphasizes similarities between a technology's operation and human capacities of influence, much as the term 'artificial intelligence' does with capacities of pattern recognition, logic, attention, and other aspects of human intelligence. If we adapt an early definition of artificial intelligence given by McCarthy and colleagues in 1955 (McCarthy et al. 2006), we might say that a persuasive technology is one which is 'behaving in ways that would be considered' persuasive 'if a human were so behaving'. While this is a useful starting point, when we turn to specific cases, the lines of comparison between technological persuasion and interpersonal persuasion quickly become hard to draw. And, of course, this anthropomorphic definition does not clarify the central definitional question of what counts as persuasion in the first place.

There is a wide sense in which any technology can be described as 'persuasive' by virtue of its mere influence. That is not the kind of persuasiveness I am interested in here. That common form of influence relates to questions of a technology's 'affordances', 'biases', or 'intentionality' and is an important subject of analysis in its own right—one that has widely been recognized as morally salient (Winner 1980; Borgmann 1987; Latour 1992, Floridi and Sanders 2004; Verbeek 2011). A slightly narrower sense of 'persuasive', though one that is still too wide for our purposes here, involves technologies that are designed *intentionally* to influence. Similarly, to be a 'persuasive' technology, it is not sufficient that a technology merely *enable* persuasion, for example, a forum or email

client that allows users to send persuasive messages to one another. In other words, a persuasive technology is persuasive because of its form, not because of its content.

Here it is particularly important to distinguish between persuasive technology as a category of technology and Persuasive Technology as a field of study. In the early years of the twenty-first century, B.J. Fogg's (Fogg 2002; Fogg et al. 2007) work in applying novel insights from behavioural psychology to the design and analysis of personal computing technologies catalysed an interdisciplinary subfield known as Persuasive Technology. (Hereafter, to avoid confusion I will capitalize the term Persuasive Technology when referring to the field and use lower case when referring to the type of technology.) In the years since, the field has proved useful for advancing research into persuasive technologies across application domains and disciplinary boundaries as well as bringing attention to the mechanisms of influence common to these systems (Fogg 2002; Oinas-Kukkonen and Harjumaa 2008).

Within the field of Persuasive Technology, the term 'persuasive technology' has broadly been defined by the presence of three necessary conditions: there must be (a) an intention on the part of its designers to (b) change, reinforce, or otherwise shape human behaviour, attitudes, or both, in (c) a manner that is not coercive. Condition (a) seems overly restrictive of the agents behind a technology but is salvageable if we interpret 'designers' to mean not the narrow set of agents, such as programmers and user-interface designers, who design the technology in a hands-on manner but rather interpret the term more broadly to mean any agent with a substantial influence on the resulting design. This could include leaders of an organization, those who set higher-level goals (such as business models or other organizational priorities), or similar agents who exercise systemic influence rather than direct influence on the design. Condition (b) is, for our purposes, broadly unobjectionable. The third condition, however—the requirement that a persuasive technology be noncoercive—presents three problems. First, it amounts to a negative definition of 'persuasiveness', framing it not in terms of what it is but in terms of what it is not. Second, the nature of coercion remains a subject of considerable debate, especially when (a) the force in question is nonphysical in nature and (b) the psychological avenues of influence are nonrational in nature, both of which are often the case in the context of digital technology design. Finally, the requirement that persuasive technology not coerce—regardless of how one defines coercion—is ultimately a normative condition masquerading as a descriptive one. As far as I am aware, there is no separate category of 'coercive technology' under analysis anywhere, and if a technology does coerce its user, it arguably merits more critical and ethical attention and therefore greater inclusion in the set of technologies under consideration, not less. Ultimately, the noncoercion condition is an artefact of the dual nature of the field of Persuasive Technology, which is concerned with not only enabling analysis of technologies but also informing their design. The exhortation to not coerce is an admirable normative heuristic for avoiding many types of morally objectionable design, but, as a constraint on definition and analytical scope, it ought to be jettisoned.

There is one more important aspect of persuasive technology that has gone unaddressed by definitions to date: the system's assessment of success at producing the desired

effect, that is, the measurement of the outcome and its comparison with the system's persuasive goal. Importantly, to achieve effective assessment of persuasive success the outcome must be measured at an equal or greater degree of specificity as the representation of the persuasive goal. For example, if you are using a mobile app designed to persuade you to exercise every day, but it only measures whether you have opened the app at any time in the past month, then it would not count as a 'persuasive' technology.

