ETHICS OF CARE

Care in an Era of New Technologies and Artificial Intelligence

Relationships in a Connected World

Vanessa Nurock

VOLUME 14

Care in an Era of New Technologies

and Artificial Intelligence

Cover illustration:
Felix Nussbaum (1904-1944)
Paar in surrealer Landschaft [*Surreal Landscape with Couple*], around 1939
Oil on wood, 50,7 × 65,5 cm
Felix-Nussbaum-Haus at Museumsquartier Osnabrück, Loan of the Niedersächsische Sparkassenstiftung

Ethics of Care
Volume 14

Care in an Era of New Technologies and Artificial Intelligence

Relationships in a Connected World

Vanessa Nurock

PEETERS
Leuven – Paris – Bristol, CT
2025

Translated by Catherine Porter

Published with the support of the CRHI and the IDEX
of the Université Côte d'Azur

Publié avec le soutien du CRHI et de l'IDEX
de l'Université Côte d'Azur

A catalogue record for this book is available from the Library of Congress.

ISBN 978-90-429-5201-0
eISBN 978-90-429-5202-7
D/2025/0602/6

Table of Contents

Acknowledgments IX

Introduction: The Polethical Moment 1

1 *Does the present still need us (and Notre Dame)?* 1
2 *The Enlightenment in reverse?* 3
3 *From technological experimentation to social experimentation* 5
4 *A "Sputnik moment" and technological neocolonialism* 8
5 *A normative and polethical opening* 9
6 *Enlightenment with care* 12

Part One:
Nanoethical Stakes

Chapter 1: Nanoethics 25

1.1 *Did you say "nano"?* 25
1.1.1 A controversial scientific definition 26
1.1.2 A social and political question? 29
1.2 *Toward a political definition of nanos? Risk and its avatars* 34
1.2.1 Three approaches to "accompaniment" 36
1.2.2 Three "candidates" 38
1.2.2.1 Social acceptability 38
1.2.2.2 Responsible innovation 40
1.2.3 "Safety by design" 43
1.3 *What "nanoethics" means* 44
1.3.1 Is there such a thing as nanoethics? 45
1.3.2 What is a nanoethical problem? 47
1.3.2.1 First and second order nanoethical questions 48
1.3.2.2 Nanoethics as a crossroads of applied ethics 49
1.3.3 The nanoethical problem 51

1.4 *Nanocare* 52
1.4.1 Care as a challenge to the traditional boundaries 54
1.4.2 The boundary between designers and objects 56
1.4.3 The boundary between the natural and the artificial 57
1.4.4 The boundary between the inside and the outside 60
1.4.5 Perspectives 61
1.5 *Nanoethics and relational responsibility* 62

Chapter 2: Cybergenetics 67

2.1 *The emergence and development of cybergenetics* 69
2.1.1 The *Human Genome Project* 69
2.1.2 The beginnings of cybergenetics: from the personal genome project to "recreational" genetics 74
2.2 *Cybergenetics and "recreational" genetics* 79
2.2.1 What is recreational cybergenetics? 79
2.2.2 Recreational genetics today 84
2.3 A *"gift," but for whom?* 89
2.3.1 The genetic DTC (super)market: direct to consumers = direct to companies 89
2.3.2 Cybergenetics and health data: the gift of (big) data 96
2.4 *Recreation and re-creation: the world according to "direct to consumer" cybergenetics* 103
2.4.1 Reshaping collective and individual identities: from self-decoding to self-narration 103
2.4.2 Recreascientific? 112
2.4.3 Curating data vs. caring about/for/with data: a "factishization" of cybergenetics? 116
2.4.4 From the reconfiguration of the concept of patient to the reconfiguration of the individual 124
2.4.5 Rethinking relationships rather than autonomy 124
2.4.6 A CyberGenoPanopticon? 128
2.4.7 For a subject in relationships 134

Part Two:
Artificial Intelligence and Ethical Experimentation

Chapter 3: Is an AI Ethics Possible? 141

3.1 *Beyond boundaries or toward new borders?* 141
3.1.1 Defining artificial intelligence: an ironic quest 142
3.1.2 Between "Gabor's law" and the Collingridge dilemma: how to ask the question of control (or not)? 149
3.1.3 State of the art of AI ethics 152
3.1.3.1 A new Hippocratic Oath? 153
3.1.3.2. An applied ethics for AI? 155
3.1.3.3 A "fair by design" AI 157
3.2 *Coding Morality?* 158

Chapter 4: Two Paradigmatic Case Studies of "Ethical" AI – Autonomous Machines and Relational AI 169

4.1 *Machines said to be "autonomous" and the automatization of ethics* 169
4.1.1 Why autonomous machines? Artificial intelligence and the smart city 169
4.1.2 "Ought" implies "can" 177
4.1.3 The blurring of the inside/outside dichotomy and the modification of the empathic relation 180
4.1.4 Patterns of life or forms of life? 192
4.1.5 Toward a mechanized moral code? 198
4.2 "*Relational*" *AI (machines of "care")* 210
4.2.1 Artificial love 211
4.2.2 Extra-ordinary assistants 214
4.2.3 AI for care? 219

Chapter 5: The Artificialist Fallacy 229

5.1 *Will it still be possible to conceptualize what is possible?* 229
5.1.2 Roomba-ization of human lives 229

5.1.2 An experiential and conceptual reconfiguration? 234
5.1.3 Shortfalls of dilemmas 236
5.2 *From the Apparatgeist to forms of life* 238
5.2.1 The Apparatgeist 238
5.2.2 From the Apparatgeist to forms of life 240
5.3 *Toward an artificialistic fallacy?* 242
5.3.1 The naturalistic fallacy 243
5.3.2 Naturalization according to Bourdieu 245
5.3.3 The artificialist fallacy 245

Conclusion: Back to the Enlightenment? 255

1 *Relations and patriarchy* 258
2 *Two central problems: new generations and deresponsibilization* 260
3 *A relational responsibility?* 261
4 *Toward a polethics of care for NBIC?* 264
5 *An ethics by design* 267

Bibliography 278

Acknowledgments

While writing a book always requires a certain form of solitude, exchanges, encounters, and discussions with others are what really make writing possible. The impetus for the present book undoubtedly dates back to the era of my doctoral studies, more than fifteen years ago, when the question of "naturalizing" morality had already struck me as inseparable from that of "artificializing" morality. Many of those early exchanges and encounters have been ongoing and have helped to shape the arguments developed here.

Sandra Laugier in particular was immensely helpful as I drafted the initial version of this work and prepared it for publication. She has my warmest thanks for her patience and her benevolent exigency.

I owe a particular debt to the late Ruwen Ogien, whose contribution to moral philosophy was primordial in itself, and who, with finesse and intelligence, humor and elegance, helped a whole generation of philosophers engage with topics that had found little acceptance within academic fields in France. Although I was able to discuss only a few aspects of the first chapter with him, I often asked myself, as I went on, what elements and arguments he would have challenged. His singular voice is sorely lacking in the French philosophical landscape today.

I want to thank Catherine and Raphaël Larrère for more than twenty years of wonderful exchanges, and for all that they have taught me.

I am also grateful to Marie-Hélène Parizeau for very fruitful and inspiring discussions on AI.

I have had the good fortune to be surrounded by exceptional individuals who work on the the thematics of care in the Francophone world. It would be difficult to cite them all, and I beg forgiveness from those I have omitted for want of space or memory. I am well aware that our teamwork has been nourished by a broad global network working on questions of care. In France, first of all, let me mention Sandra Laugier (again), Pascale Molinier, Patricia Paperman, Fabienne Brugère, Catherine Larrère (again). In Quebec, Sophie Bourgault, Sophie Cloutier, Stéphanie Gaudet. Many thanks, too, to Carol Gilligan, for having taken the time to discuss certain subtleties in her approach with me, and for our exchanges in the Luxembourg Garden some ten years ago, when I was starting to test the idea that the ethics of care could be particularly fruitful in the area of the new technologies, which were often considered more "solid," more "virile," even, than care for others.

Thanks also to my colleagues working on nano-related research projects with whom I was able to discuss the main ideas developed in the first chapter. In addition to those mentioned above, I want to thank Bernadette Bensaude-Vincent, Xavier Guchet, Sacha Loeve, Ronan Le Roux, Sophie Pellé, Nathalie Panissal, and Christophe Vieu, for our conversations about nanotechnologies and nanomedicine.

A first version of this book constituted the "unpublished manuscript" I presented in 2020 to earn the degree entitling me to supervise doctoral candidates (*habilitation à diriger des recherches*). Many thanks to the colleagues who agreed to read and discuss it and to serve on the jury: Ali Benmakhlouf, Fabienne Brugère, Pierre Cassou-Noguès, Marie Gaille, Emmanuel Hirsh, Marie-Hélène Parizeau, and Pierre-Yves Quiviger.

I very much appreciate my good fortune in being granted a UNESCO chair in the ethics of living beings and artificial entities, a position that allowed me to confront the limits of my thinking and to open it up to the world. Special thanks to the social and human sciences sector within UNESCO, and to Dafna Feinholz, Orio Ikebe, and Doaa Abu Elyounes for our discussions.

I have also been fortunate enough to be assigned to the Centre national de la recherche scientifique (CNRS) within a Franco-American laboratory (UMI EPIDAPO) that brings together biology and the social sciences, hosted by the University of California at Los Angeles in its Institute for Society and Genetics (ISG). My thanks to these institutions and to the many individuals who welcomed me into an exceptional environment. While I cannot name them all, I want to mention in particular my ISG colleagues, who have been very welcoming as well as David Panaggia and Joshua Foa Dienstag. In particular, David took the time to discuss with me some of his ideas about what he calls datapolitik; his influence on the present book is probably more significant than is apparent from the endnotes. Thanks also to Eric Vilain, director of EPIDAPO at the time, who authorized me to sit in on his genetics team's staff meetings, thus allowing me insight into the specificities of the American biomedical milieu. I was able to work during research seminars with a group of exceptional students who helped me get a clearer view of the landscape of cybergenetics; particular thanks go to Kate Anna Clendenen, Rushna Raza, Jenny Ding, Mackenzie Grace Casey, and Antoine Rajkovic. I also thank my colleagues from the ERAPERMERMED Miracle project for discussing on concrete ethics by design medical issues.

In Los Angeles, I also had the good fortune of meeting Carol Scott, from whose legal expertise I benefited. She accompanied me as I was discovering the surprising world of American cybergenetics through my French-tinted glasses; the chapter on cybergenetics owes her a great deal. Once back in Paris, I was able to attend a series of seminars directed by Lisa Ganett, who kindly took the time to clarify certain aspects of the subject for me.

Thanks also to my colleagues at Boston University, especially Juliet Floyd and James Katz, for our stimulating interactions and for their benevolence; and thanks to Jan Krzysztof Blusztajn and Julie Lynch for our discussions over the years, The Internet links in this book have been rechecked by Valentine Bailly, whom I thank here. Thanks also to my colleagues at the Université Cote d'Azur and especially the Centre de Recherche en Histoire des Idées (CRHI) for their support. I wish also to thank CREATES, CRHI, IDEX, Université Côte d'Azur for financial support with the translation and the Open Access publication. This work was supported by the French government through the France 2030 investment plan managed by the National Research Agency (ANR), as part of the Initiative of Excellence Université Côte d'Azur under reference number ANR-15-IDEX-01 and by ANR under the frame of Era PerMed (ANR-21-PERm-0001).

My family has had to put up more or less patiently with the material and psychological weight entailed in writing a "heavy" book of this sort; their remarks, reactions, and jokes have often helped me to see what I was doing more clearly.

The book was completed under very difficult conditions that delayed its publication. It was written essentially between 2018 and 2021, and I am well aware that new technologies have emerged since then. Still, it seems to me that the concepts proposed and the analyses developed are not outdated, and indeed that no recent work on the new technologies can be fully up to date, given the rapidly accelerating pace of innovation.

Last but not least, I wish to thank warmly Catherine Porter for providing much more than a mere 'translation' and helping me (try to) make this book more readable and, I hope, enjoyable. My thanks also go to two anonymous referees for helpful criticism and to the "Ethics of Care" team, with very special thanks to Sophie Bourgault for her follow-up with the editorial process.

This book is dedicated to my dearest cousin Tally Kritzman-Amir (1979-2022), who could not publish hers.

Earlier parts of this book have been published previously in the past few years (mainly in French) in the following publications:

Nurock, Vanessa. 2008. *Rawls, pour une démocratie juste?* Paris: Michalon.

Nurock, Vanessa. 2010. "Nanoethics: Ethics For, From, or With Nanotechnologies?" *Hyle* 16: 31-42.

Nurock, Vanessa, and Nathalie Panissal. 2016. "Teaching a 'Care' Approach to Nanotechnologies." In Bowman et al., 125-37.

Nurock, Vanessa, ed., 2019. *L'intelligence artificielle: Enjeux éthiques et politiques*. Coll. Cités. Paris: Presses Universitaires de France.

Nurock, Vanessa. 2019a. "Généalogie de la morale automatisée." In Marie-Hélène Parizeau and Soheil Kash, eds., *Robots et sociétés: Enjeux éthiques et politiques*, 31-50. Québec: Presses de l'Université Laval.

Nurock, Vanessa. 2019b. "L'intelligence artificielle a-t-elle un genre?" In Vanessa Nurock, ed., 2019, 61-74.

Nurock, Vanessa. 2019c. "Le *care* de la nanoéthique: Repenser la question des frontières." *Ethica* 22: 149-65.

Nurock, Vanessa. 2020a. "Puede prestar Cuidado la Intelligencia Artificial? *Cuadernos de Relaciones Laborales* 38:217-29.

Nurock, Vanessa. 2020b. "Across Boundaries : rethinking the Ethics of Nanotechnologies in light of Care". In Sophie Bourgault and Frans Vosman, eds., *Care Ethics in yet a Different Voice : Francophone Contributions*, Leuven, Peeters Publishers, 253-270.

Nurock, Vanessa. 2021. "The Artificialist Fallacy." In James E. Katz, Juliet Floyd, and Katie Schiepers, eds., *Perceiving the Future through New Communication Technologies: Robots, AI and Everyday Life*, 75-87. London: Palgrave Macmillan.

Nurock, Vanessa, Raja Chatila, and Marie-Hélène Parizeau. 2021. "What Does 'Ethical by Design' Mean?" In Bertrand Braunschweig and Malik Ghallab, eds., *Reflections on Artificial Intelligence for Humanity,* 171-90. Cham, CH: Springer.

Nurock, Vanessa. 2024. *Quelle Ethique pour les Nouvelles Technologies ?* Paris: Vrin.

INTRODUCTION

The Polethical Moment

1 Does the present still need us (and Notre Dame)?

Imagine a lovely day in Paris, where an American tourist is filming Notre Dame on his drone. Later, he posts the content on the Internet; somehow we come across it and discover that the drone has picked up part of a conversation between two people strolling along the banks of the Seine. They seem to be talking about the changing times. Intrigued, we decide to transcribe the conversation.[1] We'll call the speakers David and Jacqueline.

We hear David first: "In our day, it suffices to spit a bit of saliva into a test tube and send it off in the mail to a private company in order to find distant cousins whose existence we didn't even suspect, or to know whether we're at risk of developing a serious illness, or to find out what talents our children might be able to develop. It's incredible!"

Jacqueline agrees: "Yes, it's true, everything seems to be going faster today, and not just in biotechnology. When I think that we can move around in 'driverless' cars, that machines endowed with artificial intelligence can take care of our children or our grandparents, or run for office, as we've seen recently in Japan and Denmark, I sometimes have the impression that I'm living in a science fiction story, when these things are already real!"

[1] The transcription is as unethical as the recording and posting, perhaps, but remember – we're imagining!

David picks up the thread. "Facebook has finally replaced Rousseau's *Reveries of a Solitary Walker*, just as e-learning or robotic nursemaids have replaced his *Émile, or On Education*."

"Yes," Jacqueline goes on. "One could say that the skeptical philosophy that aimed to cast doubt on what we know or think we know by privileging an experimental approach no longer serves to test ideas, as was the case with thought experiments. We have just the opposite in fake news, for example; we're seeing forms of social experimentation in which technologies are put into practice without any preliminary investigation of their implications."

Continuing on their walk, the two pass in front of the Notre Dame cathedral. David muses aloud: "I'm thinking about an article I read maybe twenty years ago in *Wired* – do you remember that magazine? It was so important in the tech world that it was called the 'Rolling Stone of technology' at the time. The article title was "Why the Future Doesn't Need Us" (2000).[2] The author, Bill Joy, was a major figure in computer science with a pessimistic outlook. He claimed that developments in nanotechnologies, genetic engineering, and robotics were building a future in which there'd no longer be a place for humanity. And yet we're still here, in this present which was once the future, and we still need to connect with our past and our culture, even if we have to evolve and rebuild without presuming that everything has to be redone exactly as it was before." He pointed to the cathedral: "Look at Notre Dame. There were a lot of questions about how it was going to be rebuilt after the fire in 2019; people were forgetting that it had already been significantly reconstructed in the nineteenth century, largely owing to the public infatuation with Victor Hugo's novel *Notre Dame de Paris*. The cathedral came to symbolize Paris, not just locally but throughout the world. If the present still needs Notre Dame, and if the structure can be rebuilt to satisfy the modern imagination rather than exactly as it was before, then the question of whether the future will still need us probably needn't be

[2] Unless otherwise indicated, the URLs cited in the bibliography were verified on August 20, 2020 or later.

raised in quite the same terms. But the question of what we'll leave behind, what we'll bequeath to future generations certainly has to be addressed." Unfortunately, the recording cuts off here …

As the discussion between the imagined Parisians David and Jacqueline is meant to suggest, it is doubtless no exaggeration to assert that we are witnessing an unprecedented social revolution linked to a technological revolution – just as an earlier revolution in technology, the invention and diffusion of printing, was inextricably connected to the eighteenth-century Enlightenment and the social and political revolutions provoked by the new currents of thought. Today's revolution is anchored, as Joy's article demonstrates, in the development of what has come to be called NBIC (Nanotechnologies, Biotechnologies, Information science, and Cognitive Science) since the publication of a report by Mihail C. Rocco and William Sims Bainbridge titled *Converging Technologies for Improving Human Performance: Nanotechnology, Biotechnology, Information Technology, and Cognitive Science* (2003). One of the striking elements in Joy's article, and in many other texts on NBIC in general and artificial intelligence in particular, is the passive position attributed to humanity, as if it were the technologies themselves that were becoming active agents and thus the driving forces that make it possible to model the future.

2 The Enlightenment in reverse?

The core hypothesis that I propose to defend in this book is the following: we are currently experiencing a revolution *comparable to that of the Enlightenment*, but one that is *reversing direction*, thereby jeopardizing the advances humanity has made thanks to Enlightenment thinking.

The similarity with the previous revolution lies in particular in the domains in which the current one is embodied. If, like David, we were to choose Jean-Jacques Rousseau as our guide to eighteenth-century Enlightenment thinking, we would notice that the forms of self-discovery and self-expression found in his *Confessions* have been replaced today by messages in Facebook or Instagram, or even by the "recreative" genetic tests that tend to model a new history of the self and a new definition of individual identity as collective. The educational ideals expressed in Rousseau's *Emile* would be replaced today by vectors of a new educational model such as binge-watching or e-learning, or even by robotic babysitter/teachers. Today's version of his *Social Contract* would have to take cybergenetics, the human/machine interface, and artificial intelligence into account.

It is noteworthy that these significant changes are taking place precisely in domains – self-identification, learning, social and political policy – that were at the heart of the Enlightenment revolution. The coincidence is probably no accident, for these domains are precisely the ones at stake in today's great (r)evolutions not only in science and technology but also, and especially, in the social, ethical, and political realms. I see the current revolution as comparable to that of the Enlightenment in at least two respects. First, like that earlier revolution, it embraces the technological, scientific, social, ethical, and political realms *in a single movement*. Second, it raises or extends certain *key questions* from the Enlightenment period concerning identity, autonomy, morality, and the social contract.

This dual axis of comparison sets the stage for reconsideration of two fundamental – and fundamentally related – issues: power structures on the one hand, human relations on the other. Indeed, if the new revolution is indisputably comparable to that of the Enlightenment, it nevertheless constitutes its *inverse twin*. The inversion has to do with the fact that, whereas the Enlightenment revolution led to the spread of philosophical ideas including a way of thinking about freedom and autonomy that made it possible to begin to construct a more egalitarian, not to say democratic, society, the current revolution is based above all on a technicist ideology that, on the contrary, *reproduces or reconstructs forms of hierarchy while relying on forms of connection capable of eclipsing or even excluding or annihilating human relations*. As Henry Kissinger wrote in an article titled "How the Enlightenment Ends": "The Enlightenment started with essentially philosophical insights spread by a new technology. Our period is moving in the opposite direction. It has generated a potentially dominating technology in search of a guiding philosophy" (2018, 14). It is as though for some time now we have been seeking a guiding vision in technology by looking to engineers or computer scientists, or even to machines themselves for philosophical, ethical, and political principles. Or, to put it another way, this inverse twin of the Enlightenment operates as if ethics (or politics) were problems that the new technologies could and should solve.

A common thread in much of the discourse surrounding these new technologies, including Joy's in "Why the Future Doesn't Need Us," is the notion of the self-fulfilling prophecy, a motif suggesting that the future is already here. In this book I adopt the opposite stance, emphasizing that the future is in fact *not yet* here. On the contrary, as philosophers have been telling us at least since Augustine, it is the present that demands our attention. Even if – especially if – the future is being sketched out in the present, it remains precisely that which is to come.[3] Why then do we find ourselves in this paradoxical, oxymoronic situation in which the future *seems* already to have arrived? Why do we not try instead to conceptualize the future we want to see on the basis of what matters to us, while we attend to the present with care?

3 From technological experimentation to social experimentation

One of the basic hypotheses of this book, which seeks to respond at least in part to that paradox, is that this new stage of human history is simultaneously a technical advance and an experiment that is *also—and crucially—social*. It is moreover an experiment that the anticipatory discourse of self-fulfilling prophecies serves to *conceal*. Viewing the implementation of an emergent technology as a social experiment is hardly a new approach. Indeed, it is a widespread practice today: in the Netherlands, for example, as Ibo van de Poel makes clear in "Society as a Laboratory to Experiment with New Technologies" (2017). With the term "social experiment," Poel means to

[3] Translator's note: The French word *avenir*, "future," is formed from the phrase *à venir*, "to come."

convey a three-part notion of experimentation: first, experimentation *in* society, in real life; second, experimentation *on* society, which requires rethinking certain of its frameworks (legal in particular); and, last but not least, experimentation *by* society. The crucial challenge, as he makes clear, is how to determine the conditions under which such experimentation can be morally acceptable.

But the nuances Poel sets forth and the crucial problem he raises rarely come to the fore when the "social, ethical, and political" impacts of emergent technologies are under consideration, if they are considered at all. Rather than seeing the introduction of a technological innovation as an experiment, the guiding assumption seems to be that, as soon as a given technology is available, it will necessarily be used. This principle – almost a petition of principle – is sometimes mistakenly called "Gabor's law," despite the fact that the Nobel Prize winner Dennis Gabor, inventor of the hologram, was actually critical of such a technicist position ([1963] 1964, *Inventing the Future*). Were this understood, a possible public debate about this position, or indeed a necessary debate in response to public concerns, would not pertain to a so-called law.

The model underlying this techno-determinist rhetoric is the classic assertion that "there is no other choice," a claim that constitutes a leitmotif of today's uninhibited neocapitalism. Often bound up with digitalized management of society, the claim is frequently traced back to a famous statement by Margaret Thatcher: "So in a sense we do have to do it. Because there really is no alternative," sometimes abbreviated as TINA.[4] The defining feature of the TINA principle is that it restricts the field of possibilities by positing some form of dilemma and selecting a solution on the grounds that there simply is

[4] Thatcher made this statement during a press conference at No. 10 Downing Street on June 25, 1980. What is interesting here is that the logic of TINA goes hand in hand with the digitalization of society based on a form of social neo-Darwinism. (While TINA is generally associated with Thatcher, it would of course be a mistake to limit its scope to Thatcherism or to Great Britain.) I want to thank Catherine Larrère for suggesting that the logic I am describing is similar to TINA, which she also denounces; see Larrère and Larrère 2017, *Bulles technologiques*.

no other choice. An example of its use, and also of resistance to it, can be found in the case of facial recognition technology. Widely implemented after it had reached a high degree of refinement ("There was really no choice!"), reactions from civil society eventually led to strict regulation of its use or even its prohibition in major cities in California, the home of Silicon Valley and in Boston, next door to MIT – in other words, in places where the greatest concentration of state-of-the-art laboratories and industries in relevant technological zones are found.

Still, it remains clear that we as members of civil society are rarely asked our opinion when technologies we might regard as experimental are put to work; after all, facial recognition technology had been adopted without restrictions in the United States before it was socially rejected and then prohibited in certain contexts. Adults exposed to such experimentation can refer back to a previous state of affairs, pass judgment, and even band together to resist. But these experiments do not affect adult populations alone. They also, and perhaps especially, concern the younger generations and those to come. It is not self-evident that, if some of these technologies are inserted both into children's development and into everyone's daily lives (even if the spread of such technologies varies depending on country and geographical region), it will be possible to "backpedal" without difficulty. Worse still, as French philosopher and sociologist Jacques Ellul (*The Technological Bluff* [1988] 1990) has shown, a defining feature of what he calls the "technological bluff" is that it gradually takes on the look of inoffensive banality, only to infiltrate uncontested all aspects of daily life, imposing itself with the silent power of the self-evident.

If we look closely at what Dennis Gabor actually said on the subject, we find that the self-fulfilling prophecy is far from self-evident: he affirmed that "the future cannot be predicted, but futures can be invented" (Gabor [1963] 1964, 207). At the same time, he criticized man's capacity for invention when it takes on (what he calls) the "manly" appearance of destruction. According to Gabor, then, rather than positioning ourselves in the naïvely passive posture of accepting technology and its inexorability, we need not only to invent our possible futures, but we need to invent futures that are resolutely constructive in the sense of "nurturing" rather than combative. The dynamics of control or domination are hardly the ones to privilege if we want to make our approach to the future fruitful.

4 A "Sputnik moment" and technological neocolonialism

However, the crisis in which we find ourselves today is understood by certain European authorities as technology race that resembles – to borrow an expression used by Cédric O, Secretary of State for the Digital Economy in France from 2019 to 2022[5] – a "Sputnik moment," a moment in which the United States believed (probably wrongly) that they had been technologically outdistanced by the Soviet Union in the space race.

The choice of this term is all the more revealing in that, as we shall see later on with respect to genetics in particular, the model for the conquest of unknown territories—first and foremost that of space—by Western powers is a recurring paradigm in discourse about new technologies. But this choice is also significant because it is integrated into a vision of the world in which international relations would be dominated by the idea of a return, fostered by artificial intelligence, to the regime of colonial empires. There is nothing surprising, then, in the title of an article by Nicolas Miailhe (2018), "Géopolitique de l'Intelligence Artificielle: Le retour des Empires?" Miailhe asserts straightforwardly that the United States and China are going to rule the world thanks to artificial intelligence, that Europe is threatened with "cyber-vassalization" and Africa with "cyber-colonization." What is at stake, then, is world domination, presented by Miailhe as a question of absolute urgency. As a corollary, the quest for a European model under conditions of "extreme weakness and dependency" would articulate a "quest for sovereignty, a quest for power and respect for power," although the author fears

[5] Cédric O spoke during the Global Forum on Artificial Intelligence for Humanity (GFAIH) held in Paris in October 2019. The GFAIH served as the formal launch pad for the Global Partnership on AI (GPAI). See https://gpai.ai.

that the third term may attenuate the second (Miailhe 2018, 115). It is clear that the "Sputnik moment" evoked by Cédric O is not so much China's moment as Europe's, that of France in particular, and of other future "cyber-vassals" or "cyber-colonies."

Nevertheless, it seems to me, contrary to Cédric O's assertion, that, while the terms dependency, sovereignty, and power are of course intrinsically connected, what he calls the Sputnik moment can be read differently: the Québecois philosopher Marie-Hélène Parizeau calls it a moment of "normative opening," implying a necessary moment for reflection and debate on the issue at hand.[6] To some extent this expression echoes the dynamic anticipated by Dennis Gabor.

5 A normative and polethical opening

As the importance of ethical stakes in the realm of NBIC becomes increasingly clear, we are indeed witnessing a growing demand for reflection and debate on these issues. In other words, as humans, we are becoming aware of the fact that we risk being surpassed not only technically or economically but ethically and politically, for these technologies are actively reshaping our ways of life, our political institutions, and all the relationships that underlie our democratic civilizations. To be sure, this awareness has been underway for some time.[7] For example, the technicist turn has been analyzed by several philosophers, including Jacques Ellul, who, as we have seen, denounced the dictum according to which "since it is possible, it is obligatory."[8]

[6] Personal communication from the author. I borrowed Parizeau's term in a text setting the objectives for a workshop I co-organized with her in the 2019 Global Forum on AI for Humanity (GFAIH) and in the summary of the workshop published by CIFAR, the Canadian Institute for Advanced Research (2019).

[7] In the United States, authors such as Sherry Turkle (*Alone Together: Why We Expect More from Technology and Less from Each Other*, 2011) and Cathy O'Neil (*Weapons of Mass Destruction: How Big Data Increases Inequality and Threatens Democracy*, 2016) have been ringing the alarm bell for a number of years.

[8] Here Ellul is citing a statement Jacques Soustelle made in 1960 with reference to the atomic bomb. In Ellul's rephrasing, "everything which is technique is necessarily used as soon as it is available. This is the principal law of our age" (*The Technological Society* [1954] 1964, 99).

The turn toward performance and management in politics, in play at least since the Thatcher era, has led to what sociologist Albert Ogien calls the "digitalization" of society and politics to which I alluded earlier.[9] In any case, my hypothesis is that the current turn, by bringing to this process a vertiginous new scope that threatens to lead to a point of no return, infuses the necessary normative opening with new vigor and urgency.

This is why we are perhaps beginning to understand that the political responses have to be clearly articulated with ethics in the sense of a moral politics, a politics inseparably conjoined with ethics, something that might be called a *polethics*. Instead of a "Sputnik moment, or at least alongside such a moment,[10] perhaps we ought to be speaking of a "polethical moment." How is that moment to be understood?

I understand the term polethical in a sense close to the one proffered by the philosopher and poet Michel Deguy, who asserts, from the poet's point of view, a certain complicity among three practices, those of poetry, ethics, and politics.[11] In the preface to his *Réouverture après travaux*, published in 2017, in a characterization of poetry that I would propose to extend to polethics, Deguy writes: "[I]ts

[9] See especially Ogien 1995 (*L'esprit gestionnaire: Une analyse de l'air du* temps), 2009 ("L'hôpital saisi par la quantification: Une analyse de l'usage gestionnaire de la notion de qualité"), and 2010 ("La valeur sociale du chiffre: La quantification de l'action publique entre performance et démocratie"). I have had the opportunity to discuss this intuition briefly with Albert Ogien and I thank him for his remarks, which were very helpful to me in situating the problem.

[10] For the logic of power relations does not suffice to account for the complexity of what is in play here, just as the polethical question is obviously not the only one at stake. I think, though, that failing to address that question is to sidestep an important part of the problem of defining the moment.

[11] I am taking up Robert Harvey's lovely expression from "Les yeux dans les yeux": "Poetry – Politics – Ethics." "Michel Deguy sets these three practices in a conspiratorial relationship under the umbrella term polethics. In other words, these are three asceticisms for which one must have the eyes to see: to see what lies ahead, to see what lies behind, to see again, to foresee, to see the world, to see the other" (Harvey 2016, 266). By setting Deguy's work in resonance with Baudelaire's, Harvey also emphasizes the importance of the thematics of the gaze in Deguy's work and the role that empathy plays in this polethics, along with attention to the most vulnerable – themes that are also central to my own conception of polethics as a practice.

discursivity is not 'scientific.' The operation is one of opening, opening itself, reopening. In correspondence with openness to the opening of the world, to the *illumination* of the world …" (13).

In the similar sense in which I am using the word, then, the term polethics invites us to conceive of a robust articulation between ethics and politics: that articulation is the focal point of this book. This close articulation between ethics and polethics is far from self-evident in the philosophical tradition – indeed, the tradition often radically separates the two. Still, the connection is sometimes assumed by fundamental figures in political theory such as John Rawls. Furthermore, the linkage lies at the heart of feminist theory, where certain key thinkers in the ethics of care have rightly emphasized the need for such an articulation. For example, from her first work (*A Different Voice*) to her last (*A Human Voice*), Carol Gilligan has woven ethics and politics closely together, showing in particular that the ethics of care is devalorized because it is traditionally considered as the preserve of women. Thus we see why a feminist approach has to call into question the established order from social and psychological standpoints as well as from the standpoint of history. Similarly, Joan Tronto stresses the extent to which the binary division between ethics (devalorized/feminine) and politics (valorized/masculine) is problematic. Fiona Robinson, for her part, shows clearly in *The Ethics of Care: A Feminist Approach to Human Security* and in *Globalizing Care* that stressing the articulation between the ethics and the politics of care constitutes a fruitful approach for thinking through the problems she is addressing on a global level. Nevertheless, it seems important to me to go further still, by suggesting that polethics presupposes the articulation not only of ethics and politics but also of *poetics*. As I see it, the poetic, creative dimension may well be a particularly valuable and productive element for nourishing the interaction between ethics and politics, for opening up the field of possible relationships and sparking Deguy's "illumination of the world." While the standpoint from which I speak is not that of the poet, our two angles of approach are nevertheless deeply connected. The creative dimension is at the heart of the ethics and politics of care, which together open up perspectives that have been left in the shadows up to now. And to go even further, we need to explore certain technological fringes that require precisely a reconsideration of our classic categories and lead to the emergence of an unprecedented world that often seems to flirt with science fiction.

Moreover, the word "moment," borrowed here from Cédric O, is not philosophically neutral. As it is used in contemporary philosophy, first and foremost by Frédéric Worms (who has made extensive use of the concept; see for example Worms 2009), a "moment" is characterized by reflection on a set of common questions that may veer into sometimes unexpected relations and controversies. In a given "moment," then, thinking converges around common problems and is inscribed in a meaning-making relational network so as to constitute a philosophical problem.

To sum up the aims of the present book against this background, then, I would say that it attempts to grasp *the polethical stakes of the new technologies that I examine here, and to propose a constructive normative framework in which to situate them, while focusing on the ethics and politics of care*. This framework is intended to facilitate analysis of the issues without proposing ready-made responses. It is part of an effort to renew ethics by rethinking its frontiers and its modalities, while focusing on a diverse set of philosophical relations rather than a singular philosophy of domination and control.

6 Enlightenment with care

I attempt to respond to the question of the polethical stakes of the new technologies by using a method inspired by the British Enlightenment philosopher George Berkeley, a method whose duality is seemingly contradictory but is actually complementary: it conjugates a synoptic vision with a myopic one.[12]

[12] The classic interaction between myopia and synoptical vision was proposed in George Berkeley in *An Essay Towards a New Theory of Vision* (1709) and studied by Geneviève Brykman 2010 in "Courte vue et vision synoptique chez Berkeley" (2010).

On one side, then there is a "synoptic" vision, aiming to grasp the big picture and to consider the ethical stakes at the heart of the NBIC quartet, within which I have chosen to focus on three fields in particular: nanotechnologies, cybergenetics, and artificial intelligence (AI). Now, if we return to Kissinger's schema of a technology entailing a theoretical framework, it seems especially important to adopt a synoptic approach to interrogate the conception of the world that subtends the development of NBIC. For we have to contend here not only with a convergence among these technologies, as work by Roco and Bainbridge and others has emphasized, but with a system in which each element reinforces the others, resulting in what Jacques Ellul, citing Alain Touraine, has called a "programmed" society ([1977] 1980, 6).[13] To clarify this system, I shall sometimes revert to some of my previous examples or analyses, not to rehash them but to underline these systemic links.

Taking seriously Ellul's assertion that the "technological bluff" relies chiefly on "the suppression of moral judgment, with the creation of a new ideology of science" (Ellul 1990, 19). I attempt to show how that suppression of moral judgment is based on a form of fallacy that leads to a confusion of the artificial with what is right or good, a form that I propose to call "artificialist fallacy," relying on analyses that emerge from classic moral philosophy. And I propose to extend Ellul's hypothesis by showing that this operation also relies on a reconsideration of the concept of *care*, an issue on which I shall focus in the conclusion.