We can therefore define a persuasive technology as: any sociotechnical system via which an intentional agent or agents intervenes, in a manner substantially involving digital technology, to effect a particular transformation in the experience, thought, or behaviour of a subject or subjects and measures the outcome of that intervention at a level of specificity sufficient to determine whether the transformation has occurred.

While not a paragon of simplicity, this definition seems conceptually sound. At any rate, it is sufficient; it is unnecessary to seek laser-like precision in our definition of persuasive technology in order to proceed with a discussion of its ethical dimensions.

What ethical issues are raised by persuasive technologies?

Because persuasion is ubiquitous in human life, persuasive technologies may intersect with many ethical considerations. Here I am interested in those ethical issues that arise specifically from the enhanced persuasiveness *per se* rather than secondary aspects which support or otherwise accompany increased persuasiveness. These secondary aspects may include, among other things, the technology's digital nature, considerations specific to a particular application domain, the role of user attributes or behaviour, or the political economy of the technology's influence in society. While many of these secondary aspects may have great ethical importance, the issues they present—which include questions of privacy, consent, or fairness, to name but a few—are already subjects of focused critical discussion in this volume and elsewhere. Furthermore, these secondary issues are so numerous that if we were to undertake an inventory and analysis of them it would require a book of its own.

Within the field of Persuasive Technology (PT), ethics remains a relatively understudied aspect of the literature (Oinas-Kukkonen 2010; Kight and Gram-Hansen 2019). Much attention to PT ethics has occurred in the context of design. Berdichevsky and Neuenschwander (1999) suggested a set of eight ethical principles for PT design using a rule-based utilitarian framework that culminated in the 'Golden Rule of Persuasion', which suggests that 'the creators of a persuasive technology should never seek to persuade a person or persons of something they themselves would not consent to be persuaded to do'. Spahn (2011) theorized about the application of discourse ethics, and Kaptein and Eckles (2010) elaborated the use of reflective equilibrium in a PT context. Oinas-Kukkonen (2010) argued that key ethics research needs

in PT included better understanding of the role of user consent, one's awareness of the persuasive process, and the role of tasks or goals in the design of persuasion as well as the importance of factors such as culture or gender in designing persuasion for user autonomy.

Broadly speaking, the ethical considerations that arise from a technology's persuasiveness per se concern questions of autonomy. Autonomy is a central component in many philosophical conceptions of well-being. It may be conceived in terms of reasons (Kant), an individual's motivational hierarchy (e.g. Frankfurt, Dworkin), or one's 'self-authorship' (Raz 1986), among other conceptions. In the context of persuasive technologies, autonomy has been identified as a particularly important ethical consideration (Smids 2012; Spahn 2011; Oinas-Kukkonen 2010; Verbeek 2009). In addition, special autonomy considerations arise in the context of use by children as well as vulnerable people (Jacobs 2020).

It may be useful to consider some of the key ethical issues raised by persuasive technologies in terms of two broad themes that are important for most conceptions of autonomy: (a) freedom of thought and attention and (b) freedom of action. These two categories are not mutually exclusive—in practice, they may and often do co-occur—but rather are two angles of analysis on the kinds of undue claims persuasive technologies may be seen to make on a person's autonomy.

Freedom of thought and attention

Freedom of thought is widely understood as necessary for autonomy and is enumerated as a fundamental freedom in Article 18 of the Universal Declaration of Human Rights (United Nations General Assembly 1948). In *On Liberty*, Mill (1989 [1859]) writes that the 'appropriate region of human liberty … comprises, first, the inward domain of consciousness', which includes 'liberty of thought and feeling; absolute freedom of opinion and sentiment on all subjects, practical or speculative'. Importantly, these freedoms pertain to not only the content but also the manner of one's inner life: they include not merely the freedom to determine one's beliefs, desires, or volitions, but also to direct the workings of one's thought as they see fit. In particular, attention has been described as the basis of conscious thought as well as will (James 1890). The 'attention economy' (Lanham 2006) that now exists to capture and monetize human attention underlies the creation of many persuasive technologies. Many dominant platforms today optimize their services to maximize users' time and 'engagement' with the system in order to sell access to their attention to other parties (e.g. in the form of advertising) as their primary business model. What is ultimately being sold in these cases is the chance to shape the minds and wills of users.

Four key themes that relate to persuasive technologies and freedom of thought are: distraction, reflection, capacity depletion, and nonrational avenues of influence.

Distraction, or the direction of a person's attention away from task- or goal-salient information, is a common effect of persuasive technology. It may occur as an

interruption, for instance, from an automated notification on a mobile device, or as a message from another person who was prompted to send it in order to increase their engagement with the service. The effect of associating a distractor with a reward has attentional effects that last well beyond that specific situation (Itthipuripat et al. 2019). Exposure to repeated notifications can even create mental habits that train users to interrupt themselves later when the technological influences are not present (Mark et al. 2008).

When a distractor directs a person's attention away from effective consideration and processing of internal information, it may undermine their reflection. Reflection is an important component of autonomy, for example, in the formation of second-order desires and volitions (Frankfurt 1988). A persuasive technology may deploy external stimuli (i.e. perceptual information) that compete with reflective information, thereby crowding out opportunities for reflection.

A persuasive technology may drain a person's cognitive capacity, such as working memory or willpower, as well. For example, in *The Morality of Freedom*, Raz (1986) gives the example of 'The Hounded Woman' to describe how capacity depletion threatens autonomy:

> A person finds herself on a small desert island. She shares the island with a fierce carnivorous animal which perpetually hunts for her. Her mental stamina, her intellectual ingenuity, her will power and her physical resources are taxed to their limits by her struggle to remain alive. She never has a chance to do or even to think of anything other than how to escape from the beast.

The Hounded Woman lacks personal autonomy because she lacks 'an adequate range of options to choose from'. Importantly, the 'lacking' here does not mean the absence of a range of options existing in the world but rather the inability to give attention to them. This case is analogous to the way a persuasive technology might relentlessly compete for or otherwise pursue a user's attention in its desire to optimize their engagement with it. (Of course, Raz's example is far more extreme in that the woman's life is in immediate danger, but the principle at work is the same.) Influencing via capacity depletion relates to the wider issue of using nonrational avenues of influence generally, for example, the exploitation of psychological biases or vulnerabilities (Smids 2012; Spahn 2011; Oinas-Kukkonen 2010; Verbeek 2009). Such biases include loss aversion (e.g. 'fear of missing out'), social comparison, the status quo bias, framing effects, appeals to consistency or scarcity, and many others (Kahneman 2011; Fogg et al. 2007). One particular group of biases widely exploited by persuasive technologies involves the psychology of reward and surprise. When a technology seeks to maximize a user's engagement with it, it can often be extremely effective to randomize the delivery of some type of reward that the user experiences. For instance, in a video game, this could take the form of randomizing the value and number of items that the player finds in loot boxes (in-game containers that the player's character opens) or randomizing posts in the infinite scrolling feed on a social media site. These mechanisms of reward and surprise are, of course, the same

ones that underlie the compulsive nature of gambling, machine gambling in particular (Schüll 2012).

Freedom of action

Technologies designed to change people's behaviour may raise a wide range of issues related to freedom of action. Four key considerations I wish to address here are: coercion, manipulation, deception, and habituation.

As noted earlier, coercion is often explicitly excluded in definitions of persuasive technology. This exclusion is praiseworthy insofar as it is intended as a normative nudge for practitioners, but its presence ultimately hinders clear philosophical and ethical analysis of persuasive technologies. This is particularly so given the uncertainty regarding whether, and if so how, the concept of coercion can be meaningfully applied to influence that is nonrational in nature. For example, Smids (2012) argues that a car's persistent, incessant beeping aimed at causing the driver to fasten their seatbelt constitutes a type of coercion in the way it wears down their capacity for self-control. This example is particularly useful not only because it reminds us that in some cases coercion may be morally obligatory, but also because it emphasizes that, even if coercive mechanisms are apparent to the user, they may still influence via indirect means.

Like coercion, manipulation is a category of influence widely seen to pose threats to autonomy not only in the context of persuasive technologies but also in wider discussions about 'big data' (Schroeder and Cowls 2014), decision architecture (Sunstein 2015), and human interaction in general (Kane, 1996; Pereboom, 2001). While there is broad disagreement about what the criteria for manipulation are, the definition given by Pennock (1972) captures its most common elements: it is influencing a person 'by controlling the content and supply of information' in such a way that the person is not aware of the manipulation (Powers 2007; Strauss 1991; Kane 1996; Pereboom 2001; Smids 2012). Blumenthal-Barby (2014) further describes it as a form of 'nonargumentative influence', and Susser et al. (2019) describe it as intentionally hidden influence.