In keeping with Berkeley's methodology, in parallel with this synoptic approach I adopt a "myopic" vision, turning to specific cases whose conclusions moreover are not always necessarily generalizable, and whose specificities are sometimes irreducible despite the system. With this approach I seek to show how the three examples I have

[13] My thanks to Marie-Hélène Parizeau, who drew my attention to the relevance of Ellul's work to decoding the question of systems. I subscribe fully to her (unpublished) analysis, offered on November 11, 2019, at UNESCO during a symposium on Human Rights and artificial intelligence.

chosen from within NBIC, first nanotechnologies, second cybergenetics, and third artificial intelligence, all participate in the movement of entrapment by the dictates of self-evidence and gradual banalization described by Jacques Ellul, thus allowing the technological system to take over discreetly, without visible violence.

Georges Canguilhem, in *The Normal and the Pathological* ([1989] 1991), argues that any foreign matter is good for philosophy. But in the pages that follow, the matter is not only foreign but also sometimes frankly *strange*.[14] We are dealing with fields that are in the process of being constituted (even though they are no longer, strictly speaking, "emergent"), but also and especially fields whose *definitions* are problematic, either because they blur certain classic categories – the normal and the pathological, but also, for example, the living and the artificial – or because the definition of the field itself is at issue. These are also fields in which certain applications, such as sexual robots, may make some readers uncomfortable, or may seem trivial (socks that don't smell bad? rug-cleaning robots?) or least not noble enough to merit the interest of philosophers. It seems to me, though, that this strange material that tends to insert itself more and more into daily life, into our ordinary lives, to the point of becoming absolutely familiar and unavoidable, urgently demands careful study by philosophy, moral philosophy in particular.

The examples I have chosen have at least two difficulties in common: in each case it is hard to tell whether we are dealing with a revolution or with an unprecedented combination of technologies, and it is hard to assign borders or even to define them. This is why in all three cases I have made an effort to clarify both the conditions of their development and their definition, despite the risk of annoying repetition. The difficulty of the task at the epistemological level is compounded by the social, ethical, and even political issues that these cases raise; to my mind, these complications suggest that the

[14] *Translator's note:* The French word *étranger/étrangère*, "foreign," contains the root word *étrange*, "strange."

seeds of the polethical moment are already contained in the constitution of the fields I am examining.

I am not a specialist in nanotechnology, cybergenetics, or artificial intelligence, and it is likely that certain inaccuracies on the technical level have slipped in. For this I apologize sincerely. But in one sense I don't apologize fully, for it seems to me that the fear of making technical errors too often keeps (at least most French) moral philosophers from intervening on the subjects I am addressing,[15] and I believe that moral philosophers have an important contribution to make, given that the issues I am raising concern the ethics of these new technologies understood on two levels. They involve not only applied or practical ethics, that is, the ethics of the new technologies in which these technologies are the *objects* of ethics, but they also involve metaethics, that is, the ethics of the new technologies in which ethics is understood to be an order of reflection that those technologies can profoundly modify, i.e. the *subject* of the process.

My own fields of competence are those of ethics and political theory, although I have often worked at the interface with the so-called hard sciences. If all philosophers are reputed to know that they know nothing, this is all the more the case for ethical or moral philosophers (I make no distinctions between the two terms ethics and morality), since in my view ethics is not a constituted body of knowledge but rather a way of interrogating the world, framing problems, and of seeking solutions. It seems to me, however, that this anchoring in ethics constitutes part of the originality of the present book: its modesty (on the scientific and technical levels), but also its ambition (on the moral and political levels).

This ambition is double. One the one hand, it is a matter of attempting to grasp some of the ethical problems that arise in the three technical fields I am studying by considering them as a system.

[15] The opposite is not always true, however: we often hear persons with little or no training in moral philosophy sum up rather imprecisely certain analyses of moral or political philosophy without ruffling many feathers – and this is a good thing, because it allows conversation to begin and advance.

This system is such that each technology implicitly supports each of the others. But the system effect entails a risk, one that I term a translational risk. If the analysis is not sufficiently refined, the condemnation of some applications could have repercussions on others and could lead to a technophobic approach, which is certainly not one I espouse. It seems important to me, in fact, to move beyond the often sterile opposition between technophobia and technophilia and to seek an ethical and political stance for the technologies that would be '*philotechnological*', in the sense that one speaks of philosophy as the love of wisdom: the goal, in other words, is to engage a quest for the wisdom of technologies while sustaining an always possible questioning of these same technologies. It is owing to this stance and the hope induced by the polethical moment, the dynamic of a normative opening, that this book is not pessimistic; on the contrary, it seeks to propose guardrails in order to preserve a constructive rather than a conservative dynamic. Furthermore, while I study the three technologies separately, the most significant features of the system in which they share become most apparent at the end of the analysis, allowing the vision of the whole that I develop in the conclusion.

On the other hand, it seems to me that the originality of these new technologies blurs the way we are accustomed to thinking, and some of the terms I mention in order to talk about them, especially in the field of artificial intelligence, further contribute to that blurring by imperceptibly orienting our way of thinking about them; this is the case, for example, for terms giving the impression that the artificial is alive or that the technological is social, ethical, or even political ("intelligence," "life cycle," trust," and so on).

While close attention to terminology is essential, then, it is not enough. Ethical reflection itself must be renewed if we are to analyze the current revolution. I propose to contribute to that effort by drawing, among other things, on a current that provoked an upheaval in ethics starting in the 1980s: the effort to conceptualize ethics from the starting point of relationships and to highlight the links between ethics and politics. This approach, called the ethics and politics of care, is a plurivocal conception, still under construction, that brings together the various senses of care (care about, care for, care to, take care of, care with, and so on), while engaging the affective, intellectual, and practical dimensions of these notions.

Whereas the ethics of care has up to now too often been categorized dismissively as "women's ethics," of conceivable interest (at best) for thinking about "womanly" topics such as vulnerabilities, healthcare, or education, the concept has too rarely been mobilized for thinking about ostensibly more valuable topics such as technologies, which are often considered more "manly" (with some rare but important and inspiring exceptions that I shall point out in the conclusion). Now, my hypothesis is that the ethics and politics of care must be fully integrated into reflection on NBIC, which seemingly seeks to neutralize and even to overturn care. For the concept of care allows the opening up of possibilities as well as the concern for future generations that must constitute one of the principal stakes of what we could call NBIC ethics.

I should like to suggest, then, that the current polethical moment that I have been evoking is a *moment of care*, not because the ethics and politics of care offer a panacea to the ethical and political dilemmas raised by these new technologies and the NBIC system, but because they often make it possible to take a clarifying look at those problems, and in so doing, to seek paths that would allow us to formulate them as well as possible, if not to solve them. It seems to me in particular that the notion of care allows us to conceptualize in new ways what is perhaps the most problematic issue of all in this polethical moment: the issue of responsibility. We can also go further than Kissinger did by suggesting that this "reversed" revolution of the Enlightenment requires a response that does not coincide with the invention of the concept of autonomy – the philosophical analysis of which has shown that it proceeded in large part from that revolution[16] – but rather with the development of the question of relations, inseparable from the idea of care.[17]

[16] As Jerome B. Schneewind has noted in *The Invention of Autonomy: A History of Modern Moral Philosophy* (Cambridge, UK: Cambridge University Press, 1997).

[17] So much so that some writers have proposed to introduce the notion of relational autonomy in order to take this problem into account. See for example Natalie

My object is not to "apply" care to the technologies but rather to shed reciprocal light on each: in other words, I am proposing to conceptualize nanoethics, cybergenet(h)ics, AI ethics, and care simultaneously. This dynamic will make it possible, I believe, not only to conceptualize the ethics of these technologies more pertinently but also to nourish the ethics and politics of care. Since care is a polyphonic domain, as we have seen, one that is not truly unified in the form of a system or a rigid school of thought, it is perhaps important to say a word here about the way I see myself in this current. My approach does not set up an opposition between care and justice, but it considers the existing theories of justice inadequate to account for the complexity of the moral world. This does not mean, as I see it, that the ethics of care would *complete* the theories of justice in the sense that the former would be subordinated to the latter. Like many feminists, I consider that the ethics of care brings to the field of ethics a fundamental perspective that it did not have before, because the ethics of care had previously been devalorized or denied.

It is difficult to "sum up" the ethics and politics of care, but as a start I can propose some general outlines. What is at stake is a pluridisciplinary feminist approach born in the field of psychology and later developed in other disciplines, especially (although not exclusively) in philosophy and political science. These approaches all consider that it is unacceptable, on the normative level, and inappropriate on the descriptive level, to exclude from the moral field certain tendencies, attitudes, and activities that have to do with care in all its dimensions (care for, care to, care about); they are conceived as a process that takes into account the care giver as well as the care receiver, and they are inscribed in a democratic process (care with). The ethics and politics of care also have in common the fact that they prioritize persons' needs, their relations, and the dimensions of their existence. It is not a question of applying principles from a position of authority (from the "top down").

Stoljar and Catriona Mackenzie, *Relational Autonomy: Feminist Perspectives on Autonomy, Agency, and the Social Self* (New York: Oxford University Press, 2000).

My approach is close to what Fiona Robinson calls a critical conception of care. She defines it this way: "The critical lens of care ethics exposes the ways in which dominant norms and discourses sustain existing power relations that lead to inequalities in the way societies determine how and on what bases care will be given and received."[18] I am also in harmony with Robinson when she stresses that care can and must be understood as a global approach proceeding from a relational ontology.

My own approach to care is anchored in a Francophone environment that has developed significantly over the last twenty years. This "Francophone care" is thus hard to characterize, but I can at least indicate three major threads, as I see them. First, there is a fundamental linguistic difficulty, since the borrowed term "care" can be translated in various ways (as a noun: *soutien*, *souci*, *traitement*, …; as a verb, *être concerné*, *tenir à*, *avoir de l'affection pour*, …), and this has complicated its spread in France: a supplementary effort has been required to articulate the various facets of care, especially so as to avoid reducing it to the medical context alone. Moreover, the Francophone care ethics and politics has been developed in a specific linguistic and cultural framework that has led to the establishment of conceptual distinctions that are not necessarily present in the Anglo-Saxon context. The notion of *connexion*, as distinct from *relation*, is a good example. Unlike the English term "connection," which is more or less synonymous with "relationship," the French term *connexion* more strongly connotes links between things or objects than links between persons; it is also used specifically in fields such as electricity and information science. For this reason, I have chosen to retain the French spelling (which may look like a typo or an archaism to some English-speaking readers) to refer to the connections (*connexions*) that have been developed in cybergenetics and information science and are not authentic relationships. Second, this Francophone care has also sought to avoid being anchored in an opposition

[18] *The Ethics of Care: A Feminist Approach to Human Security* (Philadelphia, PA: Temple University Press, 2011), 28.

between a continental, analytic philosophy and a Franco-Germanic and Anglo-Saxon philosophy; it has adopted an integrating perspective that does not deny the originality of care but does not install it in a binary that would set in in direct opposition to other approaches. It is also inscribed in a universe of discussions featuring European voices (which are not always very audible in the rest of the world for want of translation[19]) and focusing on approaches developed in the English language. Third, I would like to believe, without being entirely certain of this, that the European anchoring makes it possible to deal with new objects, as attested for example by the inspiring work of Maria Puig della Bellacasa on soils, or that of Aimee van Wynsberghe on robots, approaches to which I shall return, especially in the conclusion – approaches that show the fruitfulness of this movement of thought that is under construction.

Why embark here on a reflection on the ethics and politics of care? What do these considerations contribute to the ethical and political discussion of the technologies in question? These questions will of course be explored continually throughout the book, but I can already propose four principal paths toward answers. The first is methodological. As Fiona Robinson suggests (2011, 26), invoking Margaret Walker's work, it is important to conjugate "reflective analysis" and "critical reflection" together, which implies not considering philosophy as an abstract task but rather as an undertaking nourished by the analysis of contexts. For that reason, my work relies quite heavily on the analysis of specific examples and on work carried out in fields of the human and social sciences other than philosophy: sociology, media studies, and so on. Moreover, as Fiona Robinson emphasized, the dimension of "critical reflection" requires bringing to light power relations, even relations of domination, that are at work via the technologies I am examining. In this process, the domain of

[19] Except for one (important) book: Sophie Bourgault and Frans Vosman, eds., *Care Ethics in yet a Different Voice: Francophone Contributions* (Leuven: Peeters, 2020).

care brings to these technologies a prospective undertaking based on concrete analyses.

Second, and in a way that is inseparably complementary to the first point, the usual discourse about these technologies too often relies on a self-fulfilling prophecy (one that precludes the possibility of an alternative) and on the idea of anticipatory design (which puts in place a double lock, technological and ethical, since the future is certain and cannot be otherwise). To counter that strategy, the ethics and politics of care help us to refocus on what matters to us and to take care of a present that integrates into its project the sustainability of a world. It also allows us to avoid focusing solely on the question of risks. Rather than anticipatory design, I would like to propose an "ethics by design" in this book, while helping to raise the question of the forms of life to be constructed with the technologies at issue.

Third, these technologies raise serious questions about the notion of autonomy, around which the West has long constructed the foundations of its moral and political thought. The technologies put into play, in a different but essential way, the notions of relations and responsibility, notions about which the ethics and politics of care offer particularly salient analyses in efforts to resolve the problems on which I shall focus.

Fourth, these technologies engage disruptively with certain borders or even binaries that we have long considered self-evident (private vs. public, for example), and offer elements that allow us to interrogate and rethink these binaries. Rethinking binaries is at the kernel of the ethics of care, which can thus offer a relevant approach in this perspective. Furthermore, as Gabor has already suggested ([1963] 1964, 165), it is perhaps time to get away from valuing strife and progress – which he characterizes as "masculine" – and to attribute greater value to the dynamics of domestication and preservation by relying on an approach he qualifies as "feminine." Still, while the project strikes me as sensible, I think it important to leave behind the potentially essentializing masculine/feminine duality – which Gabor also relates to the Yin/Yang opposition. It seems more fruitful to me to concentrate on the conclusion of his argument, which

I would qualify as "feminist" in the sense that he goes beyond those oppositions to propose an approach that subverts the classic relation of domination between politics and technology. Gabor in fact sets forth what the necessary movement of innovation must be, by asserting that the central question is less feasibility than desirability, claiming that this positions the social inventor and the technological inventor on the same plane.

Instead of asking why the future no longer needs us, as Joy did in the 2000s, it may therefore be necessary to ask, with a bit of optimism, questions like these: What future do we want? What world do we wish to pass on to future generations? Or, in other words: What matters to us? What do we care about, and how we can take care to stand up for it together?

A final word, here, to explain the standpoint from which this book has been written. Intended to build a bridge between Europe and North America, it stems from several years of research conducted primarily in France, but also in Los Angeles, thanks to the support of the French National Center for Scientific Research (CNRS) and to the hospitality of the UCLA Institute for Society and Genetics. My analyses focus largely on the American history of NBIC, from a viewpoint that is rooted essentially in French and/or Anglo-Saxon philosophy and that seeks to build a theoretical bridge between Europe and North America. Thus while certain elements that may be self-evident to North American readers are sometimes spelled out in the language of a neophyte discovering (that is, both becoming aware of and exposing) what lies in plain sight, it seems to me that the very naiveté of the gaze has made it possible also to (re)pose certain essential questions.

Part One: Nanoethical Stakes

CHAPTER 1

Nanoethics

This opening chapter provides some necessary background for the analyses to follow. I aim to bring to light crucial elements of the technological system, and to point to some of the key problems posed in particular by the system summarized as NBIC. For the nanotechnologies are *hard to pin down*: they are at once difficult to conceptualize and difficult to master on the technological level. In addition, their development has been based on a *mythology* grounded in the notions of *self-fulfilling prophecy* and *anticipatory design*; this has helped them penetrate into every aspect of our daily lives, from our sunscreen lotions to the candies our children covet. To deal with the ethical issues raised by the nanotechnologies, we shall have to go beyond the question of how these technologies are to be managed, and beyond what is generally (if wrongly) called their social "acceptability": we shall have to identify and address the specific ethical questions they provoke. This undertaking requires new ways of asking such questions and of conceptualizing the ethics of nanotechnology (and, more broadly, of NBIC): a challenge that the ethics and politics of *care* can help us meet.

1.1 Did you say "nano"?

Nanotechnologies constitute the basis, in a way, for the two other technological fields on which I have chosen to focus, cybergenetics and artificial intelligence, because they enable the portability and the low cost that are central to the development of those other fields, and they make the polethical considerations that are the subject of this book clearly inseparable from the economic issues involved, even if these issues often remain in the background. But nanotechnologies are also caught up in a powerful interdisciplinarity and a worldview that shifts the classical boundaries of the sciences and technologies as much as it shifts the frontiers of the world as we have become accustomed to conceiving it.

1.1.1 *A controversial scientific definition*

It is customary to refrain, in philosophy, from beginning a disquisition with a definition, so as to avoid obstructing the work of conceptualizing and problematizing. As we have seen, however, in dealing with NBIC in general and nanotechnologies in particular, it is important to begin by raising the question of how to *define* nanos, precisely because the definition is not a given, and the very fact that it is problematic is relevant to certain difficulties that may arise for anyone interested in the ethics of the nano field.

It is tempting, of course, to reduce nanos to a matter of scale, to 10^{-9}m, that is, to a scale at the extreme minimum. The "official" definition of the nanotechnologies, as it appears in the reports of the National Nanotechnology Initiative,[1] was introduced in 1999. Its emphasis is on scale,[2] specifying a range between 10^{-9} and 10^{-7}, that is, between 1 nanomicron (nm) and 100 nm (or 0.1 microns). Nanometric objects are thus the smallest objects that humans have ever been able to manipulate consciously. As such, nanoparticles are viewed as building blocks, in a "bottom-up" conception. However, the definition in terms of scale is not very useful, owing to the vast diversity of the nano field and of nanoparticles themselves. The same nanoparticle can be developed in different forms that may have different properties. The unification of the field by scale comes up

[1] The National Technology Initiative is a research and development project that brings together the activities of American nanotechnology agencies.

[2] See, for example, National Technology Initiative n.d., "What's So Special about the Nanoscale?"

against an infinite diversity that risks leaving the field indefinite, undefined.

The "primordial" definition of nanos in terms of scale is accompanied by another essential characteristic: at the nano scale, the materials, mechanisms, or systems involved may well present properties, behaviors, or functions that are not found at a different scale. For example, at the nanometric scale, silver presents anti-bacterial properties. The fields of research and conception thus opened at the ever so tiny nano scale are, conversely, immense.

Historically, the origins of nanoscience as a discipline are closely linked to advances in instrumentation. In 1989, IBM scored a media coup, but also a technoscientific breakthrough, when its researchers spelled out the company's logo by using 35 Zenon atoms, which they had been able to manipulate one by one thanks to the new Scanning Tunneling Microscope. In electronics, it is becoming possible to build transistors starting with a single molecule, a development that will open the door to unprecedented miniaturization. It is becoming conceivable to install in the human body apparatuses so small that they can "target" or trap certain cells or organisms that have been inaccessible up to now.

Should we be talking in the singular, then, about nanotechnology, or rather in the plural, about nanotechnologies? Or should we be talking rather about nanoscience? Should we stress the disciplines involved in the conception of this science and speak of nanosciences, which are multiple, from chemistry to engineering? Or should we stress the emerging uses of this science by the nanotechnologies, of which there are already too many to count?[3] British terminology invites us to distinguish between nanoscience (in the singular), as the study of the nano field, and nanotechnologies (in the plural), thus emphasizing the plurality of applications and of possible tools,[4] from

[3] For example, biomedical technologies, information technologies, energy production and storage, materials, fabrication, instrumentation, research on food, water, the environment, security, and so on.

[4] Royal Society & Royal Academy of Engineering 2004, "Nanoscience and Nanotechnologies, Opportunities and Uncertainties."

conception to utilization. It might be most appropriate, then, to speak here of "nanotechnosciences," stressing the interweaving of science and technology within the nano field without denying the plurality of the disciplines involved. But since the term "technosciences" is somewhat controversial today, using it would risk making things even murkier.[5] So I shall avoid using it and refrain from joining the debate; instead, I shall take a prudent approach and fall back on the simple term "nano," in the singular or in the plural.

The new scale at which the nano field operates and the complex interweaving of a wide variety of sciences and technologies within that field are complemented by another no less essential characteristic, as we have seen with the example of silver: the nanos are *enabling*. This means that they allow researchers in existing scientific disciplines or technologies to surpass previous limits by allowing them to operate on a new scale, on which certain properties are modified. Thus the nanos are making it possible to develop new diagnostic and therapeutic tools, enabled not only by the new scale but also by the new properties of the materials involved. This is an essential point, for it not only underlines the idea that "nano" is not, *a priori*, a unified field, but that it is characterized rather by its capacities for interaction with fields that are *a priori* distinct. In other words, alterity is constitutive of the field's identity. It is thus logical that nanos should be developing within a multitude of already-constituted disciplines and technologies as part of a field that is neither entirely new nor entirely the same as the familiar fields with which it interacts.

This impression of *déjà vu* is heightened, moreover, by the fact that certain properties on the nano scale have been used since ancient times without having been explained or understood – consider the well-known example of the Lycurgus cup, which takes on a different color depending on whether it is illuminated by daylight or by a light source located behind it. This property owes its existence to the presence in the glass of metallic nanoparticles stemming from an alloy of gold and silver of around 79 nm. The master glassblowers of the

[5] See especially Sebbah 2010, *Qu'est-ce que la "technoscience" ?*

fourth century BCE thus worked with nano "unknowingly," so to speak. Similarly, many scientists today assert that they had already been working with nano before the term came into common use, but that adding the label "nano" to the description of a project opened the door to new financing as scientific policy agencies in the developed countries began to encourage work in the field.

1.1.2 *A social and political question?*

Given all the complexities, it is hard to avoid noting, ironically, that the nano field is characterized by a single point of agreement about its definition: an agreement to disagree. As we have seen, the absence of consensus contributes to the difficulty encountered in analyzing NBIC. To complicate things further, there is a similar lack of consensus around the question of the "birth date" of the field. It is often credited to a 1959 lecture by Richard Feynman titled "There is Plenty of Room at the Bottom" (1959); some, however, prefer to see Eric Drexler (1986, *Engines of Creation*), as the founding father of the nanos. The rise of the nanos is interestingly inseparable from a whole mythology as well as an aesthetic that has been facilitated by the invention of new microscopes with a tunneling effect, allowing researcher to visualize and manipulate nanos. The development of the nano mythology is thus inevitably entwined with the feats accomplished in the experimental domain. Should we perhaps go looking for the definition of the nanos not in science or in logic but rather in the realm of the imagination?

It is undeniable that science fiction has played an essential role in the construction of the nano field from the outset, to such an extent that a veritable nano epic has arisen. This epic shares the ideological dimension of the traditional French *chanson de geste*. One characteristic of these epic medieval narratives is to exalt the value of a political system (for example, the feudal system in medieval epics) and the relations structured by that system; in addition, and perhaps most importantly, the system establishes what François Suard, in *Chanson de geste et tradition épique en France au Moyen-Age* (1994, 25) calls a "geography of desire" that constitutes the imaginary framework of conquest. This framework unfolds according to the inseparable modalities of the time and space imagined and desired; it abolishes certain traditional borders and asserts others as intangible. Thus Drexler seeks not just to *describe* a molecular factory that is full of universal assemblers in the form of self-replicating nanobots, but also to *convince* his readers that such a factory can be created. We are seeing a mix of science and fiction from which the distinction between what is natural and what is artificial seems also to have been removed. Even more significantly, Drexler advances the idea that the fiction encountered in this context is self-realizing, and that the present is generated by the vision of the future.[6]

At the heart of this geography of desire, the perception or even the conceptualization of temporality constitutes a fundamental element allowing us to grasp the posture that has presided over the creation of the nano field. It is as though the future nanotechnologies were already virtually here, as if time were playing out in a mirror where the past is not truly past, because the nanotechnologies were already there before, and where the future is no longer "to come," but already announced as present. This restructuring of time by the nano field calls into question a fundamental concept of our mental categories, just as the notions of space and spacings are already called into question by the issues of scale and permeability (issues to which we shall return). What is more, in the nano field, theory is closely tied to application, as if the application were *already* present in the theory – or at least as if it *should* be found there. As Sacha Loeve explains it, "the achievements of the present (such as our molecular machines) are assessed and measured in relation to a concept presumed to have a future application."[7] This idea of "anticipatory design" tends to organize the present in view of a projected future, rather than

[6] See especially Dubois 2016a, "Les nanotechnologies à travers l'imaginaire collectif" and 2016b,"Between Science and Fiction – Nanotechnology Shift in American Comic Books." See also Maestrutti 2011, *Imaginaires des nanotechnologies. Mythes et fictions de l'infiniment petit.*

[7] Loeve 2009, "Le concept de technologie à l'échelle des molécules-machines: Philosophie des techniques à l'usage des citoyens du nanomonde" (43, 45).

conceiving of the future from the standpoint of the present: "the best way to predict the future is to create it" (Roco and Bainbridge 2003, *Converging Technologies for Improving Human Performance: Nanotechnology, Biotechnology, Information Technology, and Cognitive* Science, 33). And the future is organized in an inexorable sense, quite remote – despite the echo effects – from the approach advocated by Dennis Gabor, according to whom, as we have seen, "the future cannot be predicted, but futures can be invented" ([1963] 1964, 207). With "anticipatory design," Gabor's constructive pluralism is sacrificed to a monolithic conception of the future that corresponds perfectly to the mislabeled "Gabor's law" according to which what is technically possible will come to pass.

The notion of "anticipatory design" seems to have been forged by the transhumanist father of cryogenics, Robert Ettinger (1964, *The Prospect of Immortality*).[8] The displacement of this notion into the nano field is significant in itself: not only owing to the close connection there was at the time, in many people's minds (and in popular culture), between the exploration of space and the freezing of bodies, the only technologies capable, in combination, of ensuring the survival of the human race once this planet has run out of resources, but also for two other essential reasons.

First, because freezing bodies abolishes temporality: the present and the future are one, since the present is at once frozen and unfrozen in the future. This theme has been developed in a number of fictional contexts: in *Captain America*, for example. This example is not without significance, for the nano epic, like cybergenetics and artificial intelligence, are in many respects "an American story," in the sense that they are integrated into a certain vision of the nation and its relations with others, in the way Americans have of narrating their history and constituting their own mythology – although of course other nations are also developing these technologies and, more or less consciously, the same sort of mythology,

[8] See also Damour 2018a, "Le transhumanisme, inspirateur ou idiot utile du biocapitalisme? Le cas exemplaire de la vision d'Eric Drexler."

Second, because, as Franck Damour has emphasized, cryogenics entails a prophetic vision of a post-apocalypse in which the survivors will be the ones who believed in the future that was already in the present. In this sense, cryogenics is also a metaphysical form of thought and a method of reflection that underlie transhumanist logic. Let us note here that this way of organizing thought according to the logic of anticipatory design is a characteristic feature of NBIC. Now, the increasingly recurring critiques of the "unfulfilled promises" of artificial intelligence, or of the "stupidity" of those promises, rather than of artificial intelligence itself,[9] underline the inability of these technologies to produce absolute novelty; this makes the notion of self-realizing prophecies all the more convenient in that the very idea of a possible alternative is nipped in the bud, as it were, and thus rendered useless. It is important to keep this operational dimension of self-fulfilling processes in mind in the analyses that follow.

Moreover, from the standpoint of industrialists or investors, anticipatory design makes it possible to limit the field of what can be done and to ensure the viability of the system chosen. It is a matter of an implicit technological barrier, in the sense that possible innovations are filtered out by way of anticipatory design. However, as Damour points out in "Les nanotechnologies comme technologie transhumaniste" (2018b), the barrier is also ethical: by including ethical analysis at the outset, or by pretending to do so because one is seeking what is best for humanity, one is also in a sense controlling the scope of that analysis. Anticipatory design thus integrates a social dimension by transforming the world in its political as well as its epistemological aspects. As a result, a prophecy can *only* become self-realizing; its principal aim is to enter into a mythical perspective rather than to adopt a prudent future-oriented approach.

[9] See for example Broussard 2018, *Artificial Unintelligence: How Computers Misunderstand the World*, and Stiegler 2018, in a discussion held at the French National Library (November 10, 2018), under the title "Les algorithmes et la bêtise artificielle."

Might the nano field, then, be a myth? Must we believe in this myth? Above all, what would be the function of the myth? Despite what might be one's first impression, it seems likely that the myth has attained a certain everyday quality, for it fits perfectly into popular culture, where it takes on reassuring and even protective features that allow it to spread in an "ordinary" way, along with the significant development of nanotechnologies in our everyday products. The exceptional is being transformed, in a way, into a reassuring banality.

This point further increases the inherent difficulty of defining nanos, for they entail a form of strangeness not exempt from banality. Might "nano" be no more than an umbrella term, even just a buzzword? One often encounters the expression "nano-hype," but also "nano-hate," to refer to an approach to nanos that is at once hyperbolic (highly positive or highly negative) and phantasmatic, rooted at least in part in an interplay of passions that is hard to untangle from the argumentation.

Must we then distinguish between *hype* and *hope*, as David Berube proposes in *Nano-hype: The Truth behind the Nanotechnology Buzz* (2006)? In that case, it would be a matter of distinguishing between the hype that stems from overselling, from an economy of promise, and rational hopes, without falling into technophobia, or "nano-hate." This critical tripartite distinction may be seductive at first glance. But it seems to me that we must go farther than Berube and introduce a fourth term, which is in a way the hidden facet of hope. If the hopes in question are *rational*, they are not necessarily *reasonable*, that is, their objects are not necessarily desirable from an ethical standpoint. In other words, to borrow from the philosophical tradition, the key question would be to ask, in a distant echo of Immanuel Kant, not so much "what can I know?" as "what am I allowed to hope?" We would then need to reflect on what a reasonable hope might be, a hope that would be neither delirious nor exclusively rational nor separated from ethical questioning.

Furthermore, if the history of nanos is inseparable from fiction or even from science fiction, then, *a contrario*, the question of the reality of nanos takes on particular importance. This is why, as Alfred Nordmann has emphasized in “If and Then: A Critique of Speculative NanoEthics” (2007), we must not let the path of speculative philosophy (which is sometimes aligned with that of science fiction) distract us from reflection on what the nanos really are and really do in the here and now. Even if this type of reflection risks leading to rapidly outdated results, it remains nevertheless an indispensable exercise. For if it is not repeated in the here and now, the future will always be viewed as already here in the present, and thus uninflectable, and the idea that “the future no longer needs us” will be reinforced. Moreover, we have to ask whether one can even legitimately claim to be evaluating futurist scenarios in ethical terms. Would that not entail committing a double error? On the one hand, it would mean supposing that we can predict the future and that we understand what will constitute it; on the other hand, it would mean supposing that ethics has an immutable essence that cannot be transformed by the technological changes or the processes of reflection that become integral to it. In short, the effort to define nanos, whether by way of science or by way of the imaginary realm, runs into so many obstacles that it seems to end up in an impasse. Nevertheless, one possible strategy could consist in envisioning this prospective movement from the starting point of what matters to us, taking care to leave future generations with the possibility of making choices. In other words, what is problematic is not the possibility of envisioning futurist scenarios, but the way self-fulfilling prophecies do that, as opposed to prudent and attentive – care-ful in both senses – formulation of future prospects.

1.2 Toward a political definition of nanos? Risk and its avatars

If neither science nor imagination is well equipped to define nanos, do we perhaps need to turn, as a last resort, to the political field? Such an angle of approach may seem surprising at first glance. Still, the hypothesis of a political rather than a scientific starting point is not completely unfounded. Indeed, the third name often cited along with Feynman and Drexler as a founding father of nanos is that of Bill

Clinton, who launched the National Nanotechnology Initiative in 2000. Moreover, as shown by Brice Laurent, who has analyzed the controversies over defining nanos in the European context (2013, "Les espaces politiques des substances chimiques"), the process has been eminently political. In other words, as Bruno Latour pointed out in "Pour un dialogue entre science politique et science studies" (2008), we are far from the classic Weberian bipartite divide between the scientific and the political realms.

Thus we need to acknowledge at the outset that the disagreement over definitions is not only scientific, it is also social and political. Laurent thus distinguishes three strategies for defining nanos: first, "by science," second, "for the purpose of regulation," and third, "in terms of size." The first leads to the perplexities we have already been considering.[10] The second led initially to specifying nano objects and assigning them to pre-existing categories in the legislative domain. However, in 2009, the European Union adopted a specific framework for nanos that relied initially on scale but then moved to make properties a central feature of the characterization.[11] The question of properties is not merely a matter of description; it focuses on possible modifications and risks to consumers and/or to the environment. The third strategy, distribution by size, defines as nanos any materials at least 50% of whose components are at the nano scale.[12] The fundamental question that arises in this third case is the inclusion of nanos

[10] Laurent cites a member of a working group within the International Organization for Standardization (ISO) who makes the perplexities explicit in the following terms: "Nano is an abnormal group. We've never done this before. It's really about taking the beginning of the scientific basis to understand what we're talking about. … Usually, we're looking at products. But we're ignorant of what nanotech is" (Laurent 2013, 199).

[11] See European Parliament, Council of November 30, 2009, on cosmetic products, Regulation EC No 1223/2009, article 2, line k: "'nanomaterial' means an insoluble or biopersistent and intentionally manufactured material with one or more external dimensions, or an internal structure, on the 1 nm to 100 nm scale"; and European Parliament, Council of October 25, 2011, on the provision of food information by consumers, both cited in Laurent 2013, 206.

[12] European Commission Recommendation of October 18, 2011, on the definition of nanomaterial (2011/696/EU), cited in Laurent 2013, 199, 207.

in composite structures. Now, as Laurent emphasizes, one of the essential elements in the effort to define nanos is their inclusion in a risk management process. Far from an anodyne consideration, the concern for risk management attests to a certain worldview and a specific metaphysical orientation or even an ideology, as we shall see.

1.2.1 *Three approaches to "accompaniment"*

The introduction of nanotechnologies within a "risk society," which has been theorized most notably by Ulrich Beck in *Risk Society: Towards a New Modernity* (1992) stresses the shift from natural risks to manufactured risks whose consequences would need to be anticipated in advance. While, at the time nanos were developing, the French political climate was marked by issues such as contaminated blood, mercury, and mad cow disease and thus may have been particularly sensitive to questions of risk, the idea of a necessary "accompaniment" of nanotechnologies by the human and social sciences (SHS) was affirmed on the international level, especially in Europe and the United States. Virtually every prospective report from that period cites, at least in passing, the importance of articulating the development of nanotechnologies with a social or even ethical analysis of their implications. Indeed, in 2007, the Springer publishing house introduced a journal titled *Nanoethics*.

Still, the modalities of such accompaniment were hardly the object of a consensus; as has often been the case in the quest for definitions in the nano field, they were in fact subject to lively controversies. While these modalities came in many different forms, we can identify a certain number of dominant positions and concepts that make it possible to trace their cartography – and sometimes their geography, so deep are the cultural differences. Their names are not very poetical, but they are meaningful: Ethical, Legal and Social Impacts; Embedded Humanism; Risk Assessment; Public Engagement; codes of conduct; toolbox, and so on. Among these, three approaches stand out: in order of importance, they are the ELSI approach (Ethical, Legal and Social Impacts), the CTA approach (Constructive Technology Assessment), and the approach via Public Engagement.

The ELSI approach is the one that has been most widely adopted by scientific projects, which generally include a "package" devoted to the social, legal, and ethical impacts of their work, although clear distinctions between those three categories are rarely made. As Bernadette Bensaude-Vincent shows in "Which Focus for an Ethics in Nanotechnology Laboratories?" (2013a, 23), this approach is on the rise, centered as it is on the presumed or anticipated consequences of nanotechnologies – on health, the environment, personal freedoms, "human nature," civil liberties, global justice – in the form of a "checklist," a "code of conduct," or even a "toolbox" presumed to allow the development of a road map that would attest to satisfactory governance. To what end? For whom? These questions rarely arise. Without going into detail, we can infer that this approach is partial and lacks any real normative dimension. It reduces ethics to the evaluation of consequences (cost-benefit analysis), without specifying either how that evaluation is to be made or the worldview that governs it. The Constructive Technology Assessment approach, which has been very fruitful in the Netherlands in particular under the influence of professor Arie Rip and his team, appears to call for a diversity of actors in the entire process of nanotechnology design, and it proposes to "broaden their repertoire" by including, for example, imaginative discourses and representations. Finally, the approaches known as Public Engagement, often favored in Great Britain and in France, seem to imply the involvement of the concerned parties in public debates over nanotechnologies.