In persuasive technologies, design often has a special incentive to deceive the user. Such deceptive design is often described in terms of 'dark patterns', that is, design patterns that serve to trick or mislead a user in the direction of some desired outcome. Especially prevalent on e-commerce websites, dark patterns occur in a wide range of persuasive technologies (Mathur et al. 2019; Gray et al. 2018). In discussion of recommendation algorithms, Seaver (2019) usefully suggests the metaphor of 'traps' as a framework for understanding the design. Personification of artificial agents also presents a wide range of new avenues for the deployment of dark patterns, as has been discussed with, for example, the 'cuteness' of the design of some robots (Lacey and Caudwell 2019).

Habituation is one of the most common goals of persuasive design. One of the most prevalent methods for creating user habits is via strategic use of reward mechanisms. Randomizing the reward schedule for a given action increases the number of times

a person is likely to take that action (Ferster and Skinner 1957). Eyal (2013) identifies variable rewards as a crucial step in the process of creating habits and 'hooking' users on a product. This is an underlying mechanism at work behind 'infinite' scrolling feeds of information such as those employed by Twitter, Facebook, Pinterest, and countless other news, social, and entertainment websites. It is often referred to as the 'slot machine' effect because it is the foundational psychological mechanism on which the machine gambling industry relies and which generates for them over $1 billion in revenue every day in the United States alone (Rivlin 2007). The compulsive and at times even addictive effects that these designs have on users can often be severely debilitating (Schüll 2012).

What are the major questions for the future of persuasive technology ethics?

As far as autonomy is concerned, two main categories of concern loom large on the horizon of persuasive technology ethics: (1) how to mitigate the risks that persuasive technologies pose to autonomy and (2) how to use persuasive technology as a way of enhancing autonomy. Within each category, three broad layers of questions exist that pertain to (a) the design and deployment of persuasive technologies, (b) the incentives and other upstream determinants that influence their design, and (c) linguistic, conceptual, and other cultural dimensions that shape our understanding of persuasive technologies.

When it comes to mitigating potential injuries to users' freedom of thought and action, at the level of design, there are many significant questions pertaining to the ethical acceptability of appealing to nonrational and/or nonconscious avenues of persuasion. For instance, are there certain vulnerable parts of our cognition that ought to receive special protection from persuasive attempts in the same way that certain vulnerable parts of our bodies may require special protection from physical harms? At the level of upstream determinants of design, there are major questions to be addressed involving the reformation of business models in a way that respects users' attention and will. And at the level of linguistic and conceptual dimensions that shape how we understand persuasive technology, a major project looms in clarifying and advancing the vocabulary for the myriad varieties of influence, especially as they pertain to nonrational avenues of persuasion. For example, where do the boundaries lie between different forms of influence such as manipulation, suggestion, coercion, exhortation, etc.? In particular, how can these boundaries be better clarified with reference to the morally salient factors that distinguish them?

Ultimately, the aim for persuasive technology ethics should be to transcend questions of harm minimization and pursue the positive enhancement of people's autonomy. At

the level of design, this may involve advancing the technological awareness of and sensitivity to users' tasks and goals. It could also entail the use of counter-technologies to mitigate distractions, for example, whether they arise from other technologies or some other external source. Examples may include, for instance, a web browser plug-in that blocks ads or a word processor with a 'focus mode' that blocks out interruptions from other distracting programmes on the user's computer. At the level of upstream determinants of design, key questions include how business models can be created that make it valuable for design to enhance users' autonomy rather than simply not crossing a line in undermining it. Finally, at the level of how we think and talk about persuasive technologies, one could imagine the proliferation of validated metaphors of beneficent influence such as a 'coach', 'guide', 'buddy', 'partner' and so on. I deliberately use the qualification 'validated' here because it would seem particularly important for these metaphors of persuasive nuance not to themselves become tools of undue persuasion (e.g. mere corporate marketing tools or the basis for 'greenwashing' efforts) by merely being able to say, rather than having to verify, that they are in fact on the side of the user. To mitigate such deceptive use, one could imagine the development of a vocabulary and set of expectations for such persuasive metaphors which would enable users to explicitly and confidently delegate aspects of their autonomy to persuasive systems.