Nevertheless, it is clear that none of these procedures focuses on ethical issues as such; they are all anchored, rather, in a very broad social field that has for the most part very ill-defined contours. Moreover, as has often been noted and deplored, these procedures often entail scenarios very remote from reality (this is especially true of Constructive Technology Assessment); thus their approach often appears highly speculative, sometimes verging on science fiction.

1.2.2 *Three "candidates"*

Among all these approaches, three "candidates," or criteria, for embodying the ethics of nanotechnologies seem to have been privileged: the social acceptability (or acceptance) of a given nanotechnology, evidence of "responsibility" behind the innovation, and the safety (or harmlessness) of the innovation's design. In fact, these criteria strike me as three complementary positions within the logic of risk, with none of them being entirely satisfactory.

1.2.2.1 Social acceptability

The notion of social acceptability has both a political and a sociological dimension; the problem is that it is also very often assigned an ethical dimension. But what the "public" is apt to accept here and now must not be confused with what it *ought* to accept here and now (or elsewhere and later): that which is desirable in the ethical and political sense. Social acceptance inverts anticipatory design, it would seem, by orienting the future according to the present, rather than the other way around. Such a conception (as was the case for example in Roco and Bainbridge's well-known report) runs the risk of leading to social manipulation: anyone who understands where the levers of fear or acceptance are located could in principle use them to "market" a nano without paying much attention to democratic processes.

Beyond this basic political problem, the criterion of social acceptance is also misguided in treating the normative as essentially identical to the descriptive. In other words, "what is" is confused with "what ought to be." Now, the distinction between "what is" and "what ought to be" has constituted one of the pillars of moral philosophy since the early twentieth century; we shall return to this point in the chapter on artificial intelligence. The term *acceptability* is deceptive here; it is preferable to speak of *acceptance*, understood as what is accepted today, or on the verge of being accepted because it can be rendered acceptable.[13] This paradoxical relation between the

[13] On this question, see Nurock and Panissal 2016, "Teaching a 'Care' Approach to Nanotechnologies."

normative and the descriptive, on the one hand, and between the present and the future, on the other, inscribes ELSI from the start within the terms of a confusion that is by no means anodyne, since it amounts to eliminating what is "desirable" from an ethical stance in favor of what is "feasible," seemingly echoing the so-called Gabor's law. These distinctions can be summarized in the following table:

Moral acceptability versus social acceptance
(Nurock and Panissal 2016, 128)

Acceptability	**Acceptance**
What people should accept	What people are ready to accept
Normative	Descriptive
Process	State of facts
Active	Passive

It is a matter not only of projecting what "people" are able (or not) to accept, but also of surreptitiously targeting the possible points of resistance, which "melt away" under the discreet assaults of technologies that are presented as familiar, personalizing or individualizing, as in the process Ellul describes with the term "great innovation" in his *Technological Bluff* (Ellul 1990, 16-19). The ubiquity and the increasing "ordinariness" of nanotechnologies do not conflict, as we have seen, with the mythic epic surrounding nanos; on the contrary, these characteristics allow the technologies to be integrated subtly into individual and collective identities.

These observations reinforce my earlier emphasis on the difficulty of defining nanos and on the significations of anticipatory design. Moreover, they help justify the decision to approach the topic of nanotechnologies by way of moral and political philosophy, so as to tackle the question of acceptability authentically, that is, ethically, rather than by way of a dominant philosophy of science and technology onto which moral and political philosophy would eventually be grafted. My hypothesis is in fact that to escape from the toxic hype/hate binary, and also from the somewhat unsuitable triangle hype/hate/hope (for including reasonable hope is sometimes excessively reasonable and lacking in momentum), hope requires *care*. The term "care" refers here both to attention to the process and attention to what matters to us. As Paul Rabinow and Gaymon Bennett emphasize in *Designing Human Practices: An Experiment with Synthetic Biology* (2012), the problem may lie less in the attempt to draw up a list of authorized or prohibited actions than in the attempt to determine what sort of life we wish to lead.[14] In short, it comes down to the classic questions of what is a good life and how technologies can help us flourish as human beings. But however classic those questions may be, the traditional theoretical frameworks do not seem to be the most pertinent ones for the analysis of nanotechnologies.

This point is crucial, for it leads us to ask ourselves what we care about, and thus to evaluate issues not in terms of risks and benefits, or of authorizations and prohibitions, or even of impacts. It leads us rather to envision the developed technological universe in relation to a moral outlook, and to replace the perspective of social acceptability with that of ethical acceptability. But let us recall that social acceptability is not the only criterion under consideration; there are two others, each of which raises slightly different problems.

1.2.2.2 Responsible innovation

The social acceptability of the nanotechnologies is chiefly based on the idea of *responsible innovation*, the second criterion of the trio I have been considering. Generally speaking, the term is understood to designate an interrogative and transparent process that leads the social, scientific, industrial, and/or economic actors in a given enterprise to engage in a responsible and ethical dialogue for the purpose of ensuring the acceptability, desirability, and sustainability of the innovation, both in its processes and in its products.[15] The question of responsible innovation has at least three facets. First, can we define

[14] I thank Xavier Guchet for drawing my attention to this work.

[15] Here I am borrowing the classic definition as it is used in numerous reports and articles in the wake of René von Schomberg's work. See Schomberg 2013, "A Vision of Responsible Innovation."

the impacts that we wish to see emerging from innovative research? Then, can we orient innovations toward the results on whose desirability we have reached consensus? Finally, can we evaluate the impacts and results as good or bad outcomes on a moral scale?

As Clare Shelley-Egan and her colleagues have shown in "Devices of Responsibility: Over a Decade of Responsible Research and Innovation Initiatives for Nanotechnologies" (2018), the landscape of responsible innovation as applied to the nanotechnologies is complex. It brings together various types of strategies, ranging from strategies of anticipation in the form of "and if" to strategies of precaution, including a dynamic reflexivity (of actors and institutions) and reflection (of what is already known or covered by law between).[16] Despite the multiple tools that have been developed in the nano field in the name of responsible innovation, these authors emphasize that the scientific community in general acknowledges that, in practice, the use of those tools in the scientific or industrial development of nanotechnologies remains limited. Even so, the approaches via responsible innovation have helped shape the landscape in which these technologies develop, a situation that has made the technologies virtually impossible to circumvent.

Moreover, the notion of responsible innovation has found itself under attack in recent years by a theoretical critique that denounces in particular its naiveté.[17] In the first place, transparency and responsibility, characteristics central to the concept of responsible innovation, have proved very hard to pin down. In the second place, the famous Collingridge dilemma holds that, at the outset, one has little information about the impacts of a technology and a significant margin for maneuver, while once the technology is developed and in use the opposite is true (Collingridge 1980, *The Social Control of Technology*). In other words, among the difficulties raised by an emerging

[16] See Doubleday 2007, "The Laboratory Revisited: Academic Science and the Responsible Governance of Nanotechnology."

[17] See Blok and Lemmens 2015, "The Emerging Concept of Responsible Innovation: Three Reasons Why It Is Questionable and Calls for a Radical Transformation of the Concept of Innovation."

technology, we find, on the one hand, the weakness of the available information on its impacts in its initial phase, and, on the other hand, the weakness of possible controls – one can no longer truly backpedal – in its mature phase. Thus responsible innovation suffers from intrinsic defects that render it inoperative.

Last but not least, in the wake of feminist critiques, the concept of responsible innovation has been attacked for its blinders and for the way it has muted certain preoccupations or made certain populations invisible, owing to its failure to inquire into what matters to us and what is obscured in the conception or development of these technologies. In a polethical approach, such inquiries would be pursued.[18]

In other words, again echoing Paul Rabinow and Gaymon Bennett, we must also ask ourselves, in thinking about nanos, what types of presents and futures are open to us, and, among us, *for whom*. It is not enough to take into account the stakeholders or the economic consequences of "progress"; we must also determine the costs and by whom they will be borne. What is the invisible face of responsible innovation in the nanotechnologies? And is that face ethically acceptable?[19]

[18] The question of the most vulnerable and most invisible actors as an anchoring point of ethics also comes up, of course, in many other moral and political approaches that I lack the space to examine here. Let me simply note that what we are grappling with comes down implicitly to polethics as an articulation of the political and the ethical (along with the poetic). As Michel Deguy insists, "'poetic responsibility' is something I receive from Baudelaire. I receive it from the *Fleurs du mal* and not from the hymn to a great river in Germania. It is the final line, that of the couplet with which the hundredth flower closes, the flower whose admirable opening line reads: 'That kind heart you were jealous of, my nurse': 'What could I offer this most pious soul, / Watching her tears fall from their hollow holes?'" (Deguy 2017, 56; English translation of *Les fleurs du mal*, poem 100, by James McGowan in Baudelaire 1993, *Flowers of Evil*, 203, 205).

[19] See Kerr et al. 2018, "The Limits of Responsible Innovation: Exploring Care, Vulnerability and Precision Medicine."

1.2.2.3 "Safe by design"

Finally, after social responsibility and responsible innovation, we need to address the third central criterion in the traditional configuration of classic "nanoethics," the notion that something is "safe by design," or "benign by design." These expressions refer to the idea that a concept, a design, ensures the harmlessness and risk-free character of the product and the process alike. The term "design" is of course problematic and polysemantic: it can refer to a sketch, a drawing, or an intention; it can incorporate both an abstract projection and its material realization.[20] As Christopher Kelty has pointed out in "Beyond Implications and Applications: The Story of 'Safety by Design,'" innocuousness is defined not as a property of the materials in play but as a specter of risks.[21] Thus, according to Kelty, one will not ask which product is toxic, but rather which of its forms are the most toxic (and why). The consequence of this conception is an overturning of the scientific process: we pass from a vertical hierarchy of the sciences among themselves to a more horizontal and more collaborative approach, in order to broaden the questioning and to work at the sites of interaction between the structure and the functions of new materials.

Still, as Andrew Hale and his colleagues insist, in "Safe by Design: Where Are We Now?" (2007), the notion of design, however intuitive it may appear, does not necessarily clarify the discussion as much as might be supposed. Where does design begin? In the intention? In the conception? How does it relate to the prevention of undesired (but predictable) misuses? How much attention should be paid to undesired uses and abuses in order to make the products safer? The notion of "safety by design" incorporates, in a way, the debates in moral philosophy about the notion of "double effect," which implies a difference between the predictable and desired consequences on

[20] I thank Anthony Masure for our enlightening discussion on the notion of *design*.

[21] Kelty 2009, "Beyond Implications and Applications: The Story of 'Safety by Design.'"

the one hand, and the predictable but undesired consequences on the other – even if, as Hale and his colleagues note, not all accidents can be prevented by design.

It may be worth emphasizing here that traditionally, in moral philosophy, the notion of double effect comes up in examinations of the notion of responsibility; it marks the border between a deontological position (focused on intentions) and a consequentialist position (focused on outcomes). Now, as several authors have suggested, the central (although sometimes implicit) question regarding safety by design is the question of responsibility.[22] Taking that element into account, Ibo van de Poel and Zoe Robaey propose to shift the heart of the undertaking toward "design for the responsibility for safety." In the process, they stress that presuming that the possibility of safety is inherent in the process or the product is to presume the possibility of eliminating risks –all risks, to the point of being "idiot-proof." Since the total elimination of risks may well be impossible and even undesirable, van de Poel and Robaey recommend thinking in terms not of "responsible design" (in the sense of responsible innovation) but rather in terms of "design for responsibility," one that would guarantee users epistemic access to the technology in question, enabling them to act on the basis of foreknowledge and thus potentially to accept certain risks – in short, giving them ethical access. This suggestion, needless to say, opens up a host of new questions.

1.3 What "nanoethics" means

It is clear, then, that none of these three criteria offers a fully satisfactory pathway to defining nanoethics. Nevertheless, all of them point to a set of relevant questions that we can now explore with the aim of proposing a different ethical approach to nanotechnologies and to the NBIC cluster.

[22] See McCarthy and Kelty 2017, "Responsibility and Nanotechnology," and Poel and Robaey 2017, "Safe-by-Design: From Safety to Responsibility."

1.3.1 *Can one speak of "nanoethics"?*

What ethical problems are posed by the nanos, then, and how can we best approach them? The hypothesis that I seek to defend here is the following. Echoing the "empirical turn" and the "thing turn" that has led philosophers and researchers in science and technology studies (STS) to focus on laboratory practices and on the design of technologico-scientific objects, and to adopt an interventionist posture at the early stage of conception rather than at the later stage of implementation,[23] an *ethical turn* is necessary if we are to conceptualize the nanotechnologies not solely at their points of impact but rather beginning at the moment a new nano project is conceived and continuing throughout its development and implementation. In other words, rather than adopting a descriptive or prospective approach, it would seem more judicious to seek a normative approach, one that might be a form of ethics "by design" that could be deployed at every point in the process of innovation.

The idea of an ethical turn has already been suggested by philosophers such as Philip Brey and Peter-Paul Verbeek. In "Philosophy of Technology after the Empirical Turn" (2010), Brey stressed the importance of accompaniment "from within," via a Foucaldian governance of the (nano)technologies that would rely on mediation. The guiding hypothesis of this phenomenological approach, as Verbeek explained in "Technology Design as Experimental Ethics" (2014), is that "technologies do not simply *create* connections between users and their environment but they actively contribute to *constitute* them." Thus this approach is situated at three different levels: (1) anticipation, (2) evaluation, and (3) moralization. The third level constitutes a form of giving over of ethics to objects; it can be interpreted as a variant of the empirical turn.

This last level helps emphasize a characteristic of the ethical turn that is taking place within the philosophy of technologies: its action consists in a sort of decentering of ethics toward objects on the

[23] See for example Bensaude-Vincent 2013b, "Decentring Nanoethics toward Objects," and Guchet 2014, *Philosophie des Nanotechnologies*.

ground that human decisions are in any case always mediated by technology. Now, the very idea of transferring ethical reasoning and decision-making to objects is not unproblematic, precisely because it is hardly self-evident that it is even possible (let alone desirable) to delegate ethical reasoning and decision-making to a software program in an AI device; the issue of programming is all the more problematic when the device itself is capable of evolving or "learning," as is the case with "deep learning" in artificial intelligence.

This is why, however seductive the approach via the thing turn may be, it seems difficult to adopt it in the framework of an investigation into the ethical stakes of the NBIC technologies in general. The risk, in giving over ethics to objects, would be the risk of either situating them within the "natural," as partners,[24] or placing ourselves in a position of passivity, of the sort implied by Bill Joy's "why the future doesn't need us." It could even lead us to position ourselves within the reversal described by Henry Kissinger, in which we expect the guiding threads of philosophy to come henceforth from engineers, from machines themselves, or from the interaction between the two sets of actors. The ideal posture with respect to the empirical turn, in my view, would be to retain the idea of decentering but without transferring ethical reasoning and decision-making to objects; on the contrary, we would avoid all forms of artificializing morality, as I shall argue more fully in the conclusion of this book.

In sum, it appears that the three potential candidates for conceptualizing the ethics of nanotechnologies, or nanoethics, that have emerged from the approach via risk are all unsatisfactory; they all institute an approach that either avoids ethical questions (as with social acceptability) or are subject to critique on the ethical level. Thus we must continue the quest for a new nanoethical approach.

[24] This is suggested for example in Bensaude-Vincent 2013a, "Decentring Nanoethics toward Objects."

1.3.2 *What is a nanoethical problem?*

To pursue this quest, it seems appropriate to begin by asking what might actually constitute a "nanoethical problem." For a long time, this question has been confused with the specificity of nanoethical questions, formulated as follows: From the ethical standpoint, what is "nano-specific"? The answer to this question might seem self-evident, since, as we have seen, the Springer publishing house launched a journal in 2007 titled *NanoEthics*. Alongside bioethics, business ethics, and others, there must have been something new and different in nanos that led researchers in the field to constitute a distinct editorial collective.

Yet the very possibility of defining nanoethics was challenged from the outset, if not precluded.[25] At stake in the question, if we are to believe Fritz Allhoff in "On the Autonomy and Justification of Nanoethics" (2007) would be the autonomy of the field. One of the recurring objections to such a possibility is the impossibility of distinguishing the emergence of new ethical problems in the field of nano development.[26] One of the rare writers to focus on the distinguishing features of nanoethics, van de Poel does so in "How Should We Do Nanoethics? A Network Approach for Discerning Ethical Issues in Nanotechnology" (2008) by distinguishing two types of novelty: that of the questions raised and that of the normative concepts involved. First, he says, the newness of nanoethics may involve the fact that it raises questions that no other technology has raised before. Second, that novelty may be understood in the sense that we would need a new normative tool kit in order to conceptualize the nanotechnologies. But as van de Poel himself emphasizes, there may be a danger in focusing so intently on the newness of the nanotechnologies.[27]

[25] For a good synthesis, see Keiper 2007, "Nanoethics as a Discipline?"

[26] In addition to Allhoff 2007, see Ebbesen et al. 2006, "Ethics in Nanotechnology: Starting from Scratch?" and Bacchini 2013, "Is Nanotechnology Giving Rise to New Ethical Problems?"

[27] "The newness of ethical issues in nanotechnology might after all not be that relevant. There is in fact a danger in focusing so strongly on the newness of ethical issues, i.e., that attention is drawn away from ethical issues that are maybe not

I propose to go even further and advance the idea that focusing not only on the unprecedented character but especially on the specificity and the autonomy of nanoethics is counterproductive. It ends up concealing what may well constitute one of the truly original characteristics of the field: its relational aspect, as opposed to its autonomy. In a sense, then, seeking to define nanoethics in terms of its presumed autonomy gets the question backwards, seeking the answer precisely where it does not lie. Let me support this claim by summarizing and then analyzing three factors that strike me as essential to a characterization of nanoethics.

In the first place, nanoethics is characterized as much by second-order ethical questions (which are often the most pressing) as by first-order questions.

In the second place, nanoethics is located at a crossroads of applied ethics, for the problems it raises often come up in a specific branch of that field: the ethics of biomedicine, business, information science, environmental studies, neurology, and so on.

In the third place, nanoethics is integrated into the heart of an ethics of NBIC, and is thus not exclusively an ethics of the emerging sciences and technologies (NEST), as Tsjalling Swierstra and Arie Rip have suggested; they belong to metaethics as much as to applied ethics.

1.3.2.1 Nanoethical questions of the first and second order

One of the difficulties of discussing the ethical questions raised by the nanos lies in the fact that some of them, and not the least important, are not, a priori, ethical questions strictly speaking. One very telling example has to do with the toxicology of nanoparticles. This a particularly critical issue for at least two reasons. First, owing to their size, nanoparticles can spread widely and can pass through certain classic barriers: the skin, for example. Second, certain particles, carbon, for example, can be highly dangerous in the nano state; carbon

entirely new but nevertheless pressing from a social point of view" (van de Poel 2008, 33).

nanotubes offer the best-known example of this property. However, toxicity is not a matter for ethics, at least at first glance. One could easily argue that the question is a scientific one: once the toxicity of these particles has been discovered or plausibly suspected, one can prohibit them or restrict their use, by virtue of the precaution principle, for example. One could then argue that the key question is whether these nanoparticles can be proved to be toxic or not.

However, I suggest that this type of question can be characterized as a second-order ethical problem. Here I am borrowing the terminology of the moral psychologist Elliot Turiel, who in *The Culture of Morality: Social Development, Context, and Conflict* (2002) defines second-order moral questions as questions that are not linked at first glance to the ethical realm but that, in the context of certain transgressions, can raise ethical questions. For example, in the case of toxicity there is a difference between not worrying about exposing others to possibly toxic particles and making the decision to assume these risks for certain reasons. Toxicity would thus be a first-order scientific question and a second-order ethical question. Here we are circling back, at least partially and by another path, to the earlier discussion about safety by design, in the sense that safety is not necessarily the fundamental ethical or political question, and in the sense that deliberation remains possible here. Once again, then, it is necessary to find a way of addressing the problem other than through the lens of risk.

1.3.2.2 Nanoethics as a crossroads of applied ethics

The ethical questions posed by nanotechnologies are diverse and often of different orders. In a more or less random and certainly incomplete inventory, we can list questions concerning health or the environment (questions involving toxicology, for example, or prenatal genetic testing), transformations of the human body or mind (for example, human-machine interactions or enhancements), human rights (data protection, surveillance, privacy), justice and equity (for example, how to compensate for the biological advantages or disadvantages with which we may be endowed at birth, or who has access to resources), dual-use technologies (for example, those with military applications). However, it is not self-evident that each of these questions is solely concerned with the nano dimensions of the technologies used. Indeed, it is obvious that many of them can be combined with others: data protection in the context of prenatal genetic testing, to take just one example.

Moreover, these questions are usually situated at the intersection of several fields of applied ethics. For example, prenatal genetic testing is of concern, at a minimum, to bioethics, the ethics of research, business ethics, and the ethics of information science (involving the matter of big data). Human-machine interactions are of concern to neuroethics, the ethics of research, the ethics of information science, and so on.

One of the distinctive features of nanoethics, then, lies in its interactions with various fields of ethics and in the dynamics of the encounters. The notion of crossroads has two important aspects here. First, nanotechnology is characterized as a crossroads between various sciences (physics, chemistry, and so on) and technologies. Second, the notion of crossroads emphasizes an essential dynamism and fluidity.

Now the fact that nanoethics is a crossroads of applied ethics does not necessarily mean that it must be envisioned as a top-down ethical field in which one might choose to apply one of the great moral doctrines – whether the ethic of virtues, which is based on the question of what kind of person we wish to become, or consequentialism, which evaluates the moral value of an action on the basis of its results, or finally deontologism, which evaluates the moral value of an action on the basis of its intention. Rather than seeing nanoethics as an ethics *for* the nanotechnologies, it seems more appropriate to consider it as a process of co-construction that borrows from already-constituted elements of ethics but opens up further toward an ethics *with* the nanotechnologies, as I have suggested elsewhere from a different perspective (Nurock 2010, "Nanoethics: Ethics For, From, or With Nanotechnologies?").

1.3.3 *The nanoethical problem*

The *enabling* dimension of nanotechnologies also has an essential consequence for nanoethics. As Tsjalling Swierstra and Arie Rip emphasize in "Nano-ethics as NEST-ethics: Patterns of Moral Argumentation about New and Emerging Science and Technology" (2007), a distinguishing feature of nanotechnology is that it does not raise specific or novel questions but rather makes specific pre-existing questions more urgent. In short, the specificity of nanoethical questions would arise from a difference in degree rather than in kind (2007, 3-4). These authors propose to consider nanoethics as a part of the ethics of the new and emerging sciences and technologies (NEST). They characterize NEST ethics as a set of tropes and argumentative schemas common to the ethics of the new and emerging sciences and technologies, while acknowledging that certain questions may be specific to the nano field, especially when the nano scale is the source of particular problems or when human agency is delegated to nano apparatuses.

It seems preferable, however, to be specific about the broadening of the field proposed by Swierstra and Rip and to focus on NBIC, rather than enlarging it to all the new and emerging sciences and technologies, as in the NEST approach. In my view, it is important to envision nanoethics as part of the ethics of NBIC in two senses. First, as Jean-Pierre Dupuy has argued in "Some Pitfalls in the Philosophical Foundations of Nanoethics" (2007), the questions pertaining to nanos cannot be separated from the metaphysical program that subtends NBIC, nor from the central role played in this conglomerate by the "C," the cognitive sciences. As I see it, the nanos are integrated within a set of sciences and technologies capable of modifying not only our moral capacity and what we care about, but also what we hold to be moral or immoral. As such the nanos concern not only applied ethics, which constitute a crossroads, as we have seen, but also metaethics: nanoethics is capable of (re)shaping our ethics in the broad sense.

With this understanding in place, it appears pointless to focus on the autonomy of nanoethics. On the contrary: the fact of its relations with the other (N)BIC is doubtless one of its key features. Moreover, nanoethics itself does not belong to the order of facts; it has to belong to the order of things that concern us, that raise questions for us (to borrow Bruno Latour's distinction between "matters of fact" and "matters of concern"[28]), if we are to keep the type of modification suggested above – technological modifications of human moral capacities and standards – from intervening surreptitiously, as we shall see. Finally, it is significant that the definition of nanoethics is as problematic as the definition of nanos themselves.

1.4 Nanocare

If nanos astonish us, and if we have not yet succeeded in grasping how to conceptualize their ethics by way of the most prominent moral theories, it may well be because the nanos imply an overturning of *ethos*: they scramble our classic categories and necessitate a shift of focus from a society of risk, which has up to now subtended the ideological project of the classic nanoethics articulated with the ELSI approach, toward a conceptualization of *care*. As Joan Tronto stressed in *Le risque ou le Care?* (2012, 8), risk society is based on an approach that "discusses the 'unintended consequences' of social action from the vantage point of the 'eye of God' above society, its normative implications being assumed without being either explicitly formulated or explicitly excluded." The fear born of unpredictability is one of its driving forces. Now, as Jean-Pierre Dupuy (2007) has

[28] In Latour's vocabulary in French, the term *faits indiscutés* – facts that have not been subject to debate – generally corresponds to "matters of fact," while the term *faits disputés* – facts that have been or are being disputed – corresponds to "matters of concern." It seems to me that a "matter of concern" can also be understood as a matter inviting discussion, concern, perhaps even care, as we shall see later on. See Latour 2007, "Des faits indiscutables aux faits disputés." Maria Puig de la Bellacasa also makes the link between "matters of concern" and "care" in *Matters of Care: Speculative Ethics in More Than Human Worlds* (2017). It seems to me that Latour's thinking opens up the path toward this expansion, but it is not obvious that he himself shared this view.

convincingly shown, unpredictability is a characteristic of the metaphysical project behind NBIC, in the sense (inspired by Karl Popper [1935] 1959, *The Logic of Scientific Discovery*) of a non-falsifiable theoretical framework that orients the questions raised but also frames and limits them. To put it differently, if the anxiety that underlies risk societies grows out of a feeling of insufficient mastery, then perhaps one must assume the constitutive unpredictability of NBIC by placing the stakes somewhere other than in the question of mastery.

In an interesting way, Bruno Latour's 2011 analysis of Mary Shelley's *Frankenstein* ("Love Your Monsters") offers an illustration of what is at play here. As Latour argued, the problem is not that Frankenstein lacked mastery but that he abandoned his creature and was not concerned about what it could do or become. We can even go a little farther than Latour and suggest that Frankenstein also lacked concern and failed to look after his creature; he did not fully assume his relationship with the creature to which he had given life, nor did he assume that his creature might have a need for human relations. In short, he lacked *care*. The ambiguity, in Shelley's tale, is that the creature, coming to life, did not just *behave as if* it were alive but actually developed needs and emotions. The monster's experience was analogous to Pinocchio's, when he became a real human boy. However, to the best of my knowledge that experience is *not* the case with nanotechnologies or with any other technological objects – including the so-called emotional robots, as we shall see.

Perhaps the central issue that arises here, then, is that of *care*, understood not only in the sense of *matters of concern*, but also in the complementary ethical sense of care as expressed among others by Carol Gilligan in *In a Different Voice: Psychological Theory and Women's Development* (1993) and Joan Tronto in *Moral Boundaries: A Political Argument for an Ethic of Care* (1993). Care is not to be understood in the sense of caring for objects *as if* they were endowed with a relational capability,[29] but rather in the sense of analyzing the way in

[29] Let me stress that this is not the case with Frankenstein's creature, which is not in the mode of *as if*.

which such technological objects are inserted into a network within which *we* weave relations. As I seek to suggest in the pages that follow, this approach allows new light to be shed on some of the difficulties we have encountered up to now.

1.4.1 *Care as a challenge to the traditional boundaries*[30]

Before going back to our earlier analyses in order to show how the ethics of care can allow us to nourish and amplify them, I propose to linger a moment on what I mean by the ethics and politics of care. As I suggested in the introduction, care is neither a school nor a tradition but a polyphonic current of thought under construction roughly since the 1980s; this trend underwent a renewal in France around the turn of the century.[31] While it is difficult to supply a unified definition of care, I can at least suggest that it is distinguished by its relational dimension and by its challenge to the "view from nowhere." In its political orientation, as analyzed especially by Berenice Fisher and Joan Tronto in "Toward a Feminist Theory of Caring" (1990, 40), care is defined broadly as "a species activity that includes everything we do to maintain, continue, and repair our 'world' so that we can live in it as well as possible." A second essential characteristic of care is its rejection of the classic binaries: it is *at one and the same time* feeling and action, an entire diverse and complex set of practices that is perhaps best apprehended as a process.

"Care" is thus at once a verb (one can care *about*, care *for*, care *to*, and so on) and a substantive usually indicating a process (one can *give* care, *receive* care, *organize* care, and so on). It also entails an original conception of political responsibility, understood in terms of the way in which people take care of one another and also take care of the political structure that connects them with one another (defined in Tronto 2013 as care *with*). This approach incorporates a

30 Here I am returning to elements sketched out in an earlier article: see Nurock 2019c, "Le *care* de la nanoéthique: Repenser la question des frontières."

31 In particular in the wake of the collective work by Laugier and Paperman, eds., 2006, *Le souci des autres: Éthique et politique du care*.

renewed reflection on responsibility, a reflection to which we shall return in the conclusion.

A third essential element for reflection on the nanotechnologies can be found in Joan Tronto's insistence on the importance of considering a triple challenge to borderlines, or boundaries: those between the moral and the political spheres, between reason and feeling, and between what is private and what is public.[32]

The boundary between the moral and the political spheres camouflages the central role played by our moral sense in our political conceptions (and vice versa); blurring this boundary takes us beyond the post-Hobbesian idea that politics is limited to the defense of our own interests or to a purely self-centered vision driven by fear or by the apprehension of risk. The boundary between reason and feeling posits an impartial, universal approach, a view from nowhere, as the driving force of morality; breaking down this boundary implies acknowledging that one is always operating from a particular standpoint. Finally, the boundary between public and private frames the political sphere by separating it from everything that has to do with the private sphere, thereby devaluing the latter and making it invisible; removing this boundary thus means restoring value and visibility to what are traditionally viewed as private and personal concerns. This is an essential point, as we have seen, for a polethical approach, which stresses the articulation between ethics, politics, and poetics.

While the hypothesis according to which the nanotechnologies redistribute certain epistemological or even ontological givens is not a new one, I am suggesting that they simultaneously redistribute certain ethical and political givens and that they do this by raising questions closely resembling those raised by care. We can thus think of the three boundaries being contested in somewhat different terms. First, where care invites us to rethink the boundary between the political and the moral spheres, nanoethics needs to rethink the boundary between designers and objects. Second, where care challenges us to

[32] See Tronto 1993 and Paperman and Molinier, eds., 2013, *Contre l'indifférence des privilégiés: À quoi sert le care*.

rethink the boundary between reason and feeling, nanoethics puts the boundary between the living and the artificial back into play. Third, where care interrogates the boundary between what is public and what is private, nanoethics interrogates the boundary between what is inside and what is outside.

1.4.2 *The boundary between designers and objects*

The question of the boundary between a designer and an object has been a central one in reflections on the nanos, subtended in particular by questions of risk and control. This questioning, often mobilized through the exploration of myths such as that of the Golem or Frankenstein (often misunderstood, as we have seen), is based on the fear of seeing a product escape from its creator, or even turn against him and destroy him. It focuses on the power of designers and their presumed domination over their products. This is a wholly anthropocentric presumption, centered on the consequences or impacts desired by the designer. It implies that the product is above all conceived as a realization that must be controlled by its producer, in a way that is consistent with "responsible innovation." From the ethical standpoint, this focus brings into play the question of the double effect and the imputation of responsibility for consequences that are anticipated but not desired.

This instrumental conception of technological objects may appear surprising. It is arguably the case that such objects, if they encode some of our ethical or political behaviors, are also capable of transforming our ways of life and of reshaping our world, in particular by engaging new "forms of life," that is, life understood simultaneously as living (active) and lived in the sense of experienced (passive), the one able to influence the other significantly and vice versa – although the reversal is less obvious, for it is easier to envision situations in which the living can influence lived experience than the opposite.[33]

[33] Here we can evoke the notion of exposome, which envisions the entire set of exposures to which a living being is subjected throughout its life, whether these exposures are environmental, psycho-social, economic, or other. This concept, which appeared in the early 2000s, is now integrated into French public health laws.

As James Katz has noted, the versatility of the portable telephone, for example, can be understood as a form of *Apparatgeist:* a spirit in the machine that also shapes the spirit of its user.[34]

Moreover, three different but complementary arguments have been advanced to challenge this way of looking at technological objects. First, as we have seen and as Jean-Pierre Dupuy has shown, the presumption of mastery, or control, seems to miss completely the distinguishing features of NBIC's metaphysical project: unpredictability and emergence. Second, the mythology of the nanos is based on entities like Eric Drexler's universal assemblers, whose self-replication is by definition not controllable, as is insistently demonstrated by its repeated incursion into popular culture.[35] Third, we are well aware that certain nanoparticles, once they are introduced into the environments where they are to be used, present properties that differ from the ones their designer had anticipated.[36] For this reason, toxicologists distinguish *synthetic* identity from *biological* identity (*in vivo*) of the same nano-object.[37]

In sum, the nanos make it necessary to reopen the question of the boundary between designer and object for several reasons: nano objects cannot be reduced to instruments; they are in essence neither controllable nor predictable. It thus seems paradoxical, even absurd, to seek nanoethics in an approach undergirded by an analysis of risk society and the desire for control.

1.4.3 *The boundary between the natural and the artificial*

The second boundary that is called into question by the nano field is the one between the natural and the artificial. Bernadette Bensaude-Vincent points out that, owing to the importance of self-assemblage

[34] See Katz 1999, *Connections: Social and Cultural Studies of the Telephone in American Life*.

[35] The American series *Stargate* SG1 echoes these uncontrollable "replicators" in popular culture.

[36] See Lowry et al. 2012, "Transformations of Nanomaterials in the Environment."

[37] See Fadeel et al. 2013 "Bridge over Troubled Waters: Understanding the Synthetic and Biological Identities of Engineered Nanomaterials."

in the conception of nanotechnological objects, some of the properties of those objects are not directly predictable; thus these self-assembled objects can follow "their own nature," as is the case for example with so-called molecular motors (Bensaude-Vincent 2013a). Xavier Guchet demonstrates this in "Nature et artifice dans les nanotechnologies" (2008) through his analysis of the development of an analogue of a nanomotor made from the E. Coli flagella in the Laboratory for Analysis and Architecture of Systems in Toulouse, a "motor" that depends on the mechanisms of self-assembly in proteins. As he suggests, "there is no longer a distinction between making technological objects work and launching natural processes. The presence and the operations of these artifices, even though they have been fabricated by us, end up becoming indistinguishable from the presence and the operations of natural processes" (23).

We are thus moving away from the ideal of control and entering into a phase of cooperation with nanos, a phase in which we need to conceptualize an ethics *with* nanos. This change in attitude is subtended by a change in paradigm and in worldview. According to Bensaude-Vincent (2013a, 20), this change must also entail a change in our relation to technological objects: the object must no longer be viewed as a "slave" but rather as a kind of partner, just as our household pets are, or the natural elements – the sea, for a sailor, for example. Still, one of the stumbling blocks in this analysis, as I see it, is that it risks remaining enclosed within the natural/artificial binary. It might be more fruitful to acknowledge that these nano objects belong to neither of these two categories and that our relation to them proceeds from a different logic. If we shift the nanos toward the "natural" side, even by analogy, we lose both the specificity of nanos and the specificity of the relations between humans and "nature." We shall rediscover precisely this type of problem in the second part of this book, with the "companion" robots that are often shifted from the universe of technological games to that of household pets. I would suggest that this process is problematic not so much because it recategorizes the technological object – that is probably a necessary move – but because it recategorizes the natural while denying the specificity of our relation to it.

The blurring of this boundary underlines the difficulty we encounter in trying to grasp the nano field when we rely on our classic categories and ways of looking at the world. Moreover, as we have seen, our conception of nanos must not be based on their autonomy, because their identity varies as a function of their environment.[38] Nano objects behave differently depending on their environment, and their use exploits these relational properties. For example, objects made of nanocarbon do not manifest the same property when their allotropic form is modified, whether is a matter of graphite, graphene, nanotubes with simple or multiple walls, or nanodiamonds. These objects may indeed all be constituted by the same carbon atoms, but the way they are interconnected and the reactions they may have with their environment differ greatly.

What is the point of considering the challenge to the natural/artificial divide and the challenge to the reason/feeling divide side by side? It seems to me that, just as this latter blurring leads us to question particularist ethics (which are guilty, in the eyes of some, of not promoting impartiality, the famous view from nowhere, to which we shall return), it also leads us to question the possibility of an ethics based on principles that could be applied from on high at any time and in any place, an applied nanoethics based on major principles or universal maxims that could be superimposed on the nano field. But while such a priori ethics may not appear desirable, nanoethical reflection cannot be conceived, conversely, as an after-the-fact approach, bearing solely on the impacts. Precisely because the process is located at the heart of the forms in which nano objects are materialized, the ethical reflection must be capable of intervening within

[38] I am returning here to one of the conclusions I drew from the nano project 2E founded by the French National Research Agency, a project carried out with my colleagues Bernadette Bensaude-Vincent, Xavier Guchet, Sacha Loeve, Sophie Pellé, and Ronan le Roux. My analyses owe a great deal to the discussions we had within the team, even though we did not always reach the same conclusions. I thank all these colleagues for the very stimulating debates during this project and an earlier one, Nano-bio-ethics, in which Bernadette Bensaude-Vincent, Xavier Guchet, and Sacha Loeve participated, and also – last but not least – Catherine Larrère and Raphaël Larrère, to whom these reflections are particularly indebted.

that process in order to explore the significance of the changes involved.

1.4.4 *The boundary between the inside and the outside*

The third boundary blurred in the nano field is the one that separates what is inside from what is outside. Owing both to their particular properties and to their scale, nano objects are capable of crossing certain classic boundaries such as the skin or the blood-brain barrier; they can also integrate themselves into structures so fully that they cannot be differentiated or even detected. Thus it is entirely possible that nanoparticles may pass from the environment – the outside – to the inside of a human body, or may move in the opposite direction, whether through the skin or the respiratory system. Such migration can become extremely problematic when the particles in question are toxic, as is the case for example with carbon nanotubes,[39] which seem capable of harming even mitochondrial DNA.[40] This capacity to migrate is all the more problematic in that nanomaterials are very widely used today in industry without any possible control over their dissemination. Thus nanoparticles of silver, used for their antibacterial properties (for example, in socks) are apt to spread into the environment; the nanomaterials used in sunscreen creams can pass through the skin barrier, and so on.

Moreover, in the NBIC cluster, the use of nanomaterials permits passage from the outside to the inside in the form of implants, whose installation in the human body calls into question both the definition of prostheses and the definition of privacy – all the more so in that these nanos are easily forgotten and often imperceptible. If one is unaware of harboring such nanomaterials, how can they be taken into account?

[39] See Helland et al. 2007, "Reviewing the Environmental and Human Health Knowledge Base of Carbon Nanotubes."

[40] See Li et al. 2007, "Cardiovascular Effects of Pulmonary Exposure to Single-Wall Carbon Nanotubes."

Not only do the nanos call the aforementioned boundaries into question, but the three typologies are also mutually porous. The end of the natural/artificial distinction is inseparable from that of the designer/object distinction: it is as though the nanomotor, neither natural nor artificial, were neither object nor subject but its own designer. Similarly, the relational properties of the nanos, essential for blurring the categories of natural and artificial, are apt to play a determining role when the difference between inside and outside is blurred, by changing behavior when they change milieu. All of these distinctions thus seem to be interwoven and mutually reinforcing.

Finally, nanos modify temporality just as they modify space. As we have seen, the nano field calls into question the classic framework of time when it is situated, as is generally the case, under the auspices of an anticipatory logic assumed to be self-realizing.[41]

1.4.5 *Perspectives*

It should be clear by now that the similarity between the three boundaries called into question by nanoethics and those challenged by the ethics of care go beyond mere analogy, in that they share three essential propositions: (1) It does not suffice to think solely in terms of impact, safety, or risk. (2) There is no such thing as a "view from nowhere." (3) The inside and the outside are porous.

Conceptualizing nanoethics and care together thus makes it possible not only to bring to light these shifting boundaries and to escape from the classic binary divides, but also to stress the necessity of extending our ethical and political reflections beyond the questions of risk or impact alone, and beyond the question of control, in order to try to reflect on how best to deal with unpredictability, a feature that is consubstantial with NBIC. In other words, a nanoethical reflection based on care foregrounds three fundamental stakes. First, it is not enough to be concerned about impacts, no matter how significant they may be. If nanoethics is to be developed as an ethics

[41] See Mody 2004, "Small, but Determined: Technological Determinism in Nanoscience."

with nanos, the entire process of developing nano objects must be taken into account; we might call the result an ethics *by design*. Second, nanoethical reflection cannot be undertaken from a "universal" point of view or a "view from nowhere"; it must be prepared to take the dynamics of specific relationships into account. Third, this reflection must consider the porosity of the contexts in which nano objects are intended to function and the passages or transitions in which they are apt to be engaged. These upheavals are summarized in the table below:

How to rethink boundaries and categories?
(Nurock 2019, 149-165)

Care	**Nanoethics**	**General characteristics**
Boundary between the political and the moral	Boundary between the designer and the object	Thinking in terms of safety or impacts does not suffice.
Boundary between reason and feeling	Boundary between the living and the artificial	The universal point of view –the view from nowhere – does not suffice.
Boundary between the private and the public	Boundary between the inside and the outside	The inside/outside and private/public boundaries must be understood as porous.

1.5 Nanoethics and relational responsibility

As we have seen, one of the fundamental ethical and political difficulties encountered in the nano field is that of responsibility. Precisely because of its multidisciplinary character, but also because of its enabling character, and because of the fact that nanoparticles can pass through barriers that are customarily held to be secure, it is extremely difficult to assign responsibility in the nano field in any conclusive way, given that nano objects are by definition circulatory and proteiform, in the sense that they change properties in changing environments.

My hypothesis here is that what is problematic is not so much the *assignment* of responsibility but the way in which we *conceive* of responsibility in the nano field. In other words, perhaps the central question one ought to raise is not whether it is possible to impute responsibility to a human agent but rather how to envision the network of relations that involve human agents (or patients) within the nano field. It is on the basis of this non-substantive configuration of responsibility that Joan Tronto (2012) – in the wake of Soran Reader and Iris Young, as we shall see later on – proposes to envision relational responsibility.

This conception, which is still in the process of elaboration, views responsibility not as a substantive form that would derive, for example, from the application of moral principles, or as a temporal vector going from cause to effect, but rather as a process, an engagement that would be responsive as well as responsible. In the nano field, this relational responsibility can be viewed from two distinct and complementary angles.

First, each actor in the nano field is perceived as involved in a network and as having a share in responsibility for the conception, distribution, regulation, and utilization of the nano field. Responsibility can then be envisioned in the form of a situated canvas, allowing a change of focus every time one's situation of responsibility changes, but in which each actor is more or less in relationship with the others. Here, then we have to conceptualize not a static but rather a dynamic network, one that truly involves a relationship, a change in perspective every time one's focal point changes (somewhat like choral works in the artistic field, which embrace various viewpoints to constitute a weft). Depending on whether one is a consumer, a designer, a political decision-maker, an industrialist, a citizen, and so on (positions that can be occupied in a non-exclusive manner), the viewpoint and the dynamics change.

Moreover, the fact that such circulation comes into play also makes it possible for one to try to put oneself in the place of another actor in order to attempt a moral game of "musical chairs"; according to Lawrence Kohlberg in *The Philosophy of Moral Development* (1981), the most influential moral psychologist of the second half of the twentieth century, this approach constitutes the moral methodology par excellence. However, for Kohlberg, the goal is to reach an "impartial" viewpoint – the famous "view from nowhere," in his terms, or the "eye of God" to which Tronto refers – that encompasses all the others, whereas I am suggesting that the goal is rather to succeed in assuming the (potentially destabilizing) dynamics of the process. This approach disallows exempting oneself from responsibility by arguing that one is not directly responsible, not the only one responsible, or unable to impute responsibility in a mode of legible causes and consequences. By passing from risk to care, one becomes able to conceptualize a relational responsibility in the nano field.

In the second place, as Pascale Molinier and Patricia Paperman have noted, the outlook of the researchers in nanotechnologies whom they met in a laboratory in Toulouse was positioned in many respects within a relational framework in which ethical interrogation was omnipresent.[42] This was unquestionably a "model" team whose undertaking was very open to collective debate and to ethical reflection, allowing a kind of collective "moral anguish" to be expressed – contrary to what the notion of anguish might lead us to suppose.

In *Le Care Monde* (2018), Pascale Molinier proposed to identify the experience of "moral anguish" as a sign that one's habitual reference points or concepts were being challenged or overturned. But she also saw moral anguish as signaling the possibility of producing common images, because reality is both hard to grasp and hard to communicate, hard to share. I should add that it is probably not insignificant that moral anguish has found acute expression in the context of research in nanotechnologies. Because the traditional concepts are particularly blurred, mangled, even reconfigured in the nano field, as we have seen, moral anguish is ripe for development, and it drives a dynamic effort to find new ways of configuring ethical propositions

[42] The collective project called Nanocare was financed by the mission supporting interdisciplinarity of France's National Center for Scientific Research (CNRS), in which I was the principal investigator in the early 2010s.

so as to try to respond to them more appropriately. Moral anguish is thus potentially a constructive force.

The question of relational responsibility can also be raised in other fields reconfigured by nanomaterials, and in situations seemingly more trivial than those taking place in a laboratory. For example, the antibacterial use of certain nanomaterials has led to their use in products of daily life such as "innovative textiles" (the famous socks that 'don't smell'). The secondary effects of their uses on the human body (especially as the products break down over time) or on the environment (in the processes of fabrication, use, and recycling) are not yet very well known. The environmental stakes are particularly significant here, and they raise concerns about new interactions in which we might involuntarily engage with nanomaterials by integrating them into the environment, whether directly or remotely, within a cycle that is not limited to their voluntary use alone.

Moreover, if we shift our focus to the field of medicine, we find that nanomedicine is especially propitious for the constitution of relational responsibility: early diagnosis, for example, is among the advances facilitated by nanomedicine, but it risks situating us definitively in the position of patients, and perhaps for generations, as we shall soon see. With such developments, our conception of what it means to be alive, our notions of health and illness, and the responsibility we may have toward future generations are all being transformed.

CHAPTER 2

Cybergenetics

Medicine is one of the privileged fields for the application of nanotechnologies, so much so that the term "nanomedicine" has emerged to designate their alliance. The multiple uses of nanotechnologies in medicine range across targeted medications, diagnostics, imaging, and surgical tools. These technologies are also a vector in what is called personalized medicine; their use can lead to cost reductions as well as unprecedented approaches to challenging problems.

The specialized field of cybergenetics has been developed in the context of personalized medicine, at the intersection between genetics and information technology. The current chapter will focus on the flourishing of cybergenetics in general, and on the subfield known as "recreational" cybergenetics in particular. This second case study will allow us to see the bridge between certain types of biotechnologies and information science, and thus to shift from the N to the B of NBIC, occasionally touching on neighboring disciplines, especially the I, while noting the continuity of the problems that arise.

The surge in cybergenetics is inseparable from a cluster of recent transformations involving relations of care, relations between science and technology, and even relations between individuals and social networks. And it is equally inseparable from the profound modifications that have affected the way we conceptualize identity, relationships, and politics. Finally, on the historical level, this rise is inseparable from the celebrated Human Genome Project, which marks a turning point in our relation to genomics and has brought us into a "postgenomic" era, as the American sociologist Jenny Reardon, for one, suggests in her book *The Post Genomic Condition* 2017).

The "postgenomic" condition in which we have found ourselves since the 2010s, Reardon argues, implies that we are not only rushing full speed ahead in the race to decode the genome that marked the turn of the 1980s, but also that we are turning toward "the question of meaning – the question of the uses, significance, and value of the human genome sequence" Reardon 2017, 2). In addition, at a time when we are being bombarded more relentlessly than ever with a constant stream of solicitations, information, and fake news, we have to decide where we want to focus our time, energy, and attention.

I propose to take Reardon's propositions quite seriously, approaching them from a philosophical rather than a sociological standpoint, and more precisely from the standpoint of moral and political philosophy.[1] I seek to show that awareness of the vast experimentation currently taking place in cybergenetics – not only biological but also social, political, and ethical experimentation – has become indispensable; analyzing it from an inseparably ethical and political viewpoint will allow me to demonstrate what profound modifications it implies on both the collective and the individual levels. These modifications appear all the more overwhelming when we look at them more broadly within the context of NBIC; they make it increasingly urgent to review and renew the ethics that analyzes them.

The investigation that follows will thus be organized, as in the previous discussion of nanotechnologies, around four main arguments. First, the conditions under which cybergenetics has emerged (especially in the recreational form) make the field sometimes difficult to pin down, for it blurs certain customary boundaries. Second,

[1] My aim here is to develop the notion of care to which Reardon implicitly and even explicitly refers when she writes, for example: "It is my contention that in these times as we rightly turn our attention to correcting falsehood, we must also attend to the problem of deciding which elements of this troubled world-in-need deserve our all-too-limited energies. *Which should be matters for our care and concern?*" (2017, 5; emphasis added).

its development is based on a mythology that relies on self-fulfilling prophecies and facilitates its intrusion into every aspect of our daily lives. Third, the problems posed by this new field reveal the need to go beyond the exclusively risk-based approach that is widespread in the medical realm in order to conceptualize the reconfigurations that are at work in the field. Fourth, this undertaking requires new ways of formulating the questions that arise around the new cybergenetic technologies, and new ways of conceptualizing the ethics of these technologies in particular, and the ethics of NBIC in general, in terms of relations rather than autonomy, most notably in response to the ethics and politics of care.

2.1 The emergence and development of cybergenetics: a blurred definition

The conditions under which cybergenetics emerged may seem fairly clear at first glance, but they reveal a field whose definition is far from self-evident and whose intersections with other fields create zones of complexity.

2.1.1 *The Human Genome Project*

To understand the stakes of cybergenetics, one has to begin by retracing the principal steps that made it possible, starting with the Human Genome Project (HGP). As Kate O'Riordan has noted in *The Genome Incorporated: Constructing Biodigital Identity* (2010), the Human Genome Project was an event in the digital as well as the biological realm; it would not have been possible in the pre-computer era. It is symptomatic that the project was born after the space race, on the occasion of a budgetary shift from the macrocosm to the microcosm, as it were.[2] The University of California at Santa Cruz had been granted thirty million dollars to build a telescope; when that project

[2] Here again we see the theme of conquering space, as with the Sputnik effect we encountered in the introduction. This theme, often found in the background of NBIC, appears quite concretely in the case of cybergenetics.

was dropped, UCSC's Chancellor Robert Sinsheimer, a molecular biologist, sought another way to spend the grant money.

As Lisa Gannett reports in "The Human Genome Project" (2016), the origins of the project go back to the 1980s, when three researchers, Robert Sinsheimer, Renato Dulbecco, and Charles DeLisi, each acting independently, began to take seriously the possibility of sequencing the entire human genome.[3] While their idea was supported by prominent biologists, most notably James Watson, winner of the 1962 Nobel Prize in medicine and one of the co-discoverers of the double helix structure in DNA, the project was initially viewed with caution by the scientific community, whose members were concerned not only about the difficulties of the undertaking but also about the fact that it would drain both human and material resources away from other research fields. In addition, sequencing the genome would bring only a limited amount of information: it would not contribute to the understanding of the way genes work. Thus human genome sequencing struck some as a form of "hype," a term we have already encountered with reference to nanotechnologies – and the connection is probably not accidental.

Several committees were set up to examine the desirability and especially the feasibility of the project. By the late 1980s, it had become clear that the project of mapping and sequencing the entire human genome had multiple dimensions and, at an estimated cost of three billion dollars, it would be significantly bigger than the Manhattan Project that had underwritten the development of the atomic bomb, or the Apollo Project – the "giant leap for mankind" – that had put men on the moon. Its initial goal was primarily to identify all the genes (estimated at about a hundred thousand) and nucleotides (around three billion) of the human genome, but also to develop tools for analyzing the data.

[3] The pages that follow draw heavily on Lisa Gannett's publications and on a series of lectures she gave in Paris in 2019. I am very grateful for the discussions we had and the clarifications she offered me during her time in Paris.

Interestingly, concern for the project's legal, social, and ethical impacts – the famous ELSI questions we saw in the preceding chapter seem to have come directly from this aspect of the HGP[4] – was a factor in the program from the start; these questions were assigned to a working group begun in late 1989 and led by Nancy Wexler, who had worked on the genetic causes of Huntington's Disease. Anchored in an ELSI approach based on questions about risks and control, this group identified four fields to be prioritized: the quality of and access to genetic testing, the fair use of genetic information by employers and insurance companies, the confidentiality of genetic information, and the education of the public.

In Europe, research on the human genome was also advancing in the late 1980s. In France, for example, there was a huge fund-raising campaign in the form of a "*téléthon*" (emulating an American model that originated in the 1960s) for the purpose of financing genetic research; the first broadcast, on the French public channel "France Télévisions" in 1987, brought in some 27 million euros, three times the amount expected; this tells us something about public enthusiasm for genetic research, which has scarcely wavered since. The resources generated by the *téléthon* funded the creation of the "Généthon" project, which focused initially on mapping the genome; it moved on in the late 1990s to specialize in gene therapy. What we saw in France was part of a worldwide dynamic (in the wealthy countries) that mobilized significant public and private resources and aroused considerable interest among politicians and the general public alike.

It was as though, after the conquest of space, we had to complete the conquest of our biological identity. As the Smithsonian Museum of Natural History, in partnership with the U. S. National Institute of Health, declared on its website, it came down to nothing less than "unlocking life's code."[5] As we have seen, the collaboration between nanotechologies and cryogenics had the conquest of space as its

[4] See Bennett-Woods 2008, *Nanotechnology: Ethics and Society*, 62-63.
[5] https://unlockinglifescode.org/.

backdrop.[6] This is certainly not a matter of chance, given that the conquest of space was an important historical achievement of the era, but it is nevertheless the case that this background imposed certain directions, more or less deliberately, on the development of cybergenetics.

The end of the 1990s saw a veritable race to decode the genome on the part of the private sector. The International Human Genome Sequencing Consortium was joined by Celera Genomics Corporation, founded by the well-known entrepreneur Craig Venter, who made his own genome public; for a time, Venter cherished the hope of patenting certain genes, especially those connected with specific types of breast cancer.[7] It was doubtless no accident that Venter's company, which was initially called Applied Biosystems (belonging to the Perkin-Elmer Corporation), was renamed Celera (from the Latin verb *celerare* meaning to hurry, to accelerate[8]), or that his second book is called *Life at the Speed of Light* (with an obvious reference to space in its title).[9]

Victory in the race was triumphally announced by Bill Clinton at the White House on June 26, 2000, with Craig Venter and Francis Collins, the scientific patron of the Human Genome Project, at his side. However, the victorious project had not (yet) been subjected – as is normally the case in scientific circles – to peer review for publication in professional journals; thus some saw the announcement as premature, more political or economic than scientific in its import. And indeed, with cybergenetics as with nanotechnologies in general,

[6] Cryogenics often goes hand in hand with anguish not only about the end of individual lives but of the entire human species on the planet, and thus the need to be able to look to space to find other habitable planets. It is easy to find echoes of these concerns in popular culture.

[7] See Murry 1999, "Owning Genes: Disputes Involving DNA Sequence Patents."

[8] Unlike *scelera*, which is the Latin word for crimes or villains. It is as though the "natural" outlet of these technologies were to catch villains, as we shall see at the end of this chapter with the analysis of cybergenopanoptics.

[9] See Venter 2013, *Life at the Speed of Light: From the Double Helix to the Dawn of Digital Life*.

the constitution and development of the field has turned out to be influenced at least as much by politics and economics as by science.

Moreover, as Lisa Gannett notes, the staging of the HGP proclamation had particular significance in the American historical and cultural context. For Bill Clinton explicitly compared the mapping and decoding of the genome to the nineteenth-century mapping of the American West, claiming that the map produced by the Lewis and Clark expedition "defined the contours and forever expanded the frontiers of our continent and our imagination" (Clinton 2000, also Gannett 2016).

As Gannett emphasizes, the tone was set: confidence in science, aspiration to systematic knowledge that should benefit all humanity. The central idea was that one of the frontiers of knowledge had been shifted and that a new horizon was open to conquest, in political and economic as well as scientific terms. Just as in the conquest of the American West, the implications of this dynamic, this rapid (ac*Celera*ted) race ahead, and especially its impact on those it left by the wayside or simply made invisible (in the historical context, most notably Native Americans), remained – and still remain – largely unexamined.

As we saw with the debates around nanoethics, a debate has developed over the unprecedented ethical questions – we might call them *genethical* questions – raised by the advances in genomics. The classic response to such questions is that there are "no new problems," just exacerbations of the old familiar ones[10]: the problems would merely increase in degree and complexity, as medical ethics specialist George Annas suggested in 1990.[11] Nevertheless, as we shall see, the appearance of recreational cybergenetics clearly reshuffled the cards, or at

[10] See Cooper 1994, *The Human Genome Project: Deciphering the Blueprint of Heredity*.

[11] "There are probably no *unique* issues raised by the Human Genome Initiative. On the other hand, this project raises *all* of the issues in a much more focused manner (certainly a difference in degree if not in kind) and the fact that all of these issues are implicated in the project may itself make the project societally unique" (Annas 1990, "Mapping the Human Genome and the Meaning of Monster

least shed new light on them, and has led to profound ethical and political upheavals.

2.1.2 *The beginnings of cybergenetics: from the personal genome project to "recreational" genetics*

What first appeared to be a collective and humanist enterprise – as indicated by the term "Human" in the project's title – very quickly became a *personal* enterprise. Here it is essential to recall the discourse of Craig Venter, who affirmed that his desire to develop genetic research was connected to his personal family history, including the fact that his son was afflicted with schizophrenia. Similarly, Anne Wojcicki, one of the founders of 23andMe, the best-known firm in the field of "recreational" genetics, stressed that personal details – the fact that she was a Bloom syndrome carrier and had a heightened risk of breast cancer – were among her motives for creating the company.[12]

The Human Genome Project was thus paralleled from the outset with personal – even personalized – projects. A number of scientists, whose genomes Venter made public, moreover, started a movement that quickly expanded. In 2005, Harvard geneticist George Church launched the Personal Genome Project (PGP),[13] which he conceived as something like the Wikipedia of the HGP, to which anyone could contribute by publishing their own genome online.[14]

Initially a North American endeavor, the Personal Genome Project spread elsewhere starting in the 2010s: Denmark in 2011, Canada in 2012, Great Britain in 2013, Austria in 2014, South Korea in 2014, and China in 2017. This project had two distinguishing features: not only did it make public, online, the genomes of non-anonymous volunteers, but in addition to the full genotype – the sequence

Mythology," Cited in Gannett 2016). See also Annas and Elias, eds., 1992, *Gene Mapping: Using Law and Ethics as Guide*.

[12] See for example Wolfe 2014, "Anne Wojcicki's Quest for Better Health Care: The 23andMe CEO on the Promise of Genetics and the Future of Health Care."

[13] For project details, see https://pgp.med.harvard.edu/.

[14] See Frank 2011, *My Beautiful Genome: Exposing our Genetic Future, One Quirk at a Time*.

of the 23 pairs of chromosomes – it published the phenotype of each individual, including non-genetic medical results, imaging, environmental elements where available, and so on.

The question of responsibility and informed consent on the part of the volunteers was viewed by critics from the outset as the project's Achilles' heel. However, the Harvard Medical School, in promoting the project, managed to make the potential weakness a strong point by turning to a group of ten prestigious volunteers, from scientific fields for the most part, who agreed to be the first "guinea pigs," known as the PGP 10. In addition to George Church, they included scientific figures from academia (such as John Halamka from the Harvard Medical School, and Misha Angrist from the Duke Institute for Genome Science and Policy) and from the private sector (such as Keith Batchelder from Genomic Healthcare Strategies), but also someone from outside the field of genetics, Steven Pinker, a Harvard professor and researcher known for his work in cognitive science (the C of the NBIC…).[15] In a long article titled "My Genome, My Self" published in 2009 in the *New York Times Magazine* (duly relayed on the blogs of companies such as Helix[16] and 23andMe[17]), Pinker explained what had made him decide to participate, and what he saw as the prospects opened up by cybergenetics, especially in terms of psychological screening. He stressed the need to be able to choose what one was being tested for, and he revealed why he himself chose not to be informed of any results bearing on diseases such as Alzheimer's. Beyond its defense of scientific determinism (a position entirely in keeping with Pinker's usual positions), this text is striking both for the way it sets forth the possibility, even the necessity, of obtaining "scientific" knowledge about oneself, and the fact that one might wish not to do so.

[15] Steven Pinker is a fervent defender of evolutionary psychology and of the computational theory of mind; he stresses the analogy between thinking and computation or data management.

[16] https://helix.com.

[17] https://blog.23andme.com/23andme-research/steven-pinker-on-personal-genomics/.

The Harvard Medical School's intent was not only to set an example and incite other people to emulate these "models," but also to show that figures particularly well informed about genomic research could be eager to make their genome public. Nevertheless, we might note that no ethicist, indeed no philosopher of any sort, was included on that panel. It was as though a perfectly satisfactory understanding of the scientific dimension of the issue sufficed to analyze its ethical and political stakes. It was as if, echoing the reflection by Henry Kissinger cited in the introduction to this book, instead of raising the question of the desirability of a given technology in advance, the public had given scientists the role of validating that desirability. The appropriateness of the Harvard Medical School's undertaking can certainly be questioned, but when we see that a mere handful of volunteers were followed by the more than 10,000 participants found in the Personal Genome Project today, we can hardly doubt its effectiveness.

It is worth noting that Steven Pinker, the tenth volunteer in the PGP10, emphasized that the project was consistent with the overall aim of the cognitive sciences in its ameliorative dimension, thus connecting with the C of NBIC. Some recent developments that considerably amplify this aspect should therefore not be surprising. Starting in 2017, the Personal Genome Project has been associated with a developer of digital apps designed to improve cognitive performance (the company's enlightening name is Lumos Labs), in order to bring genomic and cognitive data into contact. The central idea is to correlate performances on memory tests and the time taken to complete them with variations in the genomes of the participants, with the particular goal of developing experimental models of brain degeneration. Thus the PGP is cybergenetic in more than one respect: not only because the personal data of its participants are publicly available on line, but also because that data is now inextricably connected with the development of applications.

Still, however interesting and symptomatic the PGP may be, the "official" birth date of cybergenetics is generally not tied to the wave of personalization but rather to the modification of an economic model, which is also the model for what is called *recreational genetics*. While this form of genetics is not representative of all cybergenetics, since it is limited to personal, "recreational" uses of genetic testing unconnected with the medical field, it is nevertheless so symptomatic of some of the field's characteristics (indeed, sometimes in exaggerated forms) that it is an especially interesting case to study.

It is not my intention, in these pages, to minimize the importance and potential fruitfulness of genetics for advances in medical research. Quite the opposite: because I am convinced of the potentially life-saving value of genetics, it seems essential to work on establishing guardrails that would allow the field to develop fully while preventing it from being discredited by practices that are dubious or even dangerous on the ethical level. Whereas in the case of nanotechnologies it is hard to be unreservedly enthusiastic, in the case of cybergenetics it is a wholehearted belief in the medical progress permitted by genetics and a desire to support its advancement that makes the analysis useful, even necessary. In fact, one of the risks is that the recreational uses of cybergenetics can easily be overgeneralized, owing to their broad scope, and this may lead to mistrust or even fear of the field itself. Thus I have chosen to focus here on recreational cybergenetics in the hope that an analysis of the counter-example will help reinforce its non-recreational form.

The birth of cybergenetics is generally traced to George Church's 2007 launching of the Knome Company, which offered a *complete* genetic sequencing for $350,000. The cost has gone down dramatically since then, and partial sequencing has rapidly become affordable to almost everyone.[18] Cybergenetics has become inseparable from the same "low-cost" mantra that we saw with the nanotechnologies, a fact that is hardly surprising. Still as we shall see, its economic model implies a certain number of mutations.

[18] Complete sequencing is now available for less than $1,000.

While cybergenetics is unquestionably anchored in the dynamics between personalization and the "low-cost" economy, it would be a mistake to see these factors as the only ones responsible for its emergence. We can discern at least five other significant causal factors.

First, the paradoxically weak presence of genetics in classical medicine at the turn of the century, despite the considerable scientific advances that had been made in the field, left room for the development of new actors. This evolution was marked above all by the *reconfiguration of relations at the heart of the medical and paramedical fields* brought about by the Internet, from websites to patient forums[19] and the development of e-medicine. However, a significant component of cybergenetics has nothing to do with e-medicine, and it has played on the ambiguity and the space left vacant by that reconfiguration.

Second, the costs of sequencing declined significantly (an element that we also find in the development of nanos), with testing kits available for less than $30. In other words, recreational testing now costs no more than a theater or concert ticket or a theme park entry pass.

Third, a number of new companies with solid foundations have come to the fore, for example deCODE, an Icelandic company founded by the controversial Kari Stefansson, who profited from the homogeneity of the Icelandic gene pool,[20] and 23andMe, one of whose three founders, Anne Wojcicki, was at the time married to Sergey Brin, a co-founder of Google.[21]

Fourth, the development of the Internet, social media, and apps has led to multiple forms of mutation of what might have seemed at first sight simply a spinoff, a derivative version of a paramedical product, and it has opened the door to an original configuration of

[19] One well-known site in France, *doctissimo*, was cofounded by the transhumanist Laurent Alexandre.

[20] See https://www.decode.com.company/.

[21] They later divorced, but both maintain that the links between the two companies remain intact.

cybergenetics. In particular, the development of *social networks and blogs* has facilitated the narration of personal and family stories, whether in highly individualized or more structured ways. This dimension is taken to an extreme in Mormon culture, which traditionally grants high importance to two interrelated elements: on the one hand personal narration and journaling, intended to connect individuals horizontally to their roles within the religious community, and on the other hand genealogy, which connects individuals vertically with their forebears.[22] Unsurprisingly, digital resources have proved invaluable in this context; the Mormons are known for their highly organized genealogical data banks, but also for something that (re)configures the world in a different way, the digital staging of a prototypical family framework via blogs. "Mommy blogs" depicting an idyllic family life in which all members play their roles harmoniously have found a special niche in the blogosphere; indeed, there has been much discussion about why these blogs are found so fascinating, including for a feminist readership.[23]

Fifth, the development of *apps* that put access to genetic information literally in our hands: as we shall soon see, these have added a quasi-ludic aspect to the phenomenon and a quite addictive dimension as well.

2.2 Cybergenetics and "recreational genetics"

2.2.1 *What is recreational cybergenetics?*

Not simply a derivative, a spinoff from the medical or paramedical field, recreational cybergenetics in its diverse forms occupies a more complex position.

[22] On this topic, see Feller 2007, "Media as Compromise: A Cultural History of Mormonism and New Communication Technology in Twentieth-Century America," and Avance 2015, "Constructing Religion in the Digital Age: The Internet and Modern Mormon Identities."

[23] See for example Matchar 2011, "Why I Can't Stop Reading Mormon Housewife Blogs: I'm a Young Feminist Atheist Who Can't Bake a Cupcake. Why Am I Addicted to the Shiny, Happy Lives of These Women?"

In the broad sense, cybergenetics is located at the intersection between the digital sector and genetics, thus covering quite a broad domain. It also refers more or less implicitly to an idea that was already in Norbert Wiener's mind when he proposed the term cybernetics: the idea that the human body is an integral part of the information system in terms of incoming and outgoing messages.[24]

At the heart of this field, what has come to be called *recreational genetics* occupies a particular and very important place in terms of both development and significance, for it orients a large part of the social reconfigurations I propose to examine.

Recreational genetics can be defined as genetics developed outside of the medical field proper, under the control of a welcoming system intended to serve recreational purposes. But this characterization remains quite vague and tells us very little about the way the field is configured. Thus we may prefer the characterization of *Direct to Consumer* genetics, or DTC (expanded on occasion to DTCGT, or Direct to Consumer Genetic Testing). This label stresses the absence of mediation between the genetic testing company and the person who orders a test, characterized as a consumer – hardly an anodyne label. We find the same insistence on personalization that we found in the Personal Genome Project; and indeed, one of the major DTC companies is called 23andM*e*.

In "Direct-to-Consumer Testing 2.0: Emerging Models of Direct-to-Consumer Testing" (2018). Megan A. Allyse and her colleagues point out that recreational cybergenetics differs from traditional medical genetics in numerous respects, including the person who initiates the test, the type of regulations that apply, the person who interprets the test, the norms of quality control, the goals, and the cost. The authors recapitulate these factors in a synthetic table:

[24] See Harris et al., *Cybergenetics: Health, Genetics, and New Media*, 3.

TABLE 1. Key Points of Comparison Between Direct-to-Consumer Genetic Testing and "Traditional" Medical Testing

Key aspect	Direct-to-consumer testing	Traditional medical testing
Testing initiation	Patient initiates test	Health care worker initiates test
Source of regulation	Companies are regulated as consumer products	Health care systems are regulated by industry-specific rules
Information control	Patient controls and manages genetic information	Health care system controls and manages information
Data interpretation source	Patient chooses interpreter of genetic information	Agents of the health care system interpret genetic information
Quality control	Test quality is largely unregulated	Regulations and quality control systems in place to assess test quality
End use	Interpretation can be for multiple purposes, eg, ancestry, paternity, and health	Interpretation facilitates medical management
Pricing model	Intense competition may drive down product price	Product price is tied to the health care payment model
Return of information	Information may come without clinical support or counseling services	Information is delivered by health care professionals
Data interpretation regulation	Information interpreters are typically not accredited	Information interpreters are licensed or accredited
Secondary use	For-profit selling of data for secondary use	Secondary use largely limited to health care research purposes

Source: Megan A. Allyse et al. 2018

The history of the DTCs sheds light on the landscape of these companies as it is constituted today. As Stuart Hogarth and Paula Saukko pointed out in "A Market in the Making: The Past, Present and Future of Direct-to-Consumer Genomics" (2017), the prehistory of the DTC companies was anchored in a genetics-by-correspondence undertaking initiated in 1996 by a British start-up called University Diagnostics, and in the press, radio, and television advertising put out by the American Genetics and IVF Institute for genetic tests focused on the BRCA genes, known to be responsible for breast cancers. This prehistory of cybergenetics is thus rooted not so much in digital developments as in the development of direct-to-consumer advertising and selling by mail.

Around the turn of the century, the pre-existing model began to be restructured by the Internet: shopping by mail morphed into online shopping. This first wave of DTC cybergenetics allowed DTC companies to flourish, initially concentrating on markets involving well-being and nutrition. The Sciona Company even managed, briefly, to sell its test in *The Body Shop* stores, before pressure from the NGO Genewatch brought that practice to an end, on the grounds that Sciona's advertising could give people the impression that their "good genes" might allow them to ignore important health issues.[25]

Alongside the DTC companies focused on well-being and nutrition, another set of companies began to promote genealogical genetic tests, allowing people to find out where their ancestors had come from and to discover potential family relationships with contemporaries; paternity tests were also on offer. Family Tree DNA was founded in 2000 to make genealogical tests available to Americans (and especially Jewish Americans). The company grew rapidly: buoyed by particular proclivities of American culture and society, its estimated worth went from $2.6 million in 2004 to $12.2 million in 2006.

The example of DTC genealogy companies is a somewhat special case, symptomatic of the significance of contemporary cybergenetics within American culture. One aspect of this significance can be located at the meeting point between a particular history and credo, that of Mormons, and a more general history encompassing historical, social, and political elements. We shall come back to this point, but for now I propose to stress two significant elements. Among the Mormons, first of all, tracing one's genealogy is actually an obligation, linked to the belief that it is possible to sanctify and thereby "save" ancestors who had not adhered to the faith of their descendants. The *Book of Mormon* indicates a certain number of migratory paths that believers try to demonstrate with the help of DNA analyses.[26] And these preoccupations intersect with the Internet "mommy blogs" that are the online incarnation of a certain segment of this population.

More generally speaking, genealogy is an American passion. This is not a surprising development in a nation consisting essentially of immigrants (voluntary or not) in which integration goes hand in hand with an exaltation of one's origins. Furthermore, in a country

[25] For more details, see http://www.genewatch.org/sub-425647 (accessed August 10, 2019). Sciona had to change its location in the U.S. and definitively stopped selling tests in 2009.