Looking into the future, there will be no shortage of substantive ethical questions about persuasive technologies in need of attention any time soon. As these technologies become ever more ubiquitous and indispensable in our lives, attention to their ethical dimensions—which extend into numerous areas well beyond the ones I have briefly surveyed here—will become all the more essential for promoting the well-being of users, designers, and society as a whole.

Conclusion

The ethics of persuasive technology is an emerging field with unique significance and urgency across a wide range of domains. Success in this area will depend on not only advancing the ethical analysis but also clarifying the philosophical landscape that underlies and enables it. That philosophical landscape includes questions of the varieties of forms of influence, in particular, the varieties of nonrational pathways of influence.

Since time immemorial, a great deal of the success of our individual and collective lives has depended on the success of persuasive engagements. In our time, digital technologies have dramatically extended our capacities of persuasion in ways that stand to amplify the harms as well as benefits, depending on the nature of the design. The ethics of persuasive technology, though a young area of work, is poised to inform these technologies' design if, moving forward, it can receive attention that is commensurate with its significance. Informing design in this way will be essential if we are to realize the promise of these technologies, which, if conceived correctly, we can use to persuade ourselves in better ways, and towards better ends.

References

Berdichevsky, Daniel, and Neuenschwander, Erik (1999), 'Toward an Ethics of Persuasive Technology', *Communications of the ACM* 42(5), 51–58.

Blumenthal-Barby, Jennifer S. (2014), 'A Framework for Assessing the Moral Status of "Manipulation"', in Christian Coons and Michael Weber, eds, *Manipulation: Theory and Practice* (Oxford: Oxford University Press), 121–134.

Borgmann, Albert (1987), *Technology and the Character of Contemporary Life: A Philosophical Inquiry* (Chicago, IL: University of Chicago Press).

Eyal, Nir (2013), *Hooked: A Guide to Building Habit-Forming Products* (CreateSpace).

Ferster, Charles B., and Skinner, Burrhus F. (1957), *Schedules of Reinforcement* (New York, NY: Appleton-Century-Crofts).

Floridi, Luciano, and Sanders, John W. (2004), 'On the Morality of Artificial Agents', *Minds and Machines* 14(3), 349–379.

Fogg, Brian J. (2002), 'Persuasive Technology: Using Computers to Change What We Think and Do', *Ubiquity* December, 5.

Fogg, Brian J., Cuellar, Gregory, and Danielson, David (2007), 'Motivating, Influencing, and Persuading Users', in Andrew Sears and Julie A. Jacko, eds, *The Human–Computer Interaction Handbook* (CRC press), 159–172.

Frankfurt, Harry G. (1988), *The Importance of What We Care About: Philosophical Essays* (Cambridge: Cambridge University Press).

Gray, Colin M., Kou, Yubo, Battles, Bryan, Hoggatt, Joseph, and Toombs, Austin L. (2018), 'The Dark (Patterns) Side of UX Design', *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems April*, 1–14.

Itthipuripat, Sirawaj, Vo, Vy A., Sprague, Thomas C., and Serences, John T. (2019), 'Value-Driven Attentional Capture Enhances Distractor Representations in Early Visual Cortex', *PLoS Biology* 17(8), e3000186.

Jacobs, Naomi (2020), 'Two Ethical Concerns about the Use of Persuasive Technology for Vulnerable People', *Bioethics* 34(5), 519–526.

James, William (1890), *The Principles of Psychology* (New York: Henry Holt and Company).

Kahneman, D. (2011), *Thinking, Fast and Slow* (New York: Farrar, Straus and Giroux).

Kane, Robert (1996), *The Significance of Free Will* (New York: Oxford University Press).

Kaptein, Maurits, and Eckles, Dean (2010), 'Selecting Effective Means to Any End: Futures and Ethics of Persuasion Profiling', in *International Conference on Persuasive Technology* (Berlin, Heidelberg: Springer), 82–93.

Kight, Raymond, and Gram-Hansen, Sandra B. (2019), 'Do Ethics Matter in Persuasive Technology?', in Harri Oinas-Kukkonen, Khin Than Win, Evangelos Karapanos, Pasi Karppinen, and Eleni Kyza, eds, *International Conference on Persuasive Technology* (Cham: Springer), 143–155.

Lacey, Cherie, and Caudwell, Catherine (2019), 'Cuteness as a "Dark Pattern" in Home Robots', in *2019 14th ACM/IEEE International Conference on Human–Robot Interaction (HRI)* (Institute of Electrical and Electronics Engineers), 374–381.