[26] See for example "Book of Mormon and DNA Studies," n.d.

marked by racism but also by positive discrimination or "affirmative action," genealogical tests can serve different purposes; they are used as much by the racist and nationalist right as by figures of antiracism such as Oprah Winfrey, or by Native Americans seeking to prove their membership in a specific indigenous Nation; thus these tests may allow access to certain prerogatives in the context of affirmative action policies.

After 2005, with a pronounced acceleration in 2007-2008, cybergenetics' second wave began to appear, with a reorientation of what was offered and a reinforcement of links with companies specializing in the production of tests and digital industries. This was most notably the case for companies like 23andMe or Navigenics, connected respectively to the digital giants Google and Microsoft, and to the test development companies Affymetrix and Illumina.[27]

The big companies do not always refine the tests they acquire, nor do they necessarily develop the digital structure beyond the websites (or the apps, as we shall see). This point is crucial, for the second wave of cybergenetics was marked by the importance of *big data* and the technical-scientific structure that undergirds cybergenetics. The technological structure that refines the tests but also imposes its own model for interpretation and even narration of the results is concentrated in the hands of just a few companies, of which Illumina is probably one of the most interesting.

This second wave broadened the focus on well-being and nutrition and on genealogy to include health, and this was no accident. The shift entailed a form of objectivization or even substantialization: on the one hand, a preoccupation with health constitutes a form of medicalization of daily life, drawn straight out of "healthism," as we shall see shortly. On the other hand, questions of health turned out to be at the heart of a certain way of apprehending genealogy, associating susceptibility to certain diseases with specific ethnic origins: for example, the susceptibility of Ashkenazi Jewish women to

[27] See Hogarth and Saukko 2017. 23andMe now uses Illumina's technology.

breast cancer, or the susceptibility of persons of Mexican origin to diabetes.

This expansion did not occur without controversy. Thus 23andMe (along with four other American companies) found itself on the hot seat: in 2013, it was reprimanded by the Food and Drug Administration and required to withdraw those of its tests that were specifically linked to the detection of diseases for several years. After a long process of revising its commercial policies (and presumably of intensive lobbying), the company was allowed to market tests for Bloom's disease and certain BRCA tests linked to the risk of breast cancer.[28]

2.2.2 *Recreational genetics today*

By the end of the 2010s, the third wave of cybergenetics was expanding broadly, offering a rich panorama that continues to grow. Advocates of recreational cybergenetics, Anne Wojcicki first and foremost, have been making increasingly positive claims: according to Wojcicki, the new challenge of cybergenetics is to offer the ultimate in personalization, in both diagnostic and therapeutic terms.[29] Still, it is hard to pin down the distinguishing features of this third wave, precisely because the DTC companies themselves are so diverse. The very definition of DTC endeavors is also problematic, in that certain companies propose only partial services.

[28] It is important to specify that only the tests of certain susceptibilities were authorized by the FDA – or, more precisely, were not prohibited. For more details about the tests and the limits of FDA approval, see Janssens 2018, "Opinion: No, the FDA Didn't Really Approve_23andMe's *BRCA* Test." On the modifications in 23andMe's approach, see Vlasits 2017, "How 23andMe Won Back the Right to Foretell Your Diseases."

[29] We may wonder whether we may not be on the verge, today, of a fourth wave in recreational cybergenetics, which distinguishes itself from the transhumanist desires of immortality by focusing its discourse rather on quality of life throughout one's lifetime, assuming a longer but not endless life, one in which quality and choice are privileged over quantity. See especially Wojcicki's remarks as reported in *Forbes*, where she explains that she does not believe, as transhumanists do, that we want to be immortal, but rather that we want to live a long time in good health: Carson and Chaykowski 2019, "Live Long and Prosper: How Anne Wojcicki's 23andMe Will Mine Its Giant DNA Database for Health and Wealth."

Narrowly defined, DTC companies can be said to be characterized by the absence of mediation, in at least four respects: first, they make a *direct offer* to the consumer, without going through a medical professional (by way of advertising, an internet site, sales in big box stores, and so on); second, they *receive samples directly*; third, they *analyze the samples directly*, in their own facilities; fourth, they provide *direct communication* to the consumer of the test results and their interpretation.

If we were to limit ourselves to this strict definition, we could count more than 150 DTC companies in the West in 2018.[30] Geographically, these companies are essentially based in the United States, Canada, Great Britain, Spain, and in Scandinavia. In the United States, they are concentrated in California and Utah: the latter owing to the emphasis on ancestry in the Mormon community, and the former owing to the development of biotech and digital industries in Silicon Valley.

Alongside the three major types of tests, corresponding roughly to the three focal points developed over time by DTC companies (genealogy, nutrition and well-being, health), these companies offer a myriad of other tests that may be only loosely linked to those categories. For example, there are tests that purport to determine one's risks of inheriting a specific disease (cancer, for example), one's capacity to respond to a certain type of medication, or one's degree of fitness. While these aims are perhaps unsurprising in themselves, the offers are accompanied by claims that range from more or less plausible to quite farfetched. For example, the Superhero DNA Test, is offered by the

[30] This figure is based on an empirical study carried out with students in the Institute for Society and Genetics at the University of California at Los Angeles in 2017-2018; among the participants I would like to thank in particular Kate Anna Clendenen, Rushna Raza, Jenny Ding, Mackenzie Grace Casey, and Antoine Rajkovic. A less strict definition (including all companies proposing to do genetic testing or to collect tests without medical mediation) suggests that there were 246 companies in 2016: see Philips 2016, "Only a Click Away – DTC Genetics for Ancestry, Health, Love … and More: A View of the Business and Regulatory Landscape."

Orig3n Company for $39; despite the questionable implication of its name, it can be compared to a fitness test packaged in a more amusing way, with a comic-book aesthetic.[31] Listed as out of stock in 2020, it is not clear that many people have been prepared to take it seriously, even at that price.

Packaging of the Superhero DNA Test (Source: https://shop.orig3n.com/products/superhero.)

A significant feature of the third wave of DTC cybergenetics is the diversification among several types of tests within given companies. Few now specialize in a single area; consumers are often urged to buy a set of tests, or to purchase, for a modest sum, additional elements of analysis on the basis of elements that have been retained (whether these are elements of samples or of data is not clear). Thus someone who has bought a test oriented toward genealogy is encouraged to add a test designed to detect health risks. Moreover, the companies are apt to modify their tools and procedures so that they can offer improved or "upgraded" tests to interested consumers when this is technologically possible[32] – it is of course a matter of completing or

[31] https://shop.orig3n.com/products/superhero. For an account of the testing experience, see also Pflanzer 2016, "I Took a $30 Test that Told Me If I Had 'Superhero' Genes – and It Was By Far the Most Fun Test I've Taken."

[32] See Molteni 2019, "Not Everyone on 23andMe Will Get the Latest Gene Chip Updates."

refining the previous results, not of testing their reliability, which tells us a lot about the less-than-scientific procedures of these companies. Specific technologies may be passed from one company to another, as was the case with Living DNA, which was shifted from the Illumina chip to that of Affymetrix.[33] Transfers like this imply numerous changes not only in what is tested but also in the way the data are interpreted.

As we have seen, alongside the DTC companies taken in the strict sense, the cybergenetics constellation includes a certain number of companies linked to DTC genetics that do not fulfill all the criteria proposed. We can thus distinguish at least two other types of companies: intermediary companies, which do not carry out tests but control initial communications, the reception of samples, and the communication of results, and collector companies, which download the data obtained by other companies and under some circumstances communicate those data. These functions are summarized in the table below:

Classification of DTC Companies

	DTC strict	DTC intermediary	DTC collector
Initial communications	Yes	Yes	No
Reception of samples	Yes	Sometimes	No
Analysis of samples	Yes	No	No
Follow-up communications	Yes	Yes	Sometimes

The intermediary and collector companies are very dissimilar. The intermediary companies may be small start-ups or may belong to – or depend more or less directly on – big companies; they offer an extremely varied assortment of tests as "story-telling kits." This is particularly the case for firms depending on the giant company Illumina, which holds the major part of the technological genetic test

[33] Threlkeld 2018, "Living DNA Announces Move from Illumina Microarray."

market and thus controls to a significant extent the way the tests are conceived and analyzed. Starting in 2015, Illumina helped develop a series of new companies, one of which, Helix, has become exceptionally powerful.[34]

Among the companies in the Illumina constellation, one claims that from their tests they can deduce your preferences in wine (or those of your boss!)[35]; another proposes to select a work of art for you based on your genes.[36] In all cases, the sales pitch exalts the uniqueness and the performance value of the product: the companies want to see what suits you personally so that you can be "the best version of yourself" – you or your children, for whom performance tests are also available, both before and after conception, with the goal of determining their natural talents.

Among the collector DTC companies, diversity takes different forms, depending on the way data is handled and used; the impact of the data can be characterized as primarily social or primarily political. While some of these seem to be at the borderline of DTC companies, since the cyber element takes precedence over the genetic, their increasing presence sheds light both on the prospects for development of the landscape and on the question of the post-genomic condition posited by Jenny Reardon, by showing what value individuals, families, and other actors in the private and public worlds place on their genetic tests, and what they want to do with them. But in every case, the difficulty that one encounters in seeking to define the field, with its blurred boundaries, is quite similar to the difficulties posed by the nanotechnologies.

[34] Jenny Reardon, relying on official documents made available by Helix, reports that in 2015 that company's value was estimated at $28 billion and that it produced more than 90% of consumer DNA data. See Reardon 2017, 18, and notes 104-106.

[35] This was offered by the Vinome Company up to January 2020; since then the company has gone out of business. One can find its test kit on Amazon, but the product is listed as "currently unavailable."

[36] This has been offered, for example, by Affinity (https://www.affinity-dna.com/dna-art-portraits-2/) and DNA11, https://www.dna11.com/.

2.3. A "gift," but for whom?

2.3.1 *The genetic DTC (super)market: direct to consumers = direct to companies*

If a shift to the low-cost economy is one of the trademarks of recreational cybergenetics and one of the reasons for the explosion of its landscape, we must nevertheless not misread the situation: "low-cost" has been synonymous with high profits, at least for a time.[37] It is correlated with the development of a new economic model based on data synthesizing (Harris et al. 2016, 8) and the sale or use of the data generated by genetic and other tests; these are only the visible part of the iceberg (for the consumer), and only the point of departure rather than the destination (for the companies).

The idea of "direct" cybergenetics has been developing toward an extension of the model based on the absence of medical intermediaries in a way that prolongs the ideology of "healthism." This notion that maintaining good health is the responsibility, indeed the moral obligation, of the individual, has spread widely in the United States, especially in California. It was characterized by Robert Crawford in 1980 as a form of medicalization of daily life, inseparable from looking good and being fit.[38] This movement also includes the idea that patients must take care of themselves: certain health policies have been based on this idea, especially in order to lower the length of hospital stays and send patients home faster, thus lowering health costs (for the medical system) and exercising pressure on patients and

[37] In 2019, Forbes assessed 23andMe at $2.5 billion and Anne Wojcicki's personal fortune at $690 million. See "Forbes Releases 2019 List of America's Richest Self-Made Women, A Ranking of the Most Successful Women Entrepreneurs in the Country" (2019). However, in 2024 that profit seems to have been dramatically reduced, with the possibility of a complete collapse in 2025, according to the *Wall Street Journal*. https://www.wsj.com/health/healthcare/23andme-anne-wojcicki-healthcare-stock-913468f4.

[38] Crawford 1980, "Healthism and the Medicalization of Everyday Life."

their families, even sometimes making them feel guilty for seeking medical care.[39]

Healthism has been denounced as a form of conservative free-market ideology[40] that can veer toward a disease of democracy, totalitarianism and Nazism in its state version,[41] by setting criteria for well-being (and good behavior) in addition to the criteria for differentiating between normal and pathological states. Healthism thus entails both a demedicalization of the medical field and a medicalization of daily life. Nineteenth-century novelist Émile Zola claimed that medicine had replaced the Church in the regulation of institutions, and philosopher Michel Foucault reinforced Émile Zola's claim[42]; today we might go even further in this direction by showing how it could lead to establishing a form of "cybergenopanoptics," a possibility we shall explore shortly.

Thus it is hardly surprising to see Anne Wojcicki affirm that the goal of 23andMe is not simply to communicate to its customers the results of their health tests. A 2019 ad invites potential clients to "commit to a healthier you inspired by your genes"[43]

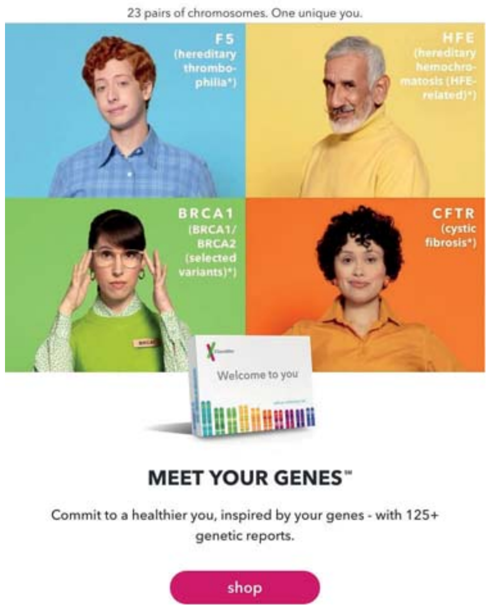
[39] See Veatch 2009, *Patient, Heal Thyself: How the New Medicine Puts the Patient in Charge*.

[40] See Cheek 2008, "Healthism: A New Conservatism?"

[41] "The pursuit of health is a symptom of unhealth. When this pursuit is no longer a personal yearning but part of state ideology, *healthism* for short, it becomes a symptom of political sickness. Extreme versions of healthism provide a justification for racism, segregation, and eugenic control since 'healthy' means patriotric, pure, while 'unhealthy' equals foreign, polluted. In the weak version of healthism, as encountered in Western democracies, the state … uses propaganda and various forms of coercion to establish norms of a 'healthy lifestyle' for all. Human activities are divided into approved and disapproved, healthy and unhealthy, prescribed and proscribed, responsible and irresponsible" (Skrabanek 1994, *The Death of Humane Medicine and the Rise of Coercive Healthism*, 15).

[42] See Michel Foucault 2008 (1975), "Panopticism" from *Discipline & Punish: The Birth of the Prison*, trans. Alan Sheridan, *Race/Ethnicity: Multidisciplinary Global Contexts* 2 (1): 1-12. https://muse.jhu.edu/article/252435/pdf.

[43] On this point, see Carson and Chaykowski 2019.



Advertisement on the 23andMe website in 2019[44]

It is hardly surprising, either, to see Wojcicki extending the perspectives of her company in two directions – toward coaching on the one hand, and the sale of medications based on genetic tests on the other – at the very moment when Apple was developing apps for its iPhones for similar purposes.[45] More precisely, on the one hand 23andMe, asserting the desire to offer consumers better control of their health, went into partnership with Lark Health, a startup that offered its clients counseling on how to manage diabetes. On the other hand, a 2015 deal with the biotech giant GSK for a four-year partnership gave 23andMe $300 million for the purpose of

[44] This ad no longer appears on the 23andMe homepage but it can still be seen on a 2020 video: https://www.ispot.tv/ad/Irk7/23andme-meet-your-genes-commit-to-a-healthier-you.

[45] See Matthew Herper 2015, "In Big Shift, 23andMe Will Invent Drugs Using Customer Data."

developing new drugs[46]; the arrangement was renewed and extended in 2022. Under the terms of this agreement, 23andMe shared the genetic data of the more than five million people who had used their testing platform to date.

The company's advertising is particularly revealing of the new modalities of relation to oneself that cybergenetics is making available: customers are to become entrepreneurs of themselves – but of the *best* version, of course, a version optimized by the advice or instructions of the cybergenetic company. Here we are undoubtedly at the peak of what Nikolas Rose calls *ethopolitics*, in *The Politics of Life Itself: Biomedicine, Power, and Subjectivity in the Twenty-first Century* (2007, 27): "If 'discipline' individualizes and normalizes, and 'biopolitics' collectivizes and socializes, 'ethopolitics' concerns itself with the self-techniques by which human beings should judge and act upon themselves to make themselves better than they are." In other words, cybergenetics finalizes a perfected phase of capitalism "in which all individuals are viewed as engineers having to invest in their own biological capital."[47]

Each of the two directions (coaching on the one hand, "made to order" drugs on the other) sheds light on the posture of DTC cybergenetics, which tends to infiltrate all aspects of our lives, from the most pathological to the most prosaic, moving further and further into intimate domains. Rather than evoking Foucault's enclosed environments as a model, perhaps we should turn to the "societies of control" described by Gilles Deleuze: in these societies, he avers, surfing replaces all other sports, in an open-ended temporality (1992 [1990] 1992, "Postscriptum on the Societies of Control"). We might add that digital surfing too, through the use of health apps, integrates

[46] https://www.gsk.com/en-gb/media/press-releases/gsk-and-23andme-sign-agreement-to-leverage-genetic-insights-for-the-development-of-novel-medicines/. It seems that it may have been precisely this turn toward pharmaceuticals that led 23andme to spend so much in that sector that its financial situation has been seriously compromised.

[47] Lafontaine 2014, *Le corps-marché: La marchandisation de la vie humaine à l'ère de la bioéconomie*, 14.

in an even more absolute manner the cyclical movement described by Deleuze, which ends up with a form of internal enclosure masked by the diversity of successive waves.

Not only does the diversity of genetic tests seem virtually limitless today, embracing even tastes in wine and intimate relationships, but the development of these new directions seems apt to allow DTC cybergenetics to play the role of spiritual and physical director, whether in ordinary and thriving states of being or under pathological conditions. Cybergenetics appears capable of inserting itself into our lives through what I call a *genethos*.[48] Yet our ways of life are reduced to certain models, certain specific patterns that schematize them; they do not aim to facilitate the emergence of life forms in which the life of forms would be creative.[49] This genethos corresponds perfectly to what Jacques Ellul describes as "encirclement by what is obvious," which he characterizes as a capacity to give people the impression "that they are more close, more familiar, more individualizing, more personal." He concludes: "Here is the true technological innovation, for it is by this basic support of the whole social body and of each individual that the system can develop without encumbrance" (1990, 18).

This possibility is not only permitted but *stimulated* by the free-market liberal model in which the DTC approach is embedded, from the ethical, political, and economic standpoints. Indeed, as Richard Scheller, a former Stanford professor and current research director at 23andMe, puts it straightforwardly: "I thought it was genius actually

[48] More broadly defined, the term genethos designates the creation of a certain way of life (an ethos) that is limited to certain pre-organized patterns by the self-fulfilling prophecy of cybergenetics. These patterns are life forms and not forms of life. It is hardly far-fetched to envision a society in which cybergenetic tests could be taken into account by insurance companies, employers, even educators and dating apps.

[49] I shall return to this point in discussing artificial intelligence, in which the genethos is embodied in a slightly different way in terms of patterns of life and habits.

that people were paying us to build the database" (cited in Carson and Chaykowski 2019).

To put it succinctly, the dollar pump actually works in both directions for the DTC companies: at the entrance and the exit alike. *Consumers have the impression that they are paying for a service, whereas they are the ones producing labor and offering goods of value.* Their benevolent contributions are situated at the junction point between two distinct elements: free labor on the Internet on the one hand, the bioeconomy and clinical labor on the other.

First, free labor is an integral part of the neo-capitalist Internet economy. As Tiziana Terranova showed clearly in "Free Labor: Producing Culture for the Digital Economy" (2000), the difference between worker and consumer tends to fade away in this context. Caught up in practices of sharing and cooperation, clients supply new data by participating in on-line surveys, forums, and so on, as we shall see.[50]

Second, this model of free labor occurs in the medical field in the forms of bioeconomy (O'Riordan 2010, 18) and biocapital,[51] as well as clinical work[52]: even if spitting into a test tube does not take much time, it remains a form of work, especially when the quantity of saliva is considerable, as attested by numerous videos posted on YouTube.[53]

This is why we can actually speak of a *biodigital* dimension (with reference to biocapital) of DTC cybergenetics. Clients offer health data and samples whose ultimate value cannot really be measured; still, their market value is clearly considerable. The fact that 23andMe offered a million free tests to African-American individuals in 2011 in an attempt to enrich its data base[54] suggests how ironic it is to

[50] These surveys are even explained on 23andMe's blog; see for example https://blog.23andme.com/news/three-new-surveys-from-23andm3-ask-new-kinds-of-questions. (accessed August 20, 2019).

[51] See Rajan 2006, *Biocapital: The Constitution of Postgenomic Life*.

[52] See Cooper and Waldby 2014, *Clinical Labor: Tissue Donors and Research Subjects in the Global Bioeconomy*.

[53] For an analysis of some of these videos, see Harris et al. 2016.

[54] The results of this program were published by 23andme in 2012 (https://blog23andme.wpengine.com/wp-content/uploads/2012/11/ASHG2012RITFposter.final_.pdf). This first program has been followed by several others aiming at

rejoice over the low costs of these tests; their true value is certainly not reflected in the fees that the companies ask their purchasers to pay.

It might well be appropriate here to conceive of that value in terms of what Céline Lafontaine calls *le corps marché*, the "body market"[55]: biocapital based on an ideology of performance, a rhetoric of promise proper to a post-mortal society inspired by the transhumanist deviations of cybernetics – which open, here, onto another aspect of the meaning to be given to the term "cybergenetics," along with a strong link to artificial intelligence.

Once a company has received its clients' data, it can proceed to sell new services and products. Consumers thus supply the companies *directly* with data with which the companies can do pretty much whatever they wish, for their model is that of implicit consent: clients who do not want their data to be exploited have to opt out explicitly – but the ways of doing so are often far from obvious. As Linnea Laestadius, Jennifer Rich, and Paul Auer emphasize in "All Your Data (Effectively) Belong to Us: Data Practices Among Direct-to-Consumers Genetic Testing Firms" (2017), on the basis of a survey of some thirty of the most significant DTC companies, in practice, cybergenetic DTC businesses take care to ensure their ownership of health data. To summarize, I would suggest that *DTC* can be understood in two senses: "Direct to Consumers" also, and perhaps above all, "Direct to Companies." The companies thus manage to exercise a form of control over the data and their interpretation, even if, as we shall see, that control remains relative.

expanding the company's database with data from people of African origins. In 2020 a debate on the origins of the people working at 23andMe revealed what is called in AI the "white guy problem," as we will see in the next chapter. See for instance https://www.statnews.com/2020/06/10/23andme-ancestry-racial-inequity-genetics/.

[55] As Lafontaine puts it: "More fundamentally, the body-market constitutes the economic infrastructure of the post-mortal society, in which the maintenance, control, improvement, and prolongation of bodily vitality have become the guarantors of the meaning given to existence" (2014, 13).

2.3.2 *Cybergenetics and health data: the gift of (big) data*

Cybergenetic DTC companies operate in a broader context marked by at least five distinct elements whose convergence creates an extremely favorable terrain for using the data received.[56]

First, the ambient discourse implies that these companies are seeking to serve the common good, whether this involves medical research or the fight against crime. In 2008, 23andM*e* thus created 23and*We*,[57] inviting its clients – and prospective clients, for $99 – to participate in scientific research, and thereby to become "part of something bigger."[58] Launching an appeal to the "23andMe community," the company stressed that it was proposing a "new" type of research: faster than the traditional forms (since it could dispense with the "administrative hassles" of academic research – including ethics committees – and because it had significant financial resources); it had closer ties with the public, and it could rely on a very substantial data base.[59] The project's legitimacy could hardly be doubted, given that this call for participants was supported by prestigious public and private partners (such as the University of Chicago, Harvard, and MIT on the academic side, Genentech and Biogen on the industrial side), and by several charitable organizations.[60] As further evidence of credibility, the company listed its more than two hundred publications,

[56] *Translator's note:* The French word for data, *données*, also means "gifts."

[57] This is "the arm of 23andMe that gives people an unprecedented opportunity to collaborate with us on cutting-edge genetic research." https://blog.23andme.com/23andme-and-you/23andwe-the-first-annual-update.

[58] "Becoming part of something bigger. Our genetic research gives everyday people the opportunity to make a difference by participating in a new kind of research – online, from anywhere. Once participants answer online survey questions, researchers link their genetic data to study topics from ancestry, to traits, to disease. These contributions help drive scientific discoveries." https://www.23andme.com/en-int/research/ (accessed August 20, 2019).

[59] "With the help of our 23andMe community we believe we can accelerate research and make an impact with our genetic data." https://www.23andme.com/en-int/research/ (accessed August 20, 2019).

[60] Examples include the National Parkinson Foundation, and the separate foundation created by Michael J. Fox, the actor known, for instance, for his role in the film *Back to the Future*. (Fox is a Parkinson's patient.)

often produced in collaboration with prominent international universities.[61] Cecile Janssens and Peter Kraft analyzed the stakes of this stance on the part of DTC companies; in "Research Conducted Using Data Obtained through Online Communities: Ethical Implications of Methodological Limitations" (2012), they demonstrated that, if 23andMe constitutes the most fully realized example of this tendency, it is not the only one.

Table 1. Examples of online research initiatives.

Initiative	Aims and Claims
PatientsLikeMe.org	"To provide a better, more effective way for you to share your real-world health experiences in order to help yourself, other patients like you and organizations that focus on your conditions."
23andMe.com	"Our research arm, 23andWe, gives customers the opportunity to leverage their data by contributing it to studies of genetics. With enough data, we believe 23andWe can produce revolutionary findings that will benefit us all."
Personal Genome Project (personalgenomes.org)	"The mission of the Personal Genome Project is to encourage the development of personal genomics technology and practices that: are effective, informative, and responsible; yield identifiable and improvable benefits at manageable levels of risk; are broadly available for the good of the general public."
DIYgenomics.com	"A non-profit research organization founded in March 2010 to realize personalized medicine through crowdsourced health studies and apps."
Genomera.com[a]	"We're crowd-sourcing health discovery by helping anyone create group health studies."
Curetogether.com[b]	"Bringing patients into research as active partners is one of our big missions at CureTogether." [21]

Quoted information was downloaded from the organizations' websites on July 1, 2012.
[a]Beta version.
[b]Acquired by 23andMe.

Source: Janssens and Kraft 2012

By positioning themselves in the world of research, these companies, like 23andWe, help us extirpate ourselves from the navel-gazing into which 23andMe might have led us by allowing us to do good works, together, and to look toward "something bigger than ourselves." However, the "link" to the public is limited to an expression of thanks by the companies, along with information about their blogs and emails – nothing more than what an academic research project would have provided. The participants have no real right to oversee the way in which their samples or their data are used. However, this does not prevent 23andMe from speaking of the "democratization" of research.

[61] https://www.23andme.com/en-int/publications/.

The second element on our list is the public/private relationship, which takes a form in American culture that distinguishes it sharply from the European model (and especially the French version). In the United States, the idea of sharing is inculcated from early childhood on: sharing toys, sharing experiences, sharing information.[62] School classes often include a regular "sharing time" during which the children are expected to "share" their hobbies, activities, and so on, and sometimes to produce and show a poster on which they share their own story ("all about me"). The opposite expectation prevails in European – and especially French – schools, where children are not supposed to talk too much about themselves in class. The development of the Internet and social media, blogs, and platforms such as Instagram have reinforced this tendency to "share."[63] The Protestant culture of "open curtains" probably contributes as well: the idea is that if one has nothing to hide, then one must show what one has. In addition, the post-2001 surveillance culture in the United States holds that nothing to hide means nothing to fear, further reinforcing the idea that failure to reveal everything may be suspect.[64]

The notion of sharing has been a fundamental aspect of the Internet since its beginnings, and it has only increased with the development of blogs, Facebook, and other digital instruments, so much so that it has practically become an injunction. Some voices have begun to be raised against that trend, arguing that while there may be nothing to hide, there is definitely something to lose: the loss of privacy entailed by the potential exploitation of health data by insurance companies or employers is a telling example.[65] The fact remains that sharing in such instances is not mutual: the DTCGT companies do not *share* their data gratis, they *sell* them.

[62] On this important facet of Anglo-Saxon (and especially American) pedagogical culture, see for example Brooke 1966, "'Sharing Time' in the Elementary School."

[63] See especially John 2013a, "The Social Logics of Sharing"; 2013b, "Sharing and Web 2.0: The Emergence of a Keyword"; and 2017, *The Age of Sharing*.

[64] See Solove 2011, *Nothing to Hide: The False Tradeoff between Privacy and Security*.

[65] See Cofone 2020, "Nothing to Hide, But Something to Lose."

A third element, then, is the purported *openness* of genetic data, which has from the outset constituted an important feature of the field of genetics, taking the form of what Jenny Reardon has called *genomic liberalism* (2017, 7). DTC companies assert a form of "democratization" of knowledge: they claim that, where medical paternalism – which Reardon characterizes as "feudal" – once held power, individuals are once again becoming masters of their own data, and can share them by following the democratic procedures of the World Wide Web – or the World *Wild* Web, given the extent to which this thrust resembles a new gold rush heading toward the Wild Wild West.

We are perhaps seeing what physician Susan Desmond-Hellmann, a CEO of the Bill and Melinda Gates Foundation from 2014 to 2020, has called "a new social contract."[66] This contract does not aim to protect certain elements as private; rather, in line with the culture of "sharing," it aims to promote sharing of personal data, suggesting that it is neither ethical nor good politics *not* to share one's health data – an idea that is contestable, at the very least, from an ethical standpoint. Nevertheless, this idea has been adopted to the letter in France, in the framework of a Health Data Hub (titled in English in the original!) designed to serve as a national biobank, housed on a private cloud, "a unique window of access to the entirety of health data supported by national solidarity."[67] This project aims to benefit public and private research (via a selection committee) by making

[66] See Hellman 2012, "Toward Precision Medicine: A New Social Contract?": "A unified group of patient advocates pushing government, academia, private industry, and caregivers to create a new social contract in which patients both contribute and benefit would be a powerful force."

[67] https://www.health-data-hub.fr. I should note that the Health Data Hub has recently been criticized by the Commission nationale de l'informatique et des libertés (CNIL) because its data are housed by the giant American company Microsoft. See Piquard and Untersinger 2022, "Coup d'arrêt pour le Health Data Hub, projet de centralisation de données médicales impliquant Microsoft." For an overview of the problems raised by this platform on the legal level (prior to the announcement of its being hosted by Microsoft), see Margo Bernelin 2019, "Intelligence Artificielle et santé: La ruée vers les données personnelles."

"sharing the rule, closure the exception."[68] While the path chosen is a different one, since France is committed to a form of regulation of the biobanks it constitutes and since its perspective is broader, given that it concerns all patients participating in the French public health system, the principles and rhetoric are very similar to those of the DTC companies: health data are not private in the sense of personal, but they are eminently privatizable, or at least apt to become a fine gift for private industry – and even, in the end, for insurers, employers, and so on. Here we can see clearly how the idea of the social contract defended by Jean-Jacques Rousseau and others is being corrupted by a misleading ideology of sharing.

The fourth element of the five I have identified is *participation*, one of the strong points of cybergenetics. This is hardly surprising, given that participation is a characteristic feature of certain health-related fields as well as of information science.[69] The DTC sites typically propose forums or discussion groups in which users can share their information and thoughts but can also establish relationships with people whose genetic profiles are similar, and especially with more or less closely related family members.

In these forums, clients offer new information, adding to what they have already provided by volunteering to answer survey questions. Clients also often receive "badges" for their participation, depending on how much the other members of the forum "like" their answers to questions.[70] In short, participating is good; making oneself "popular" is even better.

[68] See the report of the planning commission: Mission de préfiguration 2018, "Rapport Health Data Hub."

[69] For an excellent analysis of this type of participation, see Barbara Prainsack's work, especially Prainsack 2014, "Understanding Participation: The 'Citizen Science' of Genetics."

[70] "Want to earn your own badges? You get points toward badges by answering questions or posting contents that other members like." https://customercare.23andme.com/hc/en-us/articles/214116497-Tips-for-using-the-new-23andMe-Forums (accessed August 20. 2019). See also Harris et al., *Cybergenetics*, 84-85.

Furthermore, by downloading their own genetic data on the sites of DTC collector companies, not only are individuals offering their data for free, but they are also, let us recall, doing what looks very much like real work – unremunerated, needless to say. This form of neocapitalism, in which consumers have the impression of coming out ahead when they are actually exchanging gold for worthless trinkets, is a fundamental element of recreational cybergenetics. And this practice also allows the DTC companies to bypass certain legal obstacles governing genetic data, a ploy that also occurs when DNA is used in police investigations.

The participatory dimension of DTC companies puts into play a veritable invisible labor force, as do other facets of the digital realm, which subtly induce "click work" through the rhetoric of "sharing" and "participation."[71] This point is all the more interesting in that a good number of DTC companies, with 23andMe at the head of the pack, are in virtually incessant contact with their clients, and this characteristic is becoming more and more prominent with the development of apps. However, the connection seems unequal, in that the companies are not modeling transparency or concern for others; rather, they are accounting for the "consumer experience" (in which the consumer offers an opinion, and thus performs free labor) and "communication." The model for this approach to participation, as practiced by some of the largest digital companies, most notably Apple, is unquestionably that of online sales.

With the fifth element, finally, all this information is *synthesized* with other information available on the Web, often without the consumer's knowledge. Moreover, the development of apps makes it possible to combine these data not only with the traces left by earlier Internet navigations but also with mobile data gathered in real time. This is an essential point, because it allows the companies to collect data concerning its clients' biological ("objective") interiority and their experiential ("subjective") interiority, their experiences, tastes,

[71] On "click work," see especially Casilli 2019, *En attendant les robots: Enquête sur le travail du clic*.

contacts, and so on, from the past and in the present, from the inside and from the outside, this latter in particular via geolocalization. One DTC site associated with a dating app called "Pheramor" (closed down in 2019) promised that by synthesizing your biometric self with your digital self it could connect you with your soulmate, and even offered to facilitate the task by offering geographic localization, thanks to GPS. In fact, much of the interest of the data lies not simply in what they are in themselves but in the possibility of combining them, whether at the individual level (different types of data for a single individual) or at the interindividual level (the same types of data for different individuals) in order to bring patterns to light (or not), according to a logic that we have already encountered. Love life is thus reduced to schemas and leaves no place for the unpredictable – already paving the way for the revolution in this domain promised by artificial intelligence; this is hardly surprising, since Pheramor combined genetic data with other types of digital data.

Thus the social acceptability of cybergenetics has been playing since its beginnings with a form of extreme personalization and ubiquity based on the idea of community, or even the idea of a social contract 2.0 in which "sharing" benefits an industry whose fruits are not so well shared in turn. While the popularity of cybergenetics is undeniable and its social acceptability significant, it is still far from self-evident that it is acceptable from an ethical or even a political viewpoint. The gap between countries such as the United States and France, for example, is huge, since these recreational tests, so highly developed in the U.S., are simply prohibited in France.[72] The question of the ethical and political acceptability of this technology is thus, as was already the case with the nanotechnologies, a fundamental question that cannot be avoided. The question arises with all the more acuity if we consider the way cybergenetics manifests another

[72] For an overview of legislation concerning DTCs in Europe as of 2018, see Louiza Kalokairinou et al. 2018, "Legislation of Direct-to Consumer Genetic Testing in Europe: A Fragmented Regulatory Landscape." Despite the prohibition, a posting from BFM-TV in 2020 indicates that some 100,000 tests are used in France each year: https://www.bfmtv.com/societe/tests-adn-interdits-en-france-ils-sont-pourtant-realises-par-100-000-personnes-chaque-annee-1851864.html.

characteristic that we have already encountered in the nanotechnologies, namely, the self-fulfilling prophecy – even if it does so in a more logical and more pertinent way – so as to reshape basic elements of our individual and collective identities. When the notion of the self-fulfilling prophecy has free rein, it brings about without restraint Ellul's "encirclement by what is obvious," thanks to which the technological system manages to impose itself by imparting a feeling of proximity, familiarity, and personalization.

2.4 Recreation and re-creation: the world according to DTC cybergenetics and its self-realization

If a large part of the effort made by certain DTC companies points toward the desire to influence consumers' health behaviors (health taken in the broad sense here), it is far from clear that the tests in question actually have a significant impact on consumers' lifestyle habits insofar as their health is concerned.

We may well wonder whether the influence of DTC cybergenetics is not playing out elsewhere today, surreptitiously inserting itself into numerous aspects of our lives. More precisely, I seek to show that so-called recreational cybergenetics is also re-creational, and that it tends to reshape our collective and individual identities as well as our social, ethical, and political relations.

2.4.1 *Reshaping collective and individual identities: from self-decoding to self-narration*

As we have just seen, the development of digital technology has disrupted the relations between public and private spheres. With its exaggerations, its hype, it takes us into a form of self-realizing prophecies both similar to and different from the one we encountered with nanotechnologies. It is different first because it is undeniable that genetics really does allow certain types of predictions; second, because, even so, not everything is always already here, already played out. Cybergenetics and nanotechnologies are similar in the way they are undergirded by self-fulfilling prophecies, although cybergenetics goes even further in that direction, in that the digital genetic self and the "objective" genetic self are inseparably linked. Can deciphering the genome actually lead to a reshaping of identities? And if so, to what extent?

Let us begin with an issue that is a source of astonishment when, arriving from Europe, one discovers the breadth of the phenomenon known as recreational genetics in North America. Why do so many Americans participate? According to one 2017 article, 74% of the "most interested" survey respondents answered: "to learn something about my origins" (Roberts et al. 2017). According to the *MIT Technology Review,* more than 26 million people had taken recreational cybergenetic tests as of February 2019.[73] Ancestry.com alone had a genetic database of more than 10 million people in 2018[74] and 15 million in August 2019[75]; these figures give a sense not only of the size of its database but of its exponential increase.

While such figures may look enormous, they are not completely surprising if we consider that genealogy is a very widespread hobby, especially in the United States, as we have seen. Its historical evolution in that context is quite interesting. Early on, it was a leisure time activity for Whites of Anglo-Saxon Protestant origin (WASPS), at least some of whom were hoping to find ancestors on the Mayflower, as a form of nobility; later, it became a means for Black Americans to try to trace their family roots back before the period of slavery, although at least initially the DTC companies' databases suffered from limited representation of Black populations. More recently, genealogy has enabled some individuals of Native American origin to discover their roots or to reclaim, directly or indirectly, certain advantages linked to those roots, such as affirmative action for university admission.

As Steven Pinker wrote in his 2009 *New York Times* article: "It's thrilling to find yourself so tangibly connected to two millenniums of history… . [A]t some point there aren't enough ancestors to go

[73] See Regalado 2019, "More Than 26 Million People Have Taken an At-home Ancestry Test."

[74] See Holger 2018, "Ancestry DNA Review: The Largest DNA Database for Finding Relatives and Heritage."

[75] https://www.ancestry.com/corporate/about-ancestry/company-facts (accessed August 22, 2019).

around, everyone's ancestors overlap with everyone else's, and the very concept of personal ancestry becomes meaningless. I found it just as thrilling to zoom outward in the diagrams of my genetic lineage and see my place in a family tree that embraces all of humanity." Here again, as with 23andWe, the idea is to feel connected to something that is greater than ourselves but that concerns us intimately and allows us to make sense of our collective and individual identity.

Such genetic searches for genealogical purposes via DTC companies have become common currency, and they come up in blogs, books, magazine articles, and even television programs – their numbers are such that it would be impossible here to pin them down. Of course, the results are not always free of surprises. For example, when the actress Kim Trujillo made a commercial for the DTC company Ancestry.com, she stressed that she had learned that she was "26% Native American."[76] In fact, the entire 26% were Genízaros. For Trujillo, as for many North American Latinos who had taken the test and gotten their results, a whole new facet of history opened up. More or less forgotten in American history books, the Genízaros were in effect mixed-race slaves, with representation from many Native American tribes, Black Americans, and Latinos. Genizaros may have accounted for as much as a third of New Mexico's population in the eighteenth century.

As the *New York Times* reported: "The revelations have prompted some painful personal reckonings over identity and heritage. But they have also fueled a larger, politically charged debate on what it means to be Hispanic and Native American."[77] Unlike Black Americans, who often seek to go beyond slavery as the origin point of their family histories, these revelations inscribe the descendants of the Genízaros in a different and long-neglected history of slavery; some of them have come together to claim the same type of recognition that is

[76] https://www.ispot.tv/ad/wKqV/ancestrydna-kim. For another example, see Miguel Torrez's blog: https://nmgeneticgenealogy.wordpress.com.

[77] See Simon Romero 2018, "Indian Slavery Once Thrived in New Mexico. Latinos Are Finding Family Ties to It." See also Gonzales and Lamadrid, eds., 2019, *Nación Genízara: Ethnogenesis, Place, and Identity in New Mexico*.

granted to Native American tribes. This grasp of their origins is thus playing out in a specific way within American history, with links to other fragilized population groups, in a subtle interplay of comparison and differentiation.

That genetics should alter collective identities is not a new phenomenon.[78] In the late 1980s and early 1990s, researchers from the University of Arizona conducted a genetic study of members of the Havasupai Tribe living in Grand Canyon National Park. They had analyzed samples originally collected for research on diabetes, but they then used the same material without the subjects' consent for research on other subjects. The genealogical study, based on samples of 100 individuals from a total of 650 members in all (thus a significant percentage), produced information not only about degrees of kinship but also about the geographical origins of the subjects' ancestors. As it happened, the researchers' conclusions contradicted the tribe's sacred oral history, which affirmed that they were originally from the Grand Canyon and were its assigned guardians.

In a more prosaic way, these tests also redefine family relations. Paternity tests are among those most often sold (though there are maternity tests as well), and countless clients report that these tests have led them to reconsider their closest family relationships: grandparents, parents, brothers, and sisters. The resulting redefinitions can vary according to sex (and sometimes gender): paternity tests, for example, are often associated with mistrust of the (perhaps unfaithful) mother. Similarly, health tests focused on hereditary diseases often highlight certain types of transmission and are embedded in certain cultural frameworks or stereotypes. Austrian sociologists Ulrike Felt and Ruth Müller have shown in their work on medical genetic tests on the BRCA 1 and 2 genes, that genetic testing is not a neutral experience from the cultural standpoint[79] – nor is it neutral from the standpoint of gender, as we shall see.

[78] For an overview, see Prainsack and Hashiloni-Dolev 2009, *Religion and Nationhood: Collective Identities and the New Genetics*.

[79] See Felt and Müller 2011, "Tentative (Id)entities: On Technopolitical Cultures and the Experiencing of Genetic Testing."

Such cybergenetic genealogical research modifies not only the way we view our collective identities and our kinship relations; it can also modify the way others see us and their relationships with us. Cleon Brown, for example, a White police sergeant in Hastings, MI, was subjected to repeated racist taunts from his colleagues after he had let them know his genealogical test results: he was reportedly 18% African-American.[80] His colleagues finally went beyond bad jokes, and departmental relationships soured; Brown ending up suing the city, charging that he had received no support – worse still, it seems that even the mayor indulged in racist joking. It is worth noting here that the heart of the problem is not simply racist behavior but the fact that that behavior was triggered by a genetic test viewed as a more important marker of identity than direct personal contact with the individual on the part of his professional associates. As the Brown case suggests, not only familial but also professional relationships can be reshaped by these tests.

"Welcome to you" is the message spelled out on the box of the 23andMe test kit:

Source: https://www.23andme.com/en-ca/

[80] For more information about this case, see https://mlive.com/news/grand-rapids/2-17/05/white_cop_accuses_others_of+ra.html. I thank Aaron Panofsky for making me aware of this story, which he analyzes in detail in an unpublished manuscript.

As if, at the end of a trip, the DTC company's test allowed us to find ourselves! Most of the DTC companies play in a similar way on the idea of self-discovery, of finding one's underlying identity thanks to DNA kits. But the idea of connecting with one's true self, one's inner, objective, and invisible self, also permeates the rhetoric of the users. As Steven Pinker (2009) writes, users often affirm that "affordable genotyping [cybergenetic DTC testing] may offer new kinds of answers to the question 'Who am I' – to ruminations about our ancestry, our vulnerabilities, our character, and our choices in life."

An excellent example of self-redefinition through cybergenetic testing is found in Lone Frank's *My Beautiful Genome* (2011). Frank, a Danish journalist with a doctorate in neurobiology, volunteered to participate in a research study on the correlation between genes and character; at the same time, she was conducting her own journalistic investigation on the development of cybergenetics, including its recreational form. From the very beginning of her book, we learn that her own family history is troubled, that her closest relatives suffer from multiple psychiatric diseases.

One of the interesting features of this book lies in the parallels the author establishes between her own personal testimony and the anecdotes, interviews, and discoveries that she notes along the way. Her hypothesis at the outset could hardly be clearer: "Consumer genetics is about exploring and discovering yourself at a molecular level" (Frank 2011). But as Frank's testimony proceeds, the reader gradually discovers the extent to which the "data" that she has collected are reworked and readjusted in relation to her own personal history and the stakes she finds there. She concludes her book with these observations:

> Now that I have looked into my genes, the result is not a simplified self-image. On the contrary. It is rather that I'm experiencing more facets and nuances in my life. It is far more satisfying to be able to interpret myself as both a biological and a social being. My genes are not fate but cards I've been dealt, and some of those cards give me a certain amount of latitude in playing the game of life. Or, to turn another phrase, my genome is not a straitjacket but a soft sweater to fill and shape, to snuggle up and

> stretch out in. It is information I can work with and around, information that can grant me greater freedom to shape my life and my essence. It is also information that can, in its way, ease my existential burden. It tells me that I am not totally free, but neither am I completely responsible for who I am and what I have ultimately become. (155)

Her words echo those of Steven Pinker (2009): "For all the narcissistic pleasure that comes from poring over clues to my inner makeup, I soon realized that I was using my knowledge of myself to make sense of the genetic readout, not the other way around."

We could of course retort to these converging conclusions that Steven Pinker and Lone Frank are both consumers who are particularly well informed about the limits of these tests and particularly able to reflect on the link between what is innate and what is acquired – the relation between the social and the biological realms. Still, studies of very different populations – one on Black residents of the United States and Great Britain,[81] the other on the use of recreational cybergenetic testing by members of the racist American far right, based on the analysis of more than 7,000 posts on the public Stormfront discussion group[82] – go in more or less the same direction.

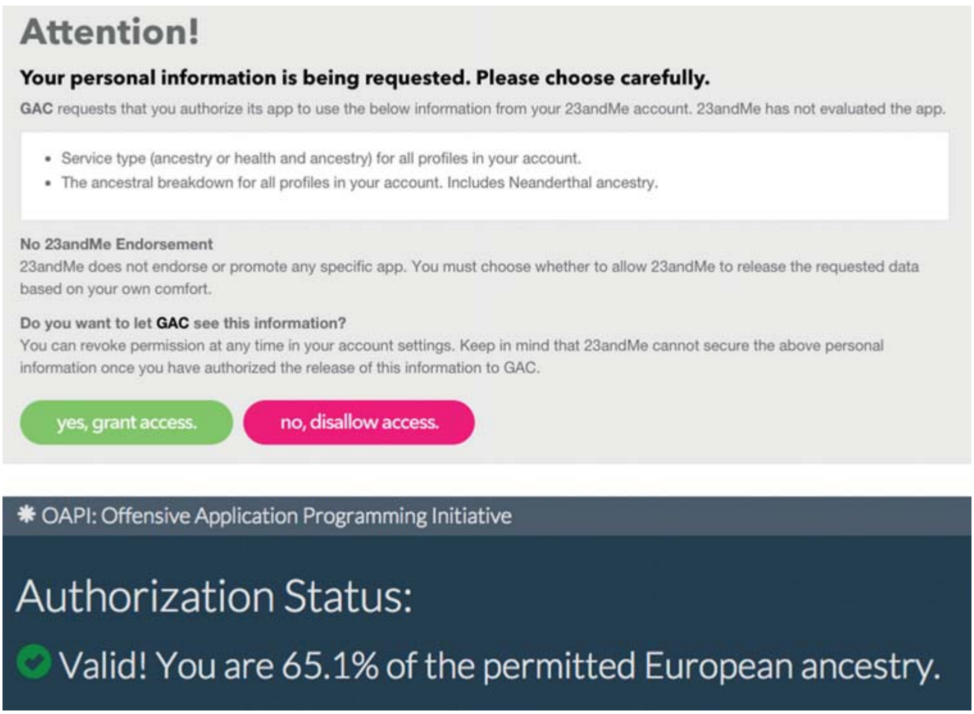
The ethnographic study Alondra Nelson carried out with Black participants in the United States and Great Britain attests to a certain caution with regard to genetic testing and to an influence – undoubtedly real, but ultimately mitigated by other elements – in the rewriting of personal identities within the collective. We witness variable combinations in which the genetic tests are conceived, in a sense, as one of the elements in the puzzle whose configuration will vary in relation to what Nelson calls the "genealogical aspirations" of the participants. The kinship of cybergenetics with a form of anticipatory design, while certainly significant at both the individual and

[81] See Alondra Nelson 2008, "Bio Science: Genetic Genealogy Testing and the Pursuit of African Ancestry."

[82] See Panofsky and Donovan 2019, "Genetic Ancestry Testing among White Nationalists: From Identity Repair to Citizen Science."

collective levels, thus remains less radical than we might have expected.

In another register, the DTC cybergenetic tests are especially popular in far right White supremacist contexts and they serve to limit access to certain forums that require prospective members to prove their degree of Whiteness by sharing their test results; in these cases, one speaks of Genetic Access Control (GAC).[83] The warning message and sample reports reproduced below are sufficiently explicit:



Genetic Access Control
Source: Screenshots from https://github.com/offapi/rbac-23andme-oauth2

[83] On this topic, see, for example, Clark 2015, "How One Coder Used 23andMe to Create a Race Wall around the Web."

As Panofsky and Donovan (2019) demonstrate, one of the uses to which White supremacist groups put these tests is verification that their members' ancestry is not Black, Native American, or Jewish (the inference being that Jews in their eyes may be White in appearance but "Black-ish" inside, in their DNA, so to speak). The tests also make it possible to establish a hierarchy in Whiteness, in which Scandinavian and Germanic ancestries rank ahead of Mediterranean origins. But research also reveals that its results are received in diverse and subtle ways, not always as caricatural as in the cases where access to a site depends on the degree of "genetic purity."

Some reactions tend in fact to propose "reparations" to individuals caught in the act, as it were, of having inadequate Whiteness. Two distinct strategies come into play when the "guilty" party asks for help from the supremacist community: the supplicant may either contest the scientific validity of the results on the grounds that everyone has mixed ancestry, or discuss the scientific value of these tests in detail and point to a large margin of error. Some of the reactions also stress the importance of self-affirmation: "Most WN's do not hold to a 'one-drop' rule. If you look White, live White, identify White, if your grand-parents and great-grand-parents looked White/lived White/identified White – that is often sufficient. … Not to mention that many WN's distrust the DNA services."[84] What can be read between the lines in these discussions is that genetic tests are received differently by different individuals, but that the recipients tend to tinker with the results in an effort to reduce the cognitive dissonance confronting them – a classic strategy, after all, for coping with this sort of gap.

However, the power to reject the test results often operates, as it were, from within the supremacist ideology, by appealing to a conspiracy theory. We can see that, whatever strategy is chosen, the use of cybergenetic tests ratifies a certain vision of race, of the self and the other, which shares in a form of self-realizing prophecy. On the one hand, where the tests are rejected, the stereotype of the

[84] Thomas Stuart, January 9, 2014, cited in Panofsky and Donovan 2019.

omnipresent deceitful Jew, in the style of the "Protocol of the Elders of Zion," is invoked; on the other hand, where the tests are debated in scientific terms, the affirmation of White Supremacy is revived, either by an invocation of the margin of error, with an appeal to the community, or by a redefinition of Whiteness, with an invitation to subscribe to a differentiated identity – which, to be sure, can include a diversity of groups among Whites. These two distinct strategies are clearly complementary: both reflect a stance of "us against the others" and reinforce the feeling of belonging to a community. That said, even if, as we have seen, the use of cybergenetic tests by White supremacists is complex, other elements intervene to complete the picture, leading us to examine the way in which these tests may contribute to a resurgence, a reshaping, or even a refounding of the notion of race in the guise of playing or tinkering.

2.4.2 *Recreascientific?*

One of the distinctive elements of DTC cybergenetics is its interactive dimension. As Kate O'Riordan stresses, interactivity, understood as biopolitical work, offers a form of pleasure and at the same time "functions as a mode for extracting value" (2010, 118). O'Riordan sees this value as essentially a product of engagement, interaction, or even a form of compulsion, though the companies cannot necessarily predict on what individual actors will choose to focus.

This potentially addictive engagement has taken on a festive dimension that has enhanced its value, I would suggest. DTC cybernetic tests have become standard holiday gifts, especially for Christmas and Thanksgiving – and while it is hard to imagine an advertising campaign promoting paternity tests for Valentine's Day (although anything is possible!), there seem to be almost no limits to the marketing strategies of these companies. We are confronting more than a passing trend, for sales keep reaching record levels on such occasions, thus allowing family databases – and thus the companies' fortunes – to grow. Even as families celebrate the joy of being together and also the links among diverse communities that are linked not by blood but by hospitality and mutual aid, by affective ties as much as kinship, the choice of such a gift may also appear paradoxical (and perhaps especially at Thanksgiving), since it foregrounds a biological rather than an emotional or relational conception of the family.[85]

These festive occasions can also take place outside of the family circle and become social or media events. 23andMe is particularly known for organizing big "spit parties," where celebrities are invited to spit without embarrassment into kits distributed by the company – which was "nice enough" to offer them "free of charge," although one can presume that the genetic data of stars may be worth their weight in gold, especially for kinship testing, not to mention the matter of paternity.

What is more, the playful and festive dimension of these products is accentuated by their interconnections, especially where genealogy is concerned, with television programs that bring in large audiences and which blur the borderlines between genres in their own way. One very popular program, "Finding Your Roots," is distributed by the Public Broadcasting Service; it invites celebrities to connect with their roots by exploring them through genealogy and genetics, with the help of the DTC cybergenetic companies. The program is complemented by summer camps called "The Seedlings" held on the campus of Pennsylvania State University: here, children aged 11 through 14 learn to construct their genealogy with the aid of cybergenetic testing; no one seems to have questioned the appropriateness of including these children's samples in data banks.[86]

The Seedlings program, conducted in collaboration with Penn State anthropologist Nina Jablonski, aims explicitly at developing these methods in middle school STEM classes, and it offers supplemental materials for teachers, so that they too can pursue the

[85] The French National Consultative Committee on Ethics has noted – with disapproval – that the use of DNA tests for family regrouping in the United States foregrounds a biological rather than a relational or emotional conception of the family. On this point, see the fascinating ethnographic study by Mélanie Gourarier (2017), "Faire la frontière dans les murs du laboratoire: Destins migratoires et usages de l'ADN aux États-Unis."

[86] https://www.fyrclassroom.org.

program in class, helping their students to ask themselves "Who am I? Who am I related to in my family and beyond, possibly thousands of years beyond? And then, what does this mean to me as a person in my life? They will realize they can discover all kinds of new information about themselves. They will know that they are actors in their own lives, empowered by the knowledge instilled in them. So we see nothing but goodness coming out of this."[87]

As we can see, this mix of addictive, festive, and scientific participation accounts in large part for the attraction of cybergenetics. It relies, we might think, on a "recreascientific" cocktail that blurs boundaries and makes the definition of cybergenetics rather difficult to pin down, all the more so in its DTC dimension, which probably constitutes the most symptomatic current example in this respect. It is nevertheless clear that this explosive cocktail corresponds quite precisely to the characterization proposed by Jacques Ellul: "Everything takes place as in a show, offered freely to a happy crowd that has no problems" (1990, 18).

However, if the addictive and festive dimension can be justified from a certain perspective, it is not obvious that cybergenetics in general, and DTC cybergenetics in particular, *fulfills its promises* on the scientific level – far from it. While this certainly does not mean that cybergenetics as such has no scientific value, it seems important to stress that we are currently witnessing a growing challenge to the way the DTC cybergenetic tests are conceived, for two principal reasons.

The first stems from the opacity of the companies' procedures. The same sample, sent to different companies, can give unaccountably different results, around 40% of which are incoherent, or false positives in the case of medical prognostications.[88] Furthermore, on the

[87] http://www.fyrclassroom.org/for-teachers/index.html.

[88] See especially Tandy-Connor et al. 2018, "False-positive Results Released by Direct-to-Consumer Genetic Tests Highlight the Importance of Clinical Confirmation Testing for Appropriate Patient Care"; American Society of Human Genetics 2019, news release October 17, "Researchers Quantify Limitations of Health Report from Direct-to-Consumer Genetic Tests"; and Imai et al. 2010, "Concordance Study of 3 Direct-to-Consumer Genetic-testing Services."

medical level, the connection between mutations and genetic diseases is not always rigorously established by commercial companies, since they do not take into account the possibilities of multiple mutations or the non-genetic dimensions that may lead to the development of such diseases.

The second reason has to do with the way the results of the tests focused on geographic origins are conceived and reported. Not only do the narratives confuse contemporary populations and nationalities, but the links established with ancient populations in order to trace a lineage rely on contemporary gene pools, as if the populations involved had not budged geographically and had not mixed together.[89]

As many commentators have noted, the story told by these tests is culturally marked to the extent that these quests are often based on a relatively recent historical span (around 500 years), a span that is directly aligned both with the history of American immigration – in an effort to return to one's roots that is more or less painful and more or less faithful to one's family history – and with the American cultural practices aimed at identifying individuals, even today, in terms of their ethnicity – a practice illegal and culturally shocking in France, for instance. Still, if these tests, at least in their Western version, retell the family story but also that of European and African populations chiefly through American lenses they do so not only for American populations but also for other populations whose stories they are equally apt to rewrite.[90]

Similarly, the scientific model used, whatever may be the presumed exemplary implication of the human and social sciences in the Human Genome Project, is based on an American-style scientific model from which the methods, procedures and nuances proper to the human sciences are absent. Here again, an "American-style"

[89] See for example Jobling et al. 2016, "In the Blood: The Myth and Reality of Genetic Markers of Identity," and Bonniol and Darlu 2014, "L'ADN au service d'une nouvelle quête des ancêtres?"

[90] It is of course quite likely that Asian tests in general and Chinese tests in particular tell a different story.

narration is called for, in the content as well as in methods and form. The problem is that the label "scientific" – which can be readily contestable, as we have seen – feeds into the recreational dimension even as it fossilizes certain types of representations. The vagueness of the term is thus valuable in the monetary sense, and also because the blurring of categories allows the DTC companies to reach a broad audience.

Thinking about this fascination with cybergenetics, the spit parties and the claim to connect us directly with our ancestors, one cannot help seeing a certain similarity with the spirit-summoning gatherings of the nineteenth century; we tend to forget that those seances were inspired, at least in some instances, by positivism and techno-scientific progress.[91] Did not Thomas Edison himself participate in that movement? The means may have been different, but the goals were somewhat similar: the reaffirmation of the bond among the participants and with their ancestors, along with a quasi-oracular power that allowed "scientific" forewarning of misfortunes to come in the form of predispositions to certain diseases. Blurring of categories and self-fulling prophesies thus go hand in hand, as was already the case with nanotechnologies.

2.4.3 *Curating of data vs. caring about/for/with data: a "factishization" of cybergenetics?*

One of the principal differences between spiritism and cybergenetics comes undeniably, however, from the *essentializing* dimension of cybergenetics: it is as if the interpretation of genetic data were acquiring a form of essentialness, of absolute truth and scientific objectivity, of concrete materiality, finally – whereas spiritism, for its part, presumably relied above all on the old oracular categories supported by science. As Steven Pinker wrote in his article for the *New York Times*:

> For better or for worse, people will want to know about their genomes. The human mind is prone to essentialism – the intuition that living

[91] See for example Edelman 2002, "Spirites et neurologues face à l'occulte (1870-1890): Une particularité française?"

> things house some hidden substance that gives them their form and determines their powers. Over the past century, this essence has become increasingly concrete. Growing out of the early, vague idea that traits are "in the blood," the essence became identified with the abstractions discovered by Gregor Mendel called genes, and then with the iconic double helix of DNA. (Pinker 2009)

The *concrete,* quasi material dimension of DNA is undeniably quite far removed from occult manifestations of spiritism, even when the latter lean on science.

This concrete dimension of cybergenetics is inseparable from its *objectivization.* Believing that one is in more or less direct contact with an ancestor is in fact quite different from discovering that one is, for example, 5% Black, 50% Irish, 30% Italian, and 23% Indian. This form of objectivization is based on a claim to scientific *objectivity* that is supported by the use of *statistics.* As Timothy Caulfield has argued, one can speak in this connection of a veritable "reification" of race.[92] This point was vigorously and somewhat urgently denounced in an open letter signed in 2018 by 67 scholars and posted on BuzzFeed, reminding their readers that race is a social construct and not a genetic given.[93] One might also add here that genetic data are not "givens," either: they are scientific results interpreted with the help of theories that are not exempt from viewpoints and interpretations. This is made clear by the fact that the origin stories proposed by the DTC genetic companies stem from a very American history, as we have seen. This point is crucial, for it is a constant in NBIC, and it shows up there in another way as well: not only do the DTC genetic tests see the world through the lenses of American culture and history, as it were, but owing to the phenomenon of globalization they can lead other peoples to reread their own collective and

[92] See Caulfield 2018a, "Is Direct-to-Consumer Genetic Testing Reifying Race?"; Caulfield et al. 2009, "Race and Ancestry in Biomedical Research: Exploring the Challenges"; and Caulfield 2018b, "Why Your DNA Test Won't Reveal the Real You": "The messaging surrounding this industry has the potential to facilitate the spread and maintenance of scientifically inaccurate and socially harmful ideas about difference."

[93] https://www.buzzfeednews.com/article/bfopinion/race-genetics-david-reich.

individual histories by putting on those same lenses unwittingly. In other words, here we have, in the absence of context, a potential form of both cultural and technological neocolonialism.

As we have seen, cybergenetics is particularly well articulated with the first three lines of analysis I have proposed. First, the context of its development and its definition blur boundaries. Second, it shares in a form of self-realizing anticipation. Third, the logic of risk is inadequate to account for what appears at first glance to be a recreation of identities, perhaps a *genethos*, even if the radical dimension of that recreation has to be tempered. These three elements will continue to be present in the analyses that follow, but it is now time to take up the fourth guiding line of analysis, which aims to show what the ethics and politics of care can contribute to thinking about cybergenetics.

A fruitful first step would be to establish a distinction between *curating* data and *caring for* data. My hypothesis is that cybergenetics wrongly passes off curation of data as a form of *care*. The notion of data curation is a classic one in the digital field. Broadly speaking, it refers to the organization and conservation of data throughout the entire cycle of their existence. However, I shall use the term in a slightly narrower sense, relying on Jenny Reardon's discussion – her use of the term stresses the absence of objectivity in the handling of data[94]– but I shall take it a little further. Thus we can associate curation with the selection of elements in order to give them coherence, or to make them tell a story; a curator of data can be compared to a curator who organizes a museum exhibit, or a designer who chooses objects for display in a high-end boutique of the "concept store" variety. Data curation is understood here as selecting and organizing data in a certain way in order to make them tell a story. It seems to me that races constitute the pinnacle of data curation, especially in DTC genetics, where the data (obtained via samples sent in the mail) have

[94] "Unlike objectivity, which carries along with it the expectation of a single true answer, curations are supposed to differ – indeed, their difference constitutes their value. This collection of paintings, of records, of genetic variants has been carefully selected by a discerning eye, not the 'God's eye' of a perspectival objectivity. This curation distinguishes the collection and makes it valuable" (Reardon 2017, 139).

no face, no individuality, but are immediately incorporated, in a de-individualized manner, into a whole that gives them their meaning. The donation of value is economic, in this case, but it can also be symbolic. Nevertheless, the symbolic dimension is secondary with respect to the economic dimension. The essentialization produced in curation is inseparable from that economic dimension: it is because people are being told who they are that they buy the product. For they think that the product determines or unveils their identity. Conversely, this form of data curation is probably capable of explaining in part why test results can vary so much from company to company: the data are probably not curated in the same way.

The essentialization produced by curation is inseparable from the reduction of individuals to their genes, something that is not found when data are handled with care, in the various forms of the term (care about, care for, care about etc.). The term "care *for* data," as understood by Clémence Pinel, Barbara Prainsack, and Christopher McKevitt, characterizes the relation between researchers and data at all levels when the data are directly collected and handled in such a way that they have a *face* as they are appropriated by researchers in their ongoing work.[95] This first appeal to ethics of care as a lens to analyze data is particularly interesting, but I believe that the idea of care could and should be mobilized in an even more fruitful manner to fertilize data ethics. In my approach to care in relation to data, five complementary aspects of care are combined. First, *caring about* data has to do with what is important, what matters, in a given data set. Second, *caring for* data has to do with the researcher's responsibility toward the data. Third, *caring with*, in relation to data in a democratic context, has to do with collective concern for issues such as privacy, control, and surveillance where personal data are involved. Fourth, special attention, special *care*, must be paid when dealing with data of *vulnerable* individuals. Fifth, it is also necessary to *take care* of the data in the sense that they should be processed with competence.

[95] See Pinel et al. 2020, "Caring for Data: Value Creation in a Data-intensive Research Laboratory."

The notion of care here explicitly includes relational and affective dimensions, in terms of donation of value. The value conferred by the researchers is both affective and scientific in the sense that the researchers are responsible for, and have to account for, the way in which they manage the data, in a scientific rather than a primarily economic mode. This responsibility thus functions in a relational rather than a purely substantial manner. It is this relational dimension, as I understand it, that gives care *for* data, about data, with data, and so on, its specificity.

The reduction that data curation entails was already pointed out – and challenged, as we shall see[96] – some time ago as a form of "geneticization." Generalized after its introduction by Abby Lippman in 1991,[97] this term designates the belief that genes are the expression of an essential nature of human beings, whose essence would thus be in a sense "contained" or "included" in the genome. This belief was gradually and very widely disseminated, from civil societies to international institutions, throughout the 1990s, most notably through the various political and legal conventions and tools developed around questions relating to human rights, which were viewed as prohibiting any modification of human genomes.[98] Geneticization is a bearer of various values having to do with scientific reliability and with a form of absolute truth. But it is also inseparable from a way of *reducing* individuals or groups to their genes – a practice that does not take into account the complexity of the field of genetics itself – and inseparable from a reduction of life to a sum of traits and risks against which there is hope of offering a degree of *forecasting* if not of *control*. What is

[96] For an excellent overview of the literature on this issue, see Weiner et al. 2017, "Have We Seen the Geneticisation of Society? Expectations and Evidence."

[97] See especially Lippman 1998, "The Politics of Health: Geneticization Versus Health Promotion"); Lippman 1993, "Prenatal Genetic Testing and Geneticization: Mother Matters for All"; Lippman 1992, "Led (Astray) by Genetic Maps: The Cartography of the Human Genome and Health Care"; and Lippman 1991, "Prenatal Genetic Testing and Screening: Constructing Needs and Reinforcing Inequities."

[98] See Rouvroy 2009, "Généticisation et responsabilité: Les habits neufs de la gouvernance néolibérale."

especially interesting here is that, as a way of reducing individuals to their traits, geneticization engenders a form of fatalism that is in part controllable. Turned toward the past via genealogy, geneticization produces a history that is presumed to be objective; turned toward the future, it offers an aspect of appropriation and control that transforms the conception of responsibility that is being implemented.

We encounter, here, a form of objectivation (understood in the double sense of objectivity and reification) or even a form of *commodification*, since, as we have seen, data are a gift of value monetarized by DTC companies; indeed, the acronym DTC can also be read as "Direct to Companies." People are transformed into things, into sellable data; in this sense, they are dehumanized. And this dehumanization plays out not through commodification alone; it is also incorporated in genealogical interpretations through percentages that are ultimately based on a logic of racial separation or even of racial purity. In this logic, races are taken to be discrete substances, whereas their mixing could be the object of a different interpretative grid and thus of a different narrative framework.

What is more, to return to the festive and often familial dimension of genetic tests viewed as the ideal gift, it is noteworthy that these tests are often offered by someone to whom you are important, someone who wishes you well. The DTC rhetoric also clearly aims to create the belief that the companies are going to look after you and let you know how you should look after yourself. In several respects, these tests play on the idea of care, understood not only as treatment but as solicitude on the part of others and solicitude for oneself. In the process, the businesses DTC companies exploit our desire for relationships and our desire for care, just as certain apps and certain forms of artificial intelligence do, as we shall see. Thus, paradoxically, it is our desire for relationship that brings us into a logic in which we are objectivized and chopped up into percentages. Our social qualities, which have helped us prosper as members of a species on the biological level, also have allowed us to build civilizations complete with tools such as cybergenetics and artificial intelligence – but these tools may well be able to turn against us.

Clearly, things are not so simple: as we have seen, this process also contributes to identity formation and subjectivization as Nikolas Rose (2007) also stresses. Subjects can discard, acquire, even transform information and can adopt various strategies for handling it, just as medical teams do[99] – activities that are incompatible with a "hard" concept of geneticization. Does this not mean, then, that we are at the heart of a paradox, finding ourselves conceptualizing simultaneously a form of objectivization and an active subjectivization? We must also note that in the shaping of new identities we endow the tests to some extent with mystical or fetishistic qualities: they allow us to connect with our own past (our lineage), our present (who we are, who are our cousins), and our future – our propensity for certain diseases, for example, or the qualities that our children may be apt to develop.

However, as Marie Gaille points out in "Pour un nouveau 'code de Nuremberg': De quelques enjeux contemporains du consentement" (2019), what needs to be reevaluated is not so much the element of paradox or mysticism but rather the fantasy of control, especially by the yardstick of "informed consent," which is often brandished in this context as a mantra. It is uncertainty rather than control that we need to acknowledge and consider inasmuch as it is "incomplete knowledge" with which we must "learn to deal collectively and individually," she writes, rather than with "false knowledge."[100] Here

[99] On this topic, see also Hedgecoe 2009, "Geneticization: Debates and Controversies."

[100] "A second aspect relating to the contemporary conditions of consent undoubtedly lies in a renewed vision of 'uncertainty.' Where knowledge is concerned, uncertainty is most often what one is seeking to reduce. One wishes to make informed decisions, and the more uncertain elements there are, the less well-founded the decision appears. Now the evolution of biomedical research implies, at least in certain of its aspects, that one is dealing with an irreducible uncertainty. Rather than making that situation of uncertainty an obstacle to decision-making in the field of research or in medical settings, it may be appropriate to learn to deal with it, collectively and individually: uncertainty is not false knowledge, it is incomplete knowledge which, despite its incompleteness, constitutes a precious element for grasping the stakes of an experiment or a medical decision" (Gaille 2019, 604).

again we find the idea according to which what I have proposed to call "philotechnology" must be able to relinquish the ideal of control in order to assume its "questing" dimension, just as philosophy is a quest for wisdom and not wisdom constituted as knowledge.

In order to get a better grasp on what is happening here, perhaps we can turn to the concept of "factish," invented by Bruno Latour ("On the Cult of the Factish Gods" [2010]). The term "factish" (French *faitiche*) allows Latour to overcome the dichotomy between fact (*fait*) and fetish (*fétiche*),[101] between the active and the passive. Latour compares this interface to the Greek "middle voice," which is neither active nor passive but both at once (Latour 2010, 56).[102] As he writes, "[T]he construal 'factish' authorizes us not to take too seriously the ways in which subjects and objects are conventionally conjoined: that which sets into action never has the power of causation – whether it be a master subject or a causal object. That which is set into action never fails to transform the action, giving rise neither to the objectified tool nor to the reified subject" (2010, 56).

These tests are in fact situated between subject and object, and the clients taking them are at once passive and active. Through the tests they are objectified, commodified, and dehumanized even as they act as subjects, transforming the tests' predictions or qualifications. In this respect, one of the most ironic examples is probably the one already mentioned in which the American far-right uses the recreational tests while loosening the category limits and reinterpreting the results.

As was the case with the nanos and will be the case with artificial intelligence, we see that cybergenetics cannot be conceptualized within the classic dualisms, whose boundaries it blurs. This blurring is summed up in the following chart, where the markers used for the nanos reappear on the left:

[101] *Translator's note: the words faitiche* and *fétiche* are homonyms in French.

[102] See also Latour 2000, "Factures/fractures: De la notion de réseau à celle d'attachement."

The Blurring of Dualisms by Cybergenetics

Nanos (dualisms)	Cybergenetics
Living/not living	Data-ization, commodification
Private/public	Sharing, participation
Active/passive	Factish

The concept of *factish* also has the advantage of pointing out the link between these tests and destiny: whereas one etymology of the French word *fétiche*, linking it with *fée*, "fairy," traces it to the Latin *fatum* (Latour 2010, 3-4), the word "factish" conveys the idea that my destiny may or may not be of my own making, insofar as it is presented as an objective fact, a given, a form of oracle that has been "made to speak."[103] Again we find the idea of self-fulfilling prophecy, as we saw with the nanos and will see with artificial intelligence. But a fetish also makes it possible, according to Latour, to call back into question the dichotomy between freedom and determinism, between autonomy and heteronomy. Finally, Latour says, what counts is less what separates us – into discrete, autonomous entities, we might add – than what connects us, and the value of the bonds that link us, that attach us: "Preceding the flag of Liberty, forever raised to guide the people, we would be well advised to carefully discriminate, among the attaching things themselves, those that will procure good and durable ties" (2010, 57). It is this attention to "good and durable ties" on which we shall now focus and to which we shall return in conclusion.

2.4.4 *From the reconfiguration of the concept of patient to the reconfiguration of the individual*

As many analyses have pointed out, one of the fundamental changes introduced by genetics in general – and by cybergenetics in particular – concerns responsibility. Two fundamental dimensions of cybergenetics serve to accentuate this change, which was already observable

[103] Here we also find, on Latour's part, a nod in passing to the fetishism of commodities denounced by Marx.

in genetics. On the one hand, cybergenetics, through its oracular dimension, introduces a new form of responsibility with respect to potential risks. On the other hand, through its relational dimension, it reconfigures our relationships, as it opens up to us a network, both vertically toward our ancestors and descendants and horizontally toward the persons, potentially in faraway places, to whom the DTC companies link us.

If early diagnosis has allowed the constitution of a new concept of patient, cybergenetics for its part transforms us all into patients "in waiting," well beyond the circle of those who might have "reasons" to consult a geneticist, for example in the context of a family secret or an interest in having children.[104] Where the old concept of patient emphasized the passivity of a person who *suffered* from some malady and was in a way *delivered over* to the paternalism of a medical team, the new concept no longer stresses either the fact of being passive or the fact of suffering (*patior*) from a malady; rather, it foregrounds the blurring of the passive/active relationship, as we have seen, and it focuses on time and patience, since the question becomes how we can act to "take in hand" our fate, which is endlessly reshaped by the prospect of preventative behaviors. The temporality of cybergenetics is totalizing: it encompasses the past, the present, and the future. The cybergenetic model thus has to be conceptualized according to the regime of chronic illness as it has been characterized by Anne-Marie Mol, in *The Logic of Care: Health and the Problem of Patient Choice* (2008). Mol depicts an ethics of care in which the dynamics of care is resolutely multidimensional and in which the subject receiving care is not an object but an active party, a member of the treatment team (although a somewhat special member, as the person primarily concerned): the team operates through adjustments and readjustments, with whatever tinkering is needed.

However, the redefinition goes beyond that of the patient alone; it operates on the level of the individual, by transforming us all essen-

[104] See Timmermans and Buchbinder 2010, "Patients-in-Waiting: Living Between Sickness and Health in the Genomics Era."

tially into patients in the sense in which we are all carriers of genetic anomalies: we do not have a disease, we *are* a genetic constitution. As William Stempsey suggests in "The Geneticization of Diagnostics" (2006): "If one's identity is to have a particular genetic constitution, and one's particular genetic constitution is in fact a genetic disease, then one's very identity is disease. We no longer have a disease, we are a disease."

2.4.5 *Rethinking relationships rather than autonomy*

Cybergenetics also puts the question of responsibility back into play by insisting on the importance of taking charge of one's potentialities and under some circumstances making them known to the persons with whom we "share" them. The response to a test is thus in the literal sense a respons-ability, a form of reactivity that concerns us not just as individuals but as participants in relationships. Conversely, the fact of sharing our genetic material also implicates those with whom we share it, whether we know them or not, whether they are already born, not yet born, or even deceased – and given the numerous DTC data banks, there is cause for concern. The popular French expression "my American uncle" takes on a paradoxical new meaning: instead of referring to an unknown wealthy uncle from whom one would miraculously inherit a fortune, it is on the contrary an unknown relative, an American cousin, perhaps, who has given his genetic data, but also some of yours, to the DTC companies!

Contrary to the stress placed on autonomy by the DTC companies, which appear to presume that individuals have control of their own health data since their testing has not had to be mediated by their medical team, we have to acknowledge that the genetic subject constitutes exemplarily, a subject positioned within a set of relationships. It is not autonomy but relationships that best define the genetic subject.

As we have seen, relationships with the persons with whom we share our genetic particularities – because they are affiliated with us on the biological level – are among the fundamental ethical and political stakes of DTC genetic tests. The most immediate case of course would be genome sequencing by your identical twin. This happened to a woman named Arielle, when her sister Samantha decided, without Arielle's consent, to order a complete sequencing of her genome.[105] Ignorance of the implications of the action was not a factor here: the sisters had known that they were monozygotic twins since a genetic analysis undertaken when they were fourteen years old, and both of them worked in the field of genetic research. Nevertheless, Samantha felt that the decision was hers to make, since it was "her" genome that she had sequenced; she said she felt as lucky "as Charlie winning the golden ticket to Willy Wonka's chocolate factory."

But beyond the case of identical twins, which remains fairly rare, the issue is a more general one, because it concerns the whole range of people with whom we are related genetically: past, present, and future family members, known and unknown. Thus the question of responsibility needs to be raised in familial and even generational terms.[106] While the problem arises in principle for all genetic analyses, it seems to me much more acute in the case of cybergenetics, and "recreational" cybergenetics in particular. In a medical framework, we can hope that the information will be relatively well protected, but in a commercial framework it is clear that the goal is precisely that they be protected as little as possible, and perhaps not at all. Thus genetic data linked to genealogical and digital data concern not only the person taking the test but all of that person's more or less direct kinship relations.

This responsibility takes different forms for women and men, not only in cases of uncertainty about paternity but also in terms of gender, because women are more often "in charge" of the well-being of their family, as Ulrike Felt and Ruth Müller have noted (2011, 346). The mental and physical burden of genetic responsibility would be associated biologically and socially with women, in terms of both sex and gender: in terms of sex, because the woman is considered the

[105] See Schilit and Schilit Nitenson 2017, "My Identical Twin Sequenced Our Genome."

[106] See Wallace et al. 2015, "FamilyTree and Ancestry Inference: Is There a Need for a 'Generational' Consent?" and Caulfield 2002, "Genetics, 'Family Consent' and the Law."

primary person concerned with the burden of reproduction, and in terms of gender, to the extent that the woman is socially considered responsible for the family's well-being.

At the same time, synthesizing genetic data with other types of data produces an extraordinary amount of information. For this reason, "pseudonymizing" (in which one replaces the name of an individual with a pseudonym but retains all the other elements, especially those concerning time, place, and relationships), which is particularly useful for following an individual or a group, is entirely utopian as a justification for the widespread claim that anonymity will be maintained in the use of genetic data. As a number of articles have made clear, such a claim is entirely illusory. As Heather Widdows emphasizes in *The Connected Self* (2013), the anonymization of genetic data is characterized precisely by its *reversibility*: the process can readily be reversed or thwarted, especially where genetic elements are concerned, as these are identifiers by their very nature.

The term "dataification" has been proposed by Viktor Mayer-Schönberger and Kenneth Cukier in *Big Data: A Revolution That Will Transform How We Live, Work, and Think* (2013) to bring to light this transformation of a growing number of aspects of our lives into computable data that can potentially be interconnected. The fact of synthesizing these data can make certain elements significant that would not be significant on their own; this makes it hard to determine in advance which data are significant and which are not: in principle, nothing can be deemed anodyne. This means that the argument according to which one has "nothing to hide" becomes devoid of meaning, for the question is not that of "nothing" but rather that of "everything." We have already encountered this issue in the context of curation: while every element, taken separately, may seem insignificant, one cannot know in what set it will find a place, or how.

2.4.6 *A CyberGenoPanopticon?*

It is precisely because DNA makes it possible to identify not only its source but also members of the source's family that it constitutes a resource of choice for the police, and that the DTC companies' websites have become police auxiliaries by "diverting" their primary aim and using relationships as vectors of identification in reverse – this is sometimes called "inverted genealogy." The approach has been used to solve several crimes, of which the best known is that of the Golden State Killer, discovered thanks to the work of genealogist Barbara Rae Venter (ex-wife of Craig Venter, the founder of *Illumina*). The Golden State Killer, long sought under that label by the American police, was a serial rapist suspected of twelve murders as well as around fifty rapes in the 1970s and 1980s. The police had the idea of comparing the genetic samples in their possession not to those in their own files but to those of a DTC genetics company. After a first false lead, they continued to explore that solution by using a genealogical site where people could upload their own data, called *GEDmatch*. On the basis of 67 genetic markers on the suspect's Y chromosome, they were able to identify about a hundred people genetically linked to the suspect. By synthesizing these data with others, in particular civil registries, they identified a principal suspect. After that, it was child's play to verify his identity by taking DNA samples from his car and from a handkerchief, before arresting and indicting him.

The Golden State Killer case, and the flurry of media attention it provoked,[107] demonstrates how cybergenetics can, by virtue of its relational dimension, lead to practices of surveillance and control of populations in a sense that is different from but complementary to those I have evoked up to now: its dynamic relation to healthist and performative logic takes us beyond Foucault's biopolitics to Rose's ethopolitics and to Deleuze's society of control.

Control is exercised here in two directions: first, vertically, for a single individual, and then horizontally, by integrating the links among individuals.

In the first case, by combining data from within (genetic data) with data from without (most notably digital: localization,

[107] This case stimulated a broad public debate on the collection of personal data and their use by DTC genetic companies.

preferences, narratives, clicks, and so on), one obtains a precise and nuanced vision (a portrait "in relief") of each individual involved, even in the most private spheres. This possibility is magnified by the use of apps that allow us to take the digital sphere everywhere we go, like an extension of ourselves, a prosthesis. The Pheramor dating app is a good example: using the results of genetic sampling and an investigation into digital traces, the app proposed to locate your soulmate without wasting any time (according to an ever-more-performative logic) thanks to a proprietary algorithm, while reassuring its clients that no one ("not even you yourself") would have access to their data.

In the second case, with information about genetic linkages and kinship networks, control is exercised through a different mode of identification, as was the case with the Golden State Killer. This strategy is all the more interesting in that, as we have seen, the investigation carried out by a first company had come up with a false lead. However, convinced that the method was a good one (perhaps because genes are presumed not to lie?), the police started over with a different company. Of course, when they uploaded their genetic data onto an Internet site, the users did not intend to supply genetic information to the police – such information is supposed to be under strict controls when they are in police files. In most cases, the first reaction of the DTC companies was marked by some embarrassment at the way the Golden State Killer case brought to light certain elements intrinsic to their databases, and they hastened to reassure their clients with messages such as "your data belong to you"; still the fact remained that these data could be hacked or "opened" owing to a glitch (intended or not) in the system.[108] After this first awkward situation, most of the companies committed themselves to helping the police as much as possible by using – even exaggerating – the

[108] In July 2020, GEDMatch was hacked, and a million people who had refused to allow their data to be shared with the police found themselves exposed. It is not known to what extent the leak was organized by the company or the police, but the question has come up. See Aldhous 2020, "A Security Breach Exposed More Than One Million DNA Profiles on a Major Genealogy Database."

already well-entrenched thematics of "sharing," "nothing to hide," even of "superheroes," as is the case with the company DNABargains: in its advertising, it uses the slogan "Want to Be a Crime Fighter?" along with a picture of a child in a red and blue superhero costume that leaves no ambiguity about the reference to Superman.[109]

The "vigilante"advertising of DNA Bargain[110]

Here we again encounter between the lines, although via a very different path, the idea that cybergenetics gives us access to something bigger than ourselves. Rather than flatly offering an exploration of our superpowers or a logic of hyperperformance, it evokes the possibility of participating in something good by joining the fight – a nonscientific fight – against crime, following the model of the vigilante superheroes who operate as auxiliaries of the police. It is

[109] My thanks to Lisa Gannett, who introduced me to this ad. Conversely, the Innocence Project seeks to use DNA in its efforts to rehabilitate the memory of innocent people who were condemned in court, although the undertaking strikes me as quite different, of course. See https://www.innocenceproject.org.

[110] The URL for this out-of-date advertisement is no longer active.

precisely the combination of all these aspects aiming to produce the impression of going beyond our own individuality, even while exacerbating it in a performative fashion, that undoubtedly constitutes one of the principal persuasive powers of cybergenetics.

The strategy of the police in the Golden State Killer case falls within a collection of genetic samples that had already been denounced as a "genetic panopticon," with reference to the circular structure that allows those in power to see without being seen (and that gives those seen an internalized impression of being permanently under observation) that was proposed by Jeremy Bentham and popularized by Michel Foucault. With the uses made of genetic information by the DTC companies, one more step has been taken in the direction of the panopticon, not only in terms of degree but perhaps also in terms of nature. Not only is the use of these databases for police purposes a deviation from the use to which the clients agreed initially, but it also allows the companies to go beyond the strict framework of the police toward murkier practices. The combination of genetic data with digital information from other databases could at least potentially allow tracing for a wide variety of purposes.

The hypothesis of a genetic panopticon was put forward by Anne Brunon-Ernst, in the wake of the 2013 King v. Maryland case, where a defendant had undergone a DNA test before being declared guilty.[111] If the precedent set by this case is followed, any person suspected or questioned in the context of a legal case could have genetic records kept on file. The subsequent generalization of DNA testing by American police forces and the constitution of a record containing the DNA data of people who have been suspected and questioned but not found guilty are both problematic outcomes. However, as Brunon-Ernst demonstrates, the subtlety of Bentham's model is such that this type of procedure cannot really be considered as proceeding from

[111] See especially Brunon-Ernst 2014, "Beyond the Genetic Panopticon: The Limits of Government Intervention on Citizens' Bodies." See also Brunon-Ernst 2012, "Deconstructing Panopticism into the Plural Panopticons," and Brunon-Ernst and Tussaud 2012, "Epilogue: The Panopticon as a Contemporary Icon."

a panopticon, above all because Bentham aimed his model more at prevention and rehabilitation at than punishment.

Similarly, digital surveillance is characterized by a form of "catopticon," to use Jean-Gabriel Ganascia's term, because the oversight is not exercised from above but from below.[112] Moreover, this form of surveillance is most often participatory and articulated with a desire to be seen and recognized. The use of the term "panopticon" with respect to digitized information is thus far from self-evident, and several other designations have been proposed, such as "sousveillance," "reverse panopticism,"[113] or "omniopticon,"[114] as a way of marking the specificities of this generalized digital monitoring.

Things are more complex with cybergenopanoptics, because the conversion takes place simultaneously within and without, owing to the internalization of norms in what Nikolas Rose, as we recall, has proposed to call *ethopolitics*. The healthist component of the genetic DTC companies' project and the capturing of data by these companies are transforming the landscape by articulating the dimension of participation with that of oversight, combining an internalization with an externalization.

The client of DTC companies is thus a *client* in the literal, etymological sense of the word, an individual inserted into a complex economic and political relationship.[115] Clients offer their data to the patrician who "pays" them in advice, or even in social recognition, by allowing them to participate in something bigger than themselves and thus reflecting back to them a certain self-image. In this way the

[112] See Ganascia 2009, *Voir et pouvoir: Qui nous surveille*.

[113] See Mann et al. 2003, "Sousveillance: Inventing and Using Wearable Computing "Devices."

[114] See Mitrou et al. 2014, "Social Media Profiling: A Panopticon or Omniopticon Tool?"

[115] According to the *Oxford Advanced Learner's English Dictionary*, s.v. the word "client" originated in "late Middle English: from Latin *cliens*, *client-*, variant of *cluens* 'heeding', from *cluere* 'hear or obey'. The term originally denoted a person under the protection and patronage of another, hence a person 'protected' by a legal adviser". For a complementary analysis of the issue of the patient as client, see Mol 2009.

DTC companies give their clients the impression that the companies care about them and are paying attention to them. However, this is only an appearance of care, which leads in fact to a very subtle form of domination and weakening, as we shall see.

This positioning of the client as vassal is complemented by a form of voluntary servitude, materialized by self-monitoring technologies such as activity trackers, paramedical apps, "smart" scales, and so on, all of which allow algorithmic measures of individuals to feed into intersecting data bases. As a constituent element of these modern techniques of self-qualification and self-quantification, cybergenetics thus offers a perverse deviation from utilitarianism, in which the greatest good for the greatest number is turned into an economic rather than a moral imperative, by way of the performative injunction to be "the best version of oneself" among one's virtual avatars.

2.4.7 *For a subject in relationships*

This injunction to perform brings us back to the question of what freedom is left to clients or seized by them, in an undertaking that the DTC companies often try to present as a form of empowerment. The problem is how to distinguish among various forms of freedom: if information offers an agent a capacity to act that may at some point be expanded, the freedom to act may nevertheless diminish that agent's well-being if it is not accompanied by an actual increase in the agent's power to act in a desired direction.

The issue here entails neither autonomy nor control. To paraphrase Bruno Latour once more, it is not a matter of the alternative between freedom and determinism, or between control and *laissez-faire*, or even between subject and object, for we often find ourselves, as we have seen, in an intermediate space where freedom is tempered and belief in determinism, even if it increases, remains the object of readjustments and negotiations (Latour 2000). Inasmuch as it is *factish*, cybergenetics invites us rather to rethink relationships in the context of a renewed form of liberty and responsibility. As Latour puts it, "freedom becomes the right not to be deprived of the bonds that make [our] existence possible" (Latour 2000, 202).

One of the pitfalls of cybergenetics is that, unlike e-medicine, with which it might be confused, it is subtended by a more or less deliberate healthist and clientelist logic that entails a social project in which 'connexions' take priority over relations. This pitfall is particularly obvious in the way biological connexions are highlighted as offering new kinships. It is not a matter, here, of adding so-called weak ties,[116] but of a weakening of ties, which is not the same thing. Under the appearances of empowerment and liberation, the DTC companies restore old binaries that in reality hide the submission of their clients to a new paternalism, as evidenced by the way the companies prioritize "directness" at the expense of mediations, thus concealing an inegalitarian relationship.

To counter this new submission, if we turn to the ethics of care, we can dismantle the binaries and instead promote subjects in relationships. Such a position lies at the intersection between ethics and politics, because it makes it possible to integrate cybergenetics, provided that relationships are valorized rather than replaced or subjected to new networks of domination. This is a particularly important point, because cybergenetics is in a position to play a significant role in a much broader context, one that Niobe Way and her colleagues, among them Carol Gilligan,[117] call loss of connection but that I prefer to call loss of relationships, as I will develop in the following chapters.

The problem can then be envisioned from two quite different angles. On the one hand, we are caught in a tangled network of "donation" (or appropriation) of our data, one that can lead to a

[116] The notion of "weak ties," proposed by Mark S. Granovetter (1973) in "The Strength of Weak Ties," refers to occasional social relations that do not require much investment. This theory has been mobilized to analyze digital sociability, even if that was not its initial aim. On this theme, see especially Gefen and Laugier 2020, eds., *Le pouvoir des liens faibles*.

[117] See Gilligan and Snider 2018, *Why Does Patriarchy Persist?* See also Way et al. 2018, *The Crisis of Connection: Roots, Consequences, and Solutions*. During an event celebrating the publication of her book in French translation, I asked Carol Gilligan about her use of the term "connection," and she responded that it could in fact be replaced by the term "relationship."

cyberpanopticon, as we have seen. On the other hand, this state of affairs might well encourage us to envision the situation as requiring a form of solidarity. This is what Barbara Prainsack has proposed, as she militates in favor of a relational and connected approach to data management.[118] Here the notion of "personalized medicine" risks masking an increase in inequality on the health front, a situation in which each patient could be caught up in an organization consisting in distinct strata corresponding to diverse types of treatment and monitoring, some of which might be paramedical and sometimes inseparable from the healthist approach, as we have seen. The whole problem, then, is how to exit from a form of paternalism and head toward real individual empowerment, but also to make sure that that empowerment does not end up in a form of self-centered egoism, not only because that would be morally problematic but also because it would deny the meaning of this relationality.

Thus it seems to me that the notion of empowerment is not necessarily unambiguous here, for two reasons. First, the model of empowerment presupposes that the patient is "always on," always plugged in and transmitting or receiving data, as Prainsack supposes (Prainsack 2017b, 48-50). The reference is thus always to performance, and individuals find themselves caught up, most often unwittingly, in the process that I have proposed to call *genethos*. Second, the model of empowerment raises the question in terms of strength and of power relations, yet we also need to know whether the individual concerned is or is not in a position to exercise or benefit from the power that is being offered – when medical information is being provided, for example. As we have seen, the model of the DTC presupposes that clients are given direct possession of the results that concern them, without filters and, most importantly, without the expertise of medical professionals. The empowerment resides here in the presumption

[118] See for example Prainsack 2017a, "Research for Personalised Medicine: Time for Solidarity," and Prainsack 2017b, *Personalized Medicine: Empowered Patients in the 21st Century?*

that clients are masters of their own health data – even though, as we have seen, the situation is far more complicated than that.

But perhaps the problematics of domination and possession is not the most appropriate one here. Perhaps it would be more useful to ask – as Amartya Sen did when, in *The Idea of Justice* (2009), he reproached John Rawls for not raising the question of the concrete ways in which justice is carried out – what makes individuals able and willing to exercise control over their own data. This is why the term Sen proposes, "capability" – which designates, in sum, the possibility of choosing effectively and of trying to bring into being the way(s) in which one wants to lead one's life – might seem more pertinent here. In any case, the relational dimension raises the question of capabilities in non-individualist terms, and this leads to tying together capabilities and solidarity in a relational perspective, a perspective that we shall encounter again in the conclusion through the notion of relational responsibility. The question of capabilities allows us to take up the question of choice and of the effective, concrete dimension of the exercise of justice. Nevertheless, it does not necessarily encompass the question of relations that remains central here. This is why it is important to return to the latter issue – relations as distinct from mere connexions – as engaging a particular ontology. As Fiona Robinson emphasizes (2011, 28), uncovering that relational ontology is all the more important in that it also allows us to discover new possibilities and to stress the relational dimension of responsibility. One of the key conclusions of the present chapter is thus that the operative distinction between individual + collectivity + society *in connexion* and individual + collectivity + society *in relation* has been pivotal for analyzing NBIC. We shall continue to pursue this question of connexion versus relation in the second part of this book, where we shall focus on artificial intelligence.

Part Two: Artificial Intelligence and Ethical Experimentation

CHAPTER 3

Is an AI Ethics Possible?

The second part of this book completes our exploration of the ethical stakes of NBIC by addressing the question of artificial intelligence (AI). It is organized around the four major lines of analysis developed in the preceding chapters: (1) blurred boundaries, (2) the problematic position of anticipatory design and self-fulfilling prophecies, (3) the inadequacy of the risk/control framework, and (4) the productive dimension of care. As before, these lines are apt to become intertwined; this is why they are not always dealt with in the same order from chapter to chapter, and why it is important to set forth a general approach before analyzing selected examples and then returning to an overall analysis. In the first chapters in this second part, I seek to establish the stakes of the problem and to propose an overview of the situation – hardly a simple matter.

3.1 Beyond the boundaries, or toward new boundaries?

As we have seen in the earlier chapters, one of the difficulties that arises in attempts to define the field under study is not only that the field – AI, in this case – does not respect the boundaries between the scientific-technological and the social realms, but that it also takes the form of experiments that cannibalize ethics and politics, as it were, by subjecting them to its own logic. In this way it creates new *sui generis boundaries* that can appear to reopen the question of social acceptability while at the same time blurring the distinction between social acceptance and moral acceptability by advocating the artificialization of morality.

3.1.1 *Defining artificial intelligence: an ironic quest*

In an initial approach, artificial intelligence can be characterized as a research program that, from the second half of the twentieth century on, has been seeking to reproduce and imitate the mechanisms of natural intelligence in a way that allows machines to be programmed to carry out more or less specialized tasks. More recently, researchers have been pursuing the idea that AI might make it possible to solve certain problems better than humans can.

Defining artificial intelligence is probably a more arduous task than it may seem at first glance. Given that we live in a world in which AI is omnipresent, we probably all think we know more or less what it is. Yet specialists disagree about its definition. There is a tendency to characterize it as a "research program" in order to stress simultaneously its interdisciplinary aspect and the variety that emerges in both its basic and applied dimensions. This is why, even though we are so used to turning to AI to answer our questions that some may be tempted to ask "OK, Google, what is artificial intelligence?" while others may go directly to Wikipedia, it may be more helpful to rely on a bit of "natural" intelligence and *question* the definition rather than taking it as already established. In fact, just as was the case with nanotechnologies and cybergenetics, we can see how the tensions that arise around the definition of AI constitute an interesting problem in themselves for the perspective I am adopting here.

I could begin, as others often have, by proposing to characterize AI as the intelligence developed by certain artifices, namely, computing machines, programs that can be implemented in robotic systems. But things are not that simple. If the "artificial" aspect of intelligence can be characterized more or less readily, what "intelligence" are we talking about? Alongside laughter, intelligence has for a long time appeared to be a distinguishing characteristic of human beings, even if many debates in ethology are bringing new challenges to that characterization. Still, keeping in mind the origins of AI, we could propose that it consists in making machines that simulate or stimulate the human mind, that is, machines that do what the human mind does, and do it in some cases even better than humans. The question, then, is what cognitive functions may be involved: perception, language, mathematics, logic, and so on. The fact that AI is often used for decision-making and problem-solving seems to raise the level of complexity a notch.

The origins of AI are in part philosophical, going back to Descartes, Pascal, and Leibniz in particular. Descartes' analysis of the notion of representation is often cited as one of the origins of the study of cognition and of the attention given to the processes of acquiring, shaping, and preserving knowledge. Leibniz and Pascal went further, showing that it was possible for a machine to reproduce our basic mathematical reasoning, and thus that a machine could perform automatically (that is, without human intervention, once the information to be dealt with had been entered into the machine) the fundamental operations of arithmetic: addition, subtraction, multiplication, and division. The revolution was both theoretical and technical. Technical, because it allowed humans to begin to free themselves from certain tasks that could be handled automatically; theoretical, because people were beginning to imagine that, in a way, the human mind could be conceived as a machine that handles some kinds of information automatically. The problem that arose, then, was to understand to what extent machines could substitute for or even replace human intelligence, and in what specific tasks it could do so. To a large extent this same problem confronts us still today.

Cybernetics is generally said to have been born in 1948 with the work of Norbert Wiener, who theorized the notion of feedback in relation to artificial intelligence. He showed that, for a system to deal with information internally, it had to be able to adjust its own behavior in response to the information received. A double movement was thus involved, since the data treated by the system is in a sense sent into the system in order to adjust the system's behavior to the data being treated. Thanks to this notion, we can envision another dimension of thought, one that could be reproduced artificially. Still, it is often forgotten that Wiener was interested in the social aspects of cybernetics, as evidenced in his book *The Human Use of Human Beings* (published in 1950 and revised in 1954). This is all the more noteworthy in that American society at the time was marked both by the shock the bombing of Hiroshima created in the scientific community and by the ambient McCarthyism. From the outset, Wiener related cybernetics to social issues by raising the issue of the dehumanization of human beings. He also linked the notion of cybernetics to its etymology (it derives from a Greek word meaning "steersman," "governor," "pilot"): an interesting fact in the context of certain contemporary projects based on artificial intelligence that aim to develop "self-piloting" or "driverless" cars.

Many theorists today distinguish between "weak" and "strong" AI. The weak version simply mimics human intelligence, generally for the purpose of performing routine tasks, whether on the basis of rules of inference or on the basis of formal or networked neurons, although it sometimes designed to undertake more demanding work, of the sort Meredith Broussard ironically calls "statistics on steroids" (2018, 22). The strong version, on which we shall focus more closely, is supposed to surpass human intelligence and eventually to acquire self-awareness. While the latter aim has not yet been achieved, it has been the object of major debates and of statements implying that it is already an accomplished fact. This latter is a real problem to which we shall return. In general, the AI we encounter most frequently in daily life – for example, in research results or social networks – stems from the weak form. Nevertheless, developments in AI have made it possible to move beyond algorithmic collations of expert systems to techniques of learning that rely essentially on a body of experience or accumulated data. This type of system can then be used in a "cooperative" procedure in which AI "learns." The AI involved adjusts its parameters continuously during the learning process; these adjustments enable it to function by inference. The adjustments work differently depending on which of three modalities is applied to the learning process. (1) If the learning is supervised, the categories used to organize the data are predefined. (2) If it is not supervised, then a first phase consists in categorizing the data. (3) Learning through reinforcement entails direct interaction with externally provided data rather than with an initial already-constituted data base.

Finally, with the emergence of *deep learning* around 2005, the number of neuron layers, or basic elements of calculation, has continued to increase. Each neuron is linked to the others during the learning process by adjustable connections. The resulting neuron networks are widely used to handle data – such as images – that are highly complex on the spatial and temporal levels, with limited success, as we shall see. One of the difficulties of deep learning is that the learning process is not necessarily predictable, so that one cannot know what the output of the process will be. AI in general, and deep learning in particular, relies on statistical regularities or patterns. This is the case, for example, with translation software, which establishes equivalencies on the basis of these regularities. But the patterns can also be used in various ways in the field of ethics, for example to try to program machines capable of proposing a "moral code," or of guiding military drones. Contrary to a common belief, AI is not hyperindividualized, or "made to measure"; it operates primarily through patterns and categorizations. Thus to recommend content to users, for example, shopping or streaming platforms collect the choices made by individuals and group them in fictitious categories. The contents are personalized in the same way that "personalized" medicine purports to be: by labeling each individual as a member of a particular group. In any case, the resulting impression of proximity and individualization corresponds to a description we have seen before, in Jacques Ellul's "encirclement by what is obvious" (1990, 18).

Finally, since the late 1990s, "emotional" information science, which is based on the recognition of human emotions, decision-making (along with the reasoning that takes that information into account) and the generation of affective expression by machines have opened the way to approaches and designs that, combined with facial recognition and/or deep learning, make it possible to create artificial intelligence software and relatively perfected robots destined to be our "companions," from virtual assistants to social and emotional robotic companions (or at least labeled as such) with multiple uses, from babysitters to pets for the elderly.

It may be useful here to recall that the rise of AI began in the 1930s and continued through the 1950s: its development was for the most part linked to the Second World War and the need to crack enemy codes. The challenge at the time was how to extend the mechanization of reasoning. This was the task that Alan Turing, among others, undertook starting in the 1930s. In "Computing Machinery and Intelligence" (1950), Turing set forth the idea of a "computing machine" (now known as a "Turing machine") consisting of a tape divided into cells containing various data, a read-write head, and a table of rules that governed its behavior. The rules governed the steps that the machine was to take at each moment in relation to the internal state of the machine and to the cell that had reached the head. Depending on how the internal state of the machine and the position of the head were set up at the start, the machine would handle the data contained in certain cells automatically. The distance between this device and Pascal's calculator is evident. Turing postulated that for all well-defined formal operations that the human mind could perform (solving arithmetic problems, for example, but going beyond the four elementary operations that had been the earlier limits), a computing machine existed that could handle these manipulations correctly, provided of course that there was a program permitting adequate regulation of the machine. Some have even asserted that Turing thought a computer could *pass for a human being*, could imitate and thus replace a human individual (a view that American philosopher John Searle rejected with the celebrated Chinese Room argument[1]). Today, however, the situation appears a little more complex.

[1] See "Minds, Brains, and Programs" (Searle 1980). In a (belated) response to the challenge Turing posed in 1950, Searle offered the counter-argument known as the "Chinese room." He asked the following question: can thinking be reduced to the simple manipulation of symbols? In response, he proposed a thought experiment. He imagined being in an isolated room where someone passed him questions he was

As Juliet Floyd showed in "La quête culturelle: Revisiter le test de Turing" (2019[2]), the structure of the Turing test is less psychological than "logical and social." Floyd reminds us that the Turing test was part of a social experiment and ended with a discussion among the participants. She also stresses, on the basis of an unpublished text titled "Intelligent Machinery" dating from 1948 (that is, two years before the book Wiener devoted to social questions), that Turing defined intelligence as sensitivity to three types of research. First, intellectual quests, which concern algorithms and computers; second, biological, genetic, and evolutionary quests; and third, cultural quests, which concern the techniques that human beings transmit from one generation to the next. Floyd sees cultural quests as fundamental, bringing together the two other types of quest, in a way. She writes: "Turing had predicted what happened: computer science, as a field, was going to become not only a branch of mathematics and of logic, or of the social sciences, but also a field of philosophical experimentation; an enterprise in which humanity as a whole, over a long period of time, would consider that it was engaged, a profoundly social and cultural quest, and consequently political and ethical" (2019, 17). Turing, of course, did not refer to NBIC, but he already grasped the significance of digital and genetic quests. What is more, he connected them with cultural quests and emphasized, in his conclusion, the importance of bringing intelligence and emotion together without situating the cultural quest below them.

This point is essential if we take into account the fact that artificial intelligence is not only at issue in scientific or technological

obliged to answer – but the questions were written in Chinese, a language he did not know. He had access to a program written in his mother tongue that allowed him to manipulate the Chinese symbols and thus to answer the questions without understanding their content. According to Searle, this situation was exactly the same as in Turing's experiment, and what was at stake in the test consisted in being able to supply the appropriate information either while understanding it (which would be the case if the person in the room understood Chinese) or while not understanding it. There are grounds of course for concern about the limits of this conception of thought as being a pure manipulation of symbols.

[2] For an English-language version, see Floyd 2022, "Revisiting the Turing Test."

endeavors but also, and perhaps even especially, in social, moral, and political contexts. Furthermore, it is being developed today with minimal guardrails and without clear knowledge of where it is headed. In *Alone Together,* Sherry Turkle, one of the finest analysts of what she calls the "robotic moment," writes from the privileged perspective she has at MIT: "We have agreed to an experiment where we are the human subjects" (2011, 296), and, more pointedly (in *Reclaiming Conversation: The Power of Talk in a Digital Age*), "[w]e have embarked on a giant experiment in which our children are the human subjects" (2015, 252).

A fairly recent awareness of the ethical and political dimensions of the issue has resulted in the creation of myriad discussion groups among specialists on these matters within national and international organizations.[3] It seems urgent today to envision the ethical, social, and political experimentation that takes place in the development of AI as implicit experimentation carried out on a given society and the individuals that constitute it. It is highly possible that this experimentation will change the way we experience our social, moral, and political life, our sense of ourselves, of others, and of our surroundings. The term experimentation is thus to be taken in the sense of trial or testing in the context of lived experience.

The idea according to which new technologies must be envisioned as experimentation is not a new one, as we have already seen. As Philip Brey notes in "Ethics of Emerging Technologies" (2017), the case of emerging technologies, in particular when they are enabling,[4] is often approached via the questions of uncertainty and risk. But focusing on uncertainty and risk often reduces the ethical question to a simple cost/benefit calculation, which is only one part of ethics,

[3] Moral philosophers are rarely participants in these groups, however – perhaps because the specialists see no point in involving specialists in the human and social sciences in discussions of these important scientific questions, and/or perhaps because some moral philosophers find the scientific and technological matters too daunting to want to be involved.

[4] That is, when they can be combined with other technologies and/or be developed in diverse fields, as we have seen with the nanotechnologies, for example.

often in connection with other approaches (economic in particular). To broaden the discussion by focusing on ethical questions themselves, one might propose to take the uncertain dimension of AI seriously and consider the extension of its use in our societies as social experimentation on a very large scale, which some call human-machine coevolution.[5]

Moreover, a supplementary problem is raised by the development of deep learning, because, as we have seen, it is often hard to determine how the system will evolve and what its output or results will be. It is thus not simply a matter of knowing whether the technology in question is morally acceptable – and under what conditions – but also of asking whether it is morally acceptable to try out the technology experimentally in our societies, or under what conditions that might (or might not) be the case.

3.1.2 *Between "Gabor's law" and the Collingridge dilemma: How to ask the question of control (or not)?*

The problem posed today by the question of an AI ethic – understood in the double sense of an ethics of AI and of ethically acceptable developments of AI – can be envisioned at the crossroads between two classic conceptions in the ethical analysis of sciences and technologies: the Collingridge dilemma, which I mentioned in chapter 1, and the so-called Gabor's law, evoked in the introduction.

According to what is sometimes mistakenly called Gabor's law, as we have seen, "if a technology is available, it will be used."[6] Let us recall that the Collingridge dilemma (Collingridge 1980) is a double bind. On the one hand, early in the development of a technology its consequences cannot be predicted; we are thus in a situation of insufficient knowledge. On the other hand, when a technology is advanced and integrated into society, it is harder to go back and modify it and

[5] Those who use this expression appear unconcerned about the fact that it puts humans and machines on the same plane, or that the term "evolution" is being used to characterize machines.

[6] A variant formulation reads "if we don't do it, someone else will" (because it is technically possible).

its social integration; here it is action rather than knowledge that is inadequate. We face a dilemma, since neither of the two directions is satisfactory: upstream, we cannot yet know (even if we can act); downstream, we can no longer act (even if we do know).

To say that we are at a crossroads today between the Collingridge dilemma and the so-called Gabor's law is to stress the importance of not waiting until everything that is technically possible has been achieved to take the question of AI ethics seriously: this means accepting the risk of not being able to predict all possible consequences. However, we should recall Jean-Pierre Dupuy's insistence that unpredictability is situated right at the heart of the metaphysical project of the convergence of NBIC (Dupuy 2007). In other words, even if it is likely that the longer we wait, the more clearly we will be able to see the impacts of artificial intelligence, it is also likely that we will never be in a position to predict all possible consequences.

At least at first glance, one might suppose that the issue here is the classic one of choice against a background of uncertainty, as it already exists in other technoscientific fields. Nevertheless, artificial intelligence is a special case, for at least two reasons. First, as a convergent cluster of studies emphasizes, AI is (already) so integrated into our public and private lives that it is modifying not only our personal identities but also our social and political structures. The question of forms of life we encountered with cybergenetics arises with artificial intelligence as well.

Next, there exists today an angle of approach by way of "moral machines," an approach that could lead to transferring ethical decisions to AI.[7] This is why the Collingridge dilemma, while it retains

[7] Here I should note that France is a special case, because, although it is under the umbrella of fairly protective European regulations, it has recently developed a controlling strategy by involving AI in all sectors of public life: education, especially via Parcoursup (a French Web portal designed to distribute candidates for university admission in a way that balances their interests against the places available at the various institutions of higher education), health, especially via the Health Data Hub, which we have already encountered, and access to public services, especially via Alicem (a tool, currently in the testing phase, for accessing public services that uses facial recognition technology: see https://www.interieur.gouv.fr/Actualites/

an undeniable heuristic value, does not enable a full grasp of the dynamics of artificial intelligence. Moreover, for similar reasons, the approach via risk does not suffice (even though it is probably necessary), and the question of controlling AI, which is often the angle through which ethical issues are addressed, thus seems as irrelevant as it does for nanotechnologies and cybergenetics.

Furthermore, as Stuart Russell stresses in his 2019 book *Human Compatible: Human Intelligence and the Problem of Control,* the problem is not simply that we risk one day having to observe, as gorillas observe humans, an intelligence that is more highly evolved than our own, nor is it that what we seek to do risks turning into a nightmare like Midas's when he sought to turn everything he touched into gold. The principal problem, according to Russell, is especially to determine how to proceed in such a way that the question of control cannot even be raised. This would mean reconceptualizing AI in order to make it humble, altruistic, and above all guided by human – all too human – objectives (rather than superhuman or transhuman, we might add). At the same time, then, on the one hand we would need to acknowledge human diversity and the fact that human aspirations are not always predictable; on the other hand we would need to embark on a vast movement that would be not technological but cultural, in order to promote options other than the delegation of decision-making power to machines. The call for what Russell calls human "autonomy" – in the sense of a capacity to decide freely without delegation – must thus be set squarely against the call for the autonomization of AI.

We might do well to consider Joan Tronto's suggestion that a risk society relies on an approach that "discusses the 'unintended consequences' of social action from a 'God's eye' vantage point, the normative implications of this approach being neither explicitly formulated

L-actu-du-Ministere/Alicem-la-premiere-solution-d-identite-numerique-regalienne-securisee). This is probably an unprecedented development in the Western world, in that it has been implemented by France (a centralized and centralizing nation) rather than by private enterprises alone, in view of developing public/private partnerships based on harvesting public data (as is already the case for the Health Data Hub and its project of aggregating all public health data in France).

nor explicitly defended" (Tronto 2012, 8). Its mode of operation is nourished by the fear stemming from unpredictability. But as Jean-Pierre Dupuy (2007) has shown, unpredictability is part and parcel of NBIC metaphysics, understood in the sense inspired by Karl Popper (1959) of an "unfalsifiable" theoretical framework that orients the questions raised but also frames and limits them. In other words, if the anxiety that underlies a risk society arises from the feeling of inadequate control, then perhaps the constitutive unpredictability of NBIC has to be acknowledged; the issue has to be detached from the question of control and placed elsewhere. One possible solution to this problem might be to complete, or even to replace, the approach via risk by an approach via care,[8] and to leave behind the problematics of risk and control – a problematics that is in any case virtually impossible in the case of deep learning, as we have seen – by asking first and foremost what matters to us and how we can take care not only of ourselves and others but also of the institutions that allow us to live together; Tronto (2013) calls this *care with*.

3.1.3 *State of the art of AI ethics*

The state of the art of AI ethics is not easy to characterize, for in the early 2020s we are at a key moment in which initiatives are multiplying in an effervescence that is hard to grasp synoptically. Moreover, AI ethics, viewed as a field of its own, is located today at the crossroads between academic research and "governance" by many more or less institutional committees, including at the highest international levels – most notably that of UNESCO, which has proposed a recommendation on AI ethics.[9] Without any claim of exhaustivity, alongside these international political initiatives we can identify at least

[8] As I have noted earlier, the ethics and politics of care, which originated in reflections on the question of gender, could open up the fields of ethics and politics by displacing their boundaries, loosening their binary structures, and taking into account vulnerabilities as well as relationships.

[9] The text can be found at https://unesdoc.unesco.org/ark:/48223/pf0000381137. See also the preliminary report by COMEST: https://unesdoc.unesco.org/ark:/48223/pf0000253952.

four principal approaches in the field of AI ethics: (1) the Hippocratic Oath; (2) applied ethics; (3) fairness by design; and (4) moral machines. These approaches are characterized by the dynamics of their processes: either top-down, starting from principles and moving to practices, or bottom-up, seeking to program ethics into AI, with some occasional attempts at hybridization.

3.1.3.1 A new Hippocratic Oath?

A first approach, which originated in the field of engineering, would consist in a "Hippocratic Oath" for coders. This approach is interesting in that it is presented as a bottom-up deontological project, coming from coders themselves, whether through associations,[10] individuals, or companies. The idea of a code of conduct comparable to that of medical doctors is not new: in 1991, students at the prestigious École Polytechnique Fédérale de Lausanne (EPFL) created the "Archimedean Oath," a code that commits engineers, acting both in their professional capacity and on their own behalf, to follow some fairly broad principles (in particular concerning human rights, defense of the environment, and economic justice), and it insists on the need for engineers to accept responsibility for their actions.[11] This rather general code supplements more local codes of conduct that were already in place.[12]

As Ali Abbas and his colleagues suggest in "A Hippocratic Oath for Technologies" (2019, 72), the credo behind the new Hippocratic Oath could be described as based on three pillars: "(i) proactively understanding the ethical implications of technology for all

[10] In French, see for example the declaration of "Data for Good": https:hippocrate.tech/.

[11] It is mentioned in the ethics charter of EPFL (École Polytechnique Fédérale de Lausanne): http://ethics.iit.edu/codes/EPFL%20Undated.pdf. The charter can be found at https://www.epfl.ch/about/overview/wp-content/uploads/2019/09/Charte-ethique-de-lEFPL.pdf.

[12] See for example the International Electric Export (IEE) code for engineering software: https://ethics.acm.org/wp- content/uploads/2016/07/seeri.french.code_.one_.column.pdf.

stakeholders, (ii) telling the truth about the capabilities, advantages, and disadvantages of a technology, and (iii) acting responsibly in situations you find ethically challenging."

It is undeniable that the establishment of a code of deontology in the AI sector would represent a decided advance. Two questions remain open, however, if we now consider ethics as a specific technical field that presents singular concepts and has a singular history. The first is that since, on the technical level, deontology is only a small part of ethics, ethics cannot be limited to deontology, which represents simply (even though this is already significant) a set of rules that are intentionally very general and quite limited. The second is that the very choice of the Hippocratic Oath is problematic in historical terms, for it was clear at the end of the Second World War that the Hippocratic Oath left aside the question of experimentation (and thus the question of consent), experimentation having been practiced on human beings under atrocious conditions by Nazi doctors who had nevertheless sworn that oath. Nazi medicine, under the direction of those same doctors, pursued experimental efforts in the name of a racist, hierarchical, and reductionist approach to human beings, and in the name of a (perverse) utilitarian vision on the grounds of the experiments' usefulness for improving military innovations but also for advancing research and improving the well-being of all (or at least of Aryans).[13]

The issues of experimentation and consent are two eminently problematic points in considerations about the ethics of AI, so much so that, even though the proposal of a deontological code is probably a necessity, conceptualizing such a code on the model of the Hippocratic Oath seems particularly delicate, and the argument made by Abbas and his colleagues (2009), which foregrounds that oath because it is well known, fails to take this important consideration into

[13] A great deal of research has been carried out on this subject, far too much to cite it all here. In the French context, one can consult the collective work edited by Christian Bonah (2006), *Nazisme, science et médecine*, and an article by Philippe Amiel (2009), "Expérimentations médicales: Les médecins nazis devant leurs juges."

account. One might suggest that the model of the Nuremberg laws, which bears on issues that appear essential today in the AI field, especially the issues of experimentation and consent, should be integrated into AI ethics; this model seems more pertinent than that of the Hippocratic Oath, and it is no less well known. However, the Nuremberg code is marked by its unique context and would probably not be fully suitable either, even if it has the undeniable merit of underlining the importance of taking into account the notions of responsibility and experimentation (and thus of consent). Finally, the modalities for obtaining consent are still very broadly open to discussion, not only on the technical level but also on the ethical and legal levels.[14]

3.1.3.2 An applied ethics for AI?

Another approach seeks to identify major principles for AI ethics, whether these are directly tied to the three principal ethical theories (deontological ethics, consequentialist ethics, virtue ethics) or whether they are general principles that supply the *sine qua non* criteria of an AI ethics. It would then suffice, as it were, to ascertain that these principles are well respected and applied to obtain an AI ethics.

Depending on whether the central question involves duties or principles, a so-called deontological approach or a so-called consequentialist approach is likely to be adopted – even though the break between the advocates of the two approaches may be less pronounced today, and a combined evaluation by way of principles and consequences is not out of the question. Still, it appears today that the approach by way of principles predominates, probably in part because, as I have already noted, a thoroughgoing grasp of the consequences of AI seems hard to achieve.

[14] Evidence of this can be found, for example, in the recent European directives on the General Data Protection Regulation (GDPR), and also in the installation of the Health Data Hub, which is based on the debatable notion of implicit consent. See for example Bernelin 2019.

Two principal entry points to AI ethics can be envisioned today. First, the idea of founding the ethics of AI on human rights is quite widely defended by legal experts (especially in the Anglo-Saxon context),[15] with an emphasis on the fact that some aspects of human rights are openly threatened at present by the way AI is developing – most notably the right to non-discrimination, as we shall see. A second entry point has been defended on a global scale by a group of experts under the auspices of the European Commission in a report published in April 2019.[16] This group proposes to identify the guiding principles of a trustworthy AI, namely, accountability (in the sense that one can and must account for, thus justify, one's actions), responsibility, and transparency. Accountability here is understood to be linked to interaction (with the systems involved); responsibility, to autonomy (of the systems); and transparency, to adaptability (of the systems), The notion of responsibility is understood both as responsibility *for* the system and as responsibility manifested *by* the system, and it is inseparable from the notion of intention or purpose that one can and must program into the system.[17] Interestingly, these principles are sometimes complemented by others, which seem (although the reference is rarely mentioned) to have come directly from biomedical ethics – doing no harm, doing good, maintaining autonomy, and practicing justice.

This approach via principles is inseparable from a conception of ethics according to which it derives from the major leading deontological principles – a conception closely linked with certain attempts at international governance. This is a crucial point, for it underlines

[15] This is also the point of view defended by many organizations such as the Mozilla Foundation and the Rockefeller Foundations. See for example https://www.elementai.com/news/2019/supporting-rights-respecting-ai?utm_source=twitter&utm_medium=social&utm_campaign=Brand_GR&utm_content=human_rights_bloh_11/27/2019.

[16] https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai.

[17] See Dignum 2019, *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*.

the current interdependence between ongoing research in AI ethics and economic and/or political globalization. The research at issue is closely connected either with major international institutions such as the GAFAM companies (Google [Alphabet], Apple, Facebook [Meta], Amazon, and Microsoft), as is largely the case in the United States and in major international institutions such as the European Commission or UNESCO. These two types of interactions are of course quite dissimilar: the latter can be explained by the growing awareness that AI can be analyzed only as a global undertaking. This may seem paradoxical, for the global approach must necessarily incorporate, it seems to me, the question of cultural diversity.

3.1.3.3 A "fair by design" AI

Another approach, particularly well developed in the United States, is that of an AI that would be "fair by design." The term Fairness by Design was of course intended to echo the other two conceptions "by design" that we encountered with the nanotechnologies: Safety by Design and Privacy by Design.

The implicit reference underlying Fairness by Design is to the debate over Big Data and the discriminations that Big Data are capable of engendering, in the matters of race and gender in particular. The principal systems used, especially in the legal framework, such as the well-known American software COMPAS (Correctional Offender Management Profiling for Alternative Sanctions),[18] were subtended by significant constitutive biases that ended up increasing inequalities, not only at the individual level but also at the level of groups – Black persons in particular were subject to more systematic discrimination. The problem here arises at the constitutive level: the biases of the algorithms reflect not only those of their creators but also those of the dominant groups in the culture.

The general idea, said to be inspired by Isaac Azimov, is to propose not simply positive principles but also guardrails in order to define

[18] See for example Larson et al. 2016, "How We Analyzed the COMPAS Recidivism Algorithm."

unacceptable behaviors.[19] This approach opens up the question of programming beyond its purely algorithmic aspects so as to envision representations of the world that is in play – which presupposes the possibility of integrating researchers in the social and human sciences into program development teams in order to ward off or compensate for biases.[20]

One of the difficulties of this approach is probably that it reduces the ethics of AI to considerations of justice and equity, even though that limitation has been contested from the 1980s on by the ethics of care as restricting ethical questions to approaches subtended by limited and even patriarchal conceptions of the moral world, insofar as, on the one hand, ethics cannot be reduced to questions of justice and, on the other hand, that, socially, boys and men in particular are trained to calculate moral situations in terms of an arithmetic justice (Gilligan 1982). Another pitfall is the degree to which it is centered on U.S. culture, which is itself based on a singular history and cannot necessarily be transferred unaltered into other contexts.

3.2 Coding morality?

A large number of AI teams have set themselves the task of developing "moral" machines. The logic behind the development of these AIs tends to be either top-down or bottom-up, although there are also approaches deemed hybrid.[21] I propose to focus first on a quite recent example of the bottom-up approach, in order to identify the problems it raises; we shall later consider other examples.

The approach in question is situated at the point where artificial intelligence, psychology, and cognitive linguistics converge. Anchoring

[19] See Thomas et al. 2019, "Preventing Undesirable Behavior of Intelligent Machines."

[20] See for example Abbasi et al. 2018, "Make 'Fairness by Design' Part of Machine Learning."

[21] This typology is the one most often adopted by following the model described in Allen et al. 2005, "Artificial Morality: Top-down, Bottom-up, and Hybrid Approaches." For a more recent overview, see Tolmeijer et al. 2020, "Implementations in Machine Ethics: A Survey."

their work methodologically in a descriptive (rather than a normative) view of ethics, researchers have been seeking to determine what groups of individuals are apt to find ethically acceptable, either through experiments carried out in cognitive psychology, as is the case with the "Moral Machine,"[22] or through an automated treatment of language on the basis of a corpus of diverse texts (including literary and religious works from the Renaissance to the present, and also newspaper articles and constitutional texts), as is the case with the "Moral Choice Machine" developed by German researchers.[23]

While the explicit goal of this approach is to create "ethical" or "moral" machines such as driverless cars, said to be autonomous, or military machines such as combat drones, the approach also encompasses algorithmic machines intended for a wide variety of uses. The properly descriptive aspect of this undertaking situates it as engaged in the search for what is "socially acceptable"; this distinguishes it from normative ethics, which seeks to evaluate not what individuals or groups of individuals are prepared to accept (or may be convinced to accept) but what they should or should not accept from a moral point of view.

I shall come back to the Moral Machine (an example of the top-down approach, it is designed to resolve moral dilemmas through the use of pre-determined built-in rules) in discussing the so-called autonomous machines; here, I seek to bring to light a certain number of problems posed by the "Moral Choice Machine" – it is important not to confuse the "Moral Machine" with the "Moral Choice Machine." The latter is based on a presupposition whose formulation may appear somewhat astonishing: since AI is imbued with "negative" biases, it can also be imbued with positive, that is, moral biases. To train an AI machine to make moral judgments, it would thus suffice, as it were, to inculcate in the machine the cartography of

[22] See Awad et al. 2018, "The Moral Machine Experiment."
[23] See Schramowski et al. 2020, "The Moral Choice Machine."

morality in the literal sense,[24] based on the lexical proximity of concepts and of the uses of those concepts. This cartography is based on a classic technique for natural language processing that explores lexical networks by attributing a vectorial value to sentences within a predefined vectorized space. This technique has been used to create a set of do's and don'ts that have served as the basis for machine learning, rather like a moral "compass."

This process is based in part on other attempts to automate morality that have been explored in particular in the medical field. In that field, the use of AI to help in decision-making is highly developed, especially via a database of information about diseases – including those that have not yet been well understood – and also, for example, in the area of medical imaging. In the register of decision-making assistance, the software program MedEthEx, developed by Michael Anderson, Susan Leigh Anderson, and Chris Armen, is designed to train students with the help of an interactive program largely based on the major principles of classic ethics (beneficence, non-maleficence, respect for autonomy, and justice) and supported by an algorithm of logical inference.[25] In this sense it is a top-down approach.[26] Nevertheless, unlike MedEthEx, which proceeds above all by a top-down logic (from principles to practice), the situation is somewhat different with the "Moral Choice Machine," since it is the machine that is categorizing and learning.

Thus different ways of conceptualizing moral AIs can be identified, corresponding to different forms of AI. In some cases, for example, MedEthEx, the system is asked only (so to speak) to apply rules

[24] See a summary of Schramowski et al. 2020, "The Moral Choice Machine."

[25] See https://webcampus.drexelmed.edu/medethex/index.html Unfortunately, I have not managed to get the access code to try it out. An introductory video is available on YouTube. For explanations, see Anderson et al. 2006, "MedEthEx: A Prototype Medical Ethics Advisor."

[26] The model also incorporates a so-called *prima facie* dimension, which might make it a hybrid model, but I have to admit that the way the top-down approach is moderated is not necessarily obvious. It seems to me that it might be preferable to stay with two categories (bottom-up and top-down) and to integrate the hybrid approaches into one or the other, depending on its degree of proximity.

automatically. MedEthEx is asked, as it were, to place situations in certain pre-established categories and to produce a predetermined response according to its computations. In other cases, the system can be asked to learn on the basis of data, that is, to produce simulations on the basis of categorized situations and a data base describing what human groups would deem it proper to do. With MedEthEx, however, the process is not normative but descriptive; this approach is used most notably by the inventors of the "Moral Machine."

While the inventors of the "Moral Choice Machine" are aware of a certain number of difficulties and bring certain nuances to the use of their machine or to what it makes it possible to do (for example, by emphasizing that their method shows that moral associations vary according to time and place), the fact remains that the process itself raises questions in various respects. The possibilities and conditions of its application are problematic, to be sure, but I seek to show that the issue here must not be framed solely in terms of practical ethics but must also, and even especially, be framed in meta-ethical terms.[27]

The possibility of generating a moral code starting from a database (whether the process is bottom-up, top-down, or hybrid) is, contrary to what one might think at first glance, a temptation that originated in classic philosophy. The British Enlightenment philosopher George Berkeley, whom we encountered in the introduction, hesitated at length about whether to argue that a moral system could be "demonstrated," that is, "proved," as in mathematics, or that such a demonstration was impossible. The historical reference is interesting not only because it shows that the "Moral Choice Machine" fits into a project that goes beyond the twenty-first century. Berkeley's approach to morality is also interesting because of his hesitations, which bring to light the difficulties inherent in such a quest. He examined the reasons why a moral system might be demonstrable (or generative), on the one hand, and the reasons why it might not be,

[27] For this reason, I am deliberately choosing not to discuss in detail each of the attempts to automate morality (that discussion would require a book of its own); the problem seems to me to lie not so much in the details but rather in the posture.

on the other. The steps in his argument, quite as much as his conclusions, are thus very enlightening for our own investigation.

On the one hand, Berkeley argued in "Philosophical Commentaries" that morality could be demonstrable. The rationality of the physical world went hand in hand, as he saw it, with the hypothesis of a mathematical morality whose laws would be formally comparable to those of the natural sciences, so that "Morality may be Demonstrated as mixt [applied] Mathematics" ([1707-] 1948, 92). Berkeley was probably influenced to some extent by John Locke's contention that our knowledge could be considered solidly grounded only if our ideas were in conformity with the realities of things.[28] But Berkeley took this proposition further, for mathematics constitutes a practice, applied science.[29] Thus he proposed a discipline dealing with rules of conduct that would not be pure theorization – unlike Locke's "pure" scientific morality. As with mathematics, morality is connected with signs and words. That is why Berkeley asserted that "[t]o demonstrate Morality, it seems one need only make a Dictionary of Words & see which included which, at least. This is the greatest part & bulk of the Work."[30] On the basis of this "moral dictionary," all the other propositions could then be established by induction. The real obstacle to this demonstrative moral system would arise, according to him,

[28] John Locke (1690) 2001, *An Essay Concerning Human Understanding* (Kitchener, Ontario: Batoche Books), Book IV, chapter IV, "Of the Reality of Knowledge," 465-75. Locke distinguished between simple and complex ideas. According to him, we have knowledge of "simple ideas" because they are produced in us regularly and naturally thanks to operations that take place outside of us, whereas "complex ideas" are produced and linked to each other by the mind, independently of nature. In the first case, the ideas are in conformity with things, whereas in the second, it is things that are in conformity with ideas; but both are objects of real knowledge. For Locke, morality involved complex ideas, and we would recognize to what extent morality is similar to mathematics if we could agree on sufficiently rigorous moral definitions.

[29] See especially George Berkeley (1710) 1998, *A Treatise Concerning the Principles of Human Knowledge,* 148, §121, and 152, §131.

[30] Berkeley (1707-) 1948, 84, §690.

from the power of prejudices in the matter, which necessitated a reformulation of the moral demonstrations.[31]

Berkeley's guiding thread at this point was to seek a formulation convincing enough to surpass human prejudices, and then to focus on fundamental duties. He thus introduced a fundamental distinction between positive and negative moral precepts. It was not a question of telling human beings what must be done, because action often depends on external accidents; rather, it was a question of preventing certain transgressions by establishing guardrails. One might object to Berkeley, however, that abstaining from doing something can also depend on sometimes unforeseeable external accidents, so that the difference between the two types of precepts is tenuous and involves first and foremost the question of causal responsibility for our actions.

But what might seem surprising is that, alongside this analysis of the possible demonstrability of morality, Berkeley also examined the hypothesis according to which morality would be undemonstrable, and in the end he seemed to lean toward this second hypothesis, despite the effort he had put into supporting the first. He asserted that "[w]e have no Ideas of vertues & vices, no Ideas of Moral Actions wherefore it may be Question'd whether we are capable of arriving at Demonstrations about them, the morality consisting in the Volition chiefly."[32] Thus he proposed to abandon the mathematical method, rigorously analytical, in favor of a less static conception. The advantage of the latter, according to him, is that it leaves ample room for the essentially active role of human nature, characterized by free will. As John Wild, a commentator on Berkeley, has emphasized, to

[31] Berkeley (1707-) 1948, 22, §163: "The short jejune way in Mathematics will not do in Metaphysiques & Ethiques, for yt about Mathematical propositions men have no prejudices, no anticipated opinions to be encounter'd, they not having yet thought on such matters. tis not so in the other 2 mention'd sciences, a man must not onely demonstrate the truth, he must also vindicate it against scruples & establish'd opinions wch contradict it. In short the dry strigose rigid way will not suffice. he must be more ample & copious, else his demonstration tho never so exact will not go down wth most."

[32] Berkeley [1707-) 1948, 82, §693.

crystallize morality in a static concept would amount to leaving aside the essential movement of human volition, in the dynamic moment when thought grasps itself.[33]

Berkeley's rejection of the demonstrability of a moral system consists in restoring demonstration to what he believed was its rightful place: that of a mathematical exercise such that the demonstration bears above all on the *form* of the utterances. Alongside this denunciation of the formal dimension of demonstration, Berkeley added a more psychological reason: it is hard to explain to ordinary people why we act in one way rather than another on the basis of some sort mathematical demonstration, which would likely be more complex than what Berkeley saw as the paths taken by the author of nature. Now, we quite often observe the presence in ourselves of innate dispositions and passions that we seek to satisfy. Our principles must thus be taught us by our experience, and we can state them in general terms only once they have reached their fullest development.

This idea of a moral sense that is innate in human beings, developed for the most part during the classical era by the sentimentalist philosophers of the moral sense tradition, from Shaftesbury to Smith or Rousseau by way of Hume with his idea of a "science of human nature," was taken up again in the twentieth century by John Rawls, with an astonishing echo effect: some of Rawls's arguments are similar to the dialectics Berkeley used in comparing generative and non-generative approaches to moral systems.[34]

Rawls tried to think through what he called not the moral sense but the "sense of justice," on the basis of the cognitive sciences and more particularly the generative grammar proposed by Noam Chomsky.[35] He put forward the hypothesis of a generative moral code

[33] See Wild 1962 (1936), *George Berkeley, A Study of His Life and Philosophy*, 340: "Virtues and vices involve action or passage that cannot be crystallized into a static concept or set of concepts from which we proceed to deduce analytic consequences. Such a procedure would miss the very essence of the matter – the creative will."

[34] See Rawls 1971, *A Theory of Justice*. This comparison has been very well analyzed by John Mikhail.

[35] See Chomsky 1965, *Aspects of the Theory of Syntax*.

that would be systematizable but not mechanizable. Rawls asserted that, just as Chomsky was astonished by the poverty of the linguistic stimulus (the fact that young children can understand and produce utterances that they have never heard before), one of the mysteries of moral development is that, from childhood on, we can produce moral evaluations that we have never seen formulated, nor have we previously formulated them ourselves. This is what could be considered – even if Rawls did not do so – the poverty of the moral stimulus. Just as Chomsky spoke of linguistic competence to designate our capacity to learn language, as opposed to performance, which is the effective use of language in a given concrete situation, Rawls proposed to study our moral competence, which cannot be confused with deliberately spelled-out moral rules. This is why Rawls proposed to model his approach on "Universal Grammar," which aims to analyze the universal biological properties of human nature. He saw this theory as "descriptively adequate": it allows us to account for linguistic phenomena by formulating universal principles, but also by explaining the generative mechanisms that allow us to formulate particular cases.

Rawls combined this approach based on competence with an approach inspired by the psychological theories of moral development that were predominant in his day – first that of Jean Piaget, then than of Lawrence Kohlberg.[36] This articulation is based on the idea that the theory of justice is possible only if it is not simply feasible but also desirable. Then we can be motivated by cooperation.

The sense of justice is inseparable, for Rawls, from what he calls "natural attitudes" such as trust or friendship. These natural attitudes are not physiological reactions; we can give reasons for them, typically by mobilizing moral notions. This is why the sense of justice is a blend of reason and emotion; it is neither purely rational nor purely emotional. For this reason, the sense of justice is at once motivating

[36] See for example Piaget (1932) 1965, *The Moral Judgment of the Child*. See also my own detailed study of Rawls's work: Nurock 2008, *Rawls, pour une démocratie juste?*

and justifying: it allows us both to distinguish what is just from what is unjust and to be motivated to act justly. It lies thus at the conjunction of the rational and the reasonable, while inclining more to the side of the reasonable. For in order to be just, according to Rawls, it is not a question of being capable of applying justice, as a machine might run a program; one must be able to understand justice so as to be motivated to act justly. Now it is precisely this motivation that is at the heart of decision-making and action.[37]

From this perspective, we can better understand why the idea of an artificial moral code, which seems to proceed from a generative "moral dictionary," cannot truly be called moral. There are two reasons for this. First, such codes are based on the analysis of lexical corpuses and not of arguments, which are the basis for our moral judgments – and let us note in passing that the same problem faces the "LegalTechs" who claim to use AI to automate legal decisions, especially (but not only) in administrative law.[38] Second, if this code is not based on freedom (to go back to Berkeley's argument) or on a blending of the rational and the reasonable, it cannot be confused with our moral intuitions: to make the "right" decision, one must be able to be mistaken, but also, and especially, one must rely on a dynamic that is not only logical and rational but also *reasonable*. If the moral code machine were endowed with empathy, then the question would be formulated differently, no doubt, but the very idea of a "Moral Choice Machine" shows clearly that the problem is not so much knowing how to apply the code but determining on which conception of morality it is based: is it a purely rational one, which leaves aside an essential component of the sense of justice? The fact that humans codify their moral behaviors does not mean that they are capable of freely following a moral code that they may have helped

[37] I have developed some of these points in more detail elsewhere: see Nurock 2019, "Généalogie de la morale automatisée."

[38] For a discussion of the problems raised by the introduction of AI into law, see Lacour and Piana 2019, "Faites entrer les algorithmes! Regards critiques sur la 'justice prédictive.'"

forge, nor does it mean that it is possible (or desirable) to code morality in the sense of programming it artificially.

Moreover, such an approach is equally debatable on the metaethical level, to the extent that it is implicitly subtended by a jurisprudential and deontological conception of ethics, one that considers duties and taboos. At least two arguments can be summoned to challenge that implicit conception. For one thing, the deontological approach is far from the only possible approach in ethics; it would be necessary at a minimum to make that presupposition explicit and to justify it. Moreover, a jurisprudential approach amounts to foreseeing future decisions on the basis of past decisions. Yet even if the authors of such a moral code machine claim to have taken into account the historical dimension of ethics and its possible evolutions, it is not at all obvious that future ethical decisions can be foreseen on the basis of those made in the past. Here it is a matter of conceptualizing ethics schematically, via patterns, as we shall soon see in a different context. This first point is extremely problematic, for it comes down to assuming a certain vision of ethics and doing so implicitly, without the slightest justification.

Furthermore, if the goal is to automate morality, then the consequentialist approach might have seemed to be easier to program – and indeed, there have been attempts in this direction.[39] Thus apart from the application of a tool for handling natural language, which was available, it is hard to see why and how it is possible to justify its application to the specific field of morality. It might seem more appropriate to use artificial intelligence for tasks requiring extensive calculations, as is the case for example with "act utilitarianism." This form of utilitarianism proposes to evaluate an action morally by enumerating all of its possible consequences. However, that theory is often challenged owing not to its degree of moral exigency but to the extent of the calculations required, for it is not always possible to predict all possible consequences. This type of difficulty could be

[39] See for example the "utilibot" in Cloos 2005, "The Utilibot Project: An Autonomous Mobile Robot Based on Utilitarianism."

partially resolved by artificial intelligence, at least on the theoretical level, for one of AI's strengths is its capacity to calculate. But one of the problems of act utilitarianism is precisely its extremely casuistic dimension: it proceeds step by step, and it can lead to false results if one of its parameters – not being dependent on the subject carrying out the action – changes. Thus this type of approach is also rooted in a logic of performance and absolute control. But it seems impossible to control all the parameters unless one is functioning in a completely predictable world. This is the same type of problem that comes up in the quest for driverless cars. Even if it were possible to conceive of a driverless car and if it were morally desirable to do so, one of the problems is that it would be fully functional only in a separate microcosm where no unpredictable factor could intervene, and only in the presence of other vehicles not controlled by humans.

To conclude this chapter, then, I suggest that, just as we saw in the analysis of nanotechnologies and cybergenetics, the definitional complexity of AI is such that, even though it can be viewed as a form of anticipatory design, it blurs boundaries, and it cannot be circumscribed within the categories of control or social acceptability (or, to be more precise, social *acceptance* as I have defined it in the first chapter), inasmuch as it claims to be able to embody and to automate ethics in terms that are at the very least problematic. In the following chapter, we shall go into greater detail on these points and explore, among other things, in what respects the ethics and politics of care can offer fruitful paths for conceptualizing the ethics of artificial intelligence.

CHAPTER 4

Two Paradigmatic Case Studies of "Ethical" AI – Autonomous Machines and "Relational" AI

Two uses of artificial intelligence strike me as particularly significant in several respects. First, because these uses invite us to focus on the issues of autonomy and relationships, two issues that are central for ethics and that appear especially problematic where AI is concerned. Next, because they are excellent examples of the way AI can blur classical boundaries and traditional dichotomies. Last but not least, because these two examples make it possible to show, between the lines, how care can add fruitfully to ethical reflection on AI. It is worth noting that in each case, the use of AI in so-called autonomous machines and its use in so-called relational machines, the designation applied is inaccurate.

4.1 Machines said to be "autonomous" and the automatization of ethics

4.1.1 *Why autonomous machines? Artificial intelligence and the smart city*

The notion of autonomy is central today in discussions of artificial intelligence. This centrality is doubtless in part a legacy of the prevailing futuristic thinking in the late 1970s and the 1980s, when robots were featured as automatons capable of emancipation, and when emancipation seemed to be an ideal in itself. The idea of autonomy in AI has developed since then in many fields of application, most importantly perhaps within "smart cities," but also on battlefields, in the form of so-called autonomous vehicles and drones, whether civil or military. In both cases, enthusiasts stress the fact that there is no pilot, indeed not necessarily any human being on board: the term used is "unmanned." However, unlike drones, which do not transport humans, driverless vehicles can do just that, whether their passengers are adults capable of driving themselves, adults who do not know how to drive or are physically unable to do so, or children traveling alone.

While earth-bound vehicles and drones are dissimilar in many respects, and while there are decided differences even within each category, it is nevertheless important to stress at the outset that in all cases ethical issues are at the heart of the problems raised by these so-called autonomous AI machines. Just as I earlier questioned the meaning of the term "intelligence" in the expression "artificial intelligence" (some critics speak insistently of "artificial stupidity"[1]) and noted the use of "intelligence" in the Anglo-Saxon context to refer to espionage, one can question here what is meant by "smart" in the terms "smart city" or "smart bomb."[2]

Whereas John Rawls stressed the importance of keeping the rational and the reasonable together, allying logical intelligence with moral intelligence without conflating the two, we may wonder under what conditions "smart cities," which one imagines full of drones and driverless vehicles, could be not only rational, if they actually are – no doubt the type of rationality in question would have to be specified here – but also reasonable. Here, I am seeking precisely to show

[1] Deploring the "unkept promises" of AI and even alleging its "stupidity" (as opposed to intelligence), critics stress with increasing frequency the inability of the technologies based on AI to produce irreducible novelty, thus making self-fulfilling prophecies all the more practical: the very idea of possible alternatives is nipped in the bud, as it were, and made useless. This is why it will be important, in future analyses, to keep in mind this operative dimension (operative in several respects) of anticipatory design, for it renders the thought of alternatives pointless or even impossible.

[2] As the Australian philosopher Peter Singer has written, a "smart" bomb is a bomb that explodes where it has been directed to explode: "As one navy admiral put it, 'Smart bombs' are really only 'pretty obedient bombs.' A human finds and designates the target and the bomb just goes where it is told" (2009, *Wired for War: The Robotics Revolution and Conflict in the Twenty-First Century*, 57).

that this field raises problems because it leads us to question not only the notion of autonomy itself but also the reasons why, with regard to these AI machines, we tend to insist on their autonomy.

How many TV viewers from the 1980s on have dreamed of finding themselves passengers in KITT, the intelligent car in the series "Knight Rider"? The title of the TV show refers to the name assigned to the hero, Michael Knight, a name transmitted, as it were, by the Knight Company, benefactor of the "Foundation for Law and Government." The autonomous vehicles (also called driverless) being developed today may have certain advantages, but they are probably less romantic. They are more oriented toward efficiency and performance than toward the complementarity of humans and machines, or even the insolent companionship depicted between KITT and Michael. More importantly, driverless cars would help us gain money and time – and time is money, according to the well-known adage.

Five principal benefits are attributed to driverless vehicles – although it is far from clear that these claims are justifiable. (1) Driverless vehicles would save money on gas, by adjusting both itineraries and driving styles. (2) As a result of these adjustments