Lanham, Richard A. (2006), *The Economics of Attention: Style and Substance in the Age of Information* (Chicago, IL: University of Chicago Press).

Latour, Bruno (1992), 'Where Are the Missing Masses? The Sociology of a Few Mundane Artifacts', in Wiebe E. Bijker and John Law, eds, *Shaping Technology/Building Society: Studies in Sociotechnical Change*, Vol. 1 (Cambridge, MA: MIT Press), 225–258.

Mark, Gloria, Gudith, Daniela, and Klocke, Ulrich (2008), 'The Cost of Interrupted Work: More Speed and Stress', in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (ACM), 107–110.

Mathur, Arunesh, Acar, Gunes, Friedman, Michael J., Lucherini, Elena, Mayer, Jonathan, Chetty, Marshini, and Narayanan, Arvind (2019), 'Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites', *Proceedings of the ACM on Human–Computer Interaction* 3, 1–32.

McCarthy, John, Minsky, Marvin L., Rochester, Nathaniel, and Shannon, ClaudeE. (2006), 'A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955', *AI Magazine* 27(4), 12.

Mill, John S. (1989 [1859]), *JS Mill: 'On Liberty' and Other Writings* (Cambridge: Cambridge University Press).

Oinas-Kukkonen, Harri (2010), 'Behavior Change Support Systems: A Research Model and Agenda', in *International Conference on Persuasive Technology*, 4–14.

Oinas-Kukkonen, Harri, and Harjumaa, Marja (2008), 'A Systematic Framework for Designing and Evaluating Persuasive Systems', in *International Conference on Persuasive Technology*, 164–176.

Pennock, Joland R. (1972), 'Coercion: An Overview', in *Coercion* (New York: Routledge), 1–15.

Pereboom, Derk (2001), *Living without Free Will* (Cambridge: Cambridge University Press).

Powers, Penny (2007), 'Persuasion and Coercion: A Critical Review of Philosophical and Empirical Approaches', *HEC Forum*, 19(2), 125–143.

Raz, Joseph (1986), *The Morality of Freedom* (Oxford: Clarendon Press).

Rivlin, Gary (2007), 'Slot Machines for the Young and Active', *New York Times*, 15 March, http://www.nytimes.com/2007/12/10/business/10slots.html, accessed 6 September 2022.

Schroeder, Ralph, and Cowls, Josh (2014), 'Big Data, Ethics, and the Social Implications of Knowledge Production', in *Proceedings of KDD Data Ethics Workshop*, Vol. 24, 1–4.

Schüll, Natasha D. (2012), *Addiction by Design: Machine Gambling in Las Vegas* (Princeton, NJ: Princeton University Press).

Seaver, Nick (2019), 'Captivating Algorithms: Recommender Systems as Traps', *Journal of Material Culture* 24(4), 421–436.

Smids, Jilles (2012), 'The Voluntariness of Persuasive Technology', in *International Conference on Persuasive Technology*, *Persuasive Technology. Design for Health and Safety*, 123–132.

Spahn, Andreas (2011), 'And Lead Us (Not) into Persuasion . . . ? Persuasive Technology and the Ethics of Communication', *Science and Engineering Ethics* 18, 633–650.

Strauss, David A. (1991), 'Persuasion, Autonomy, and Freedom of Expression', *Columbia Law Review* 91(2), 334–371.

Sunstein, Cass R. (2015), 'Nudging and Choice Architecture: Ethical Considerations', Yale Journal on Regulation, Forthcoming, Harvard John M. Olin Discussion Paper Series Discussion Paper No. 809, January.

Susser, Daniel, Roessler, Beate, and Nissenbaum, Helen (2019), 'Technology, Autonomy, and Manipulation', *Internet Policy Review* 8(2), 22.

United Nations General Assembly (1948), 'Universal Declaration of Human Rights', *UN General Assembly* 302(2), 14–25.

Verbeek, Peter-Paul (2009), 'Ambient Intelligence and Persuasive Technology: The Blurring Boundaries between Human and Technology', *Nanoethics* 3(3), 231–242.

Verbeek, Peter-Paul (2011), *Moralizing Technology: Understanding and Designing the Morality of Things* (Chicago, IL: University of Chicago Press).

Winner, Langdon (1980), 'Do Artifacts Have Politics?', *Daedalus* 109(1), 121–136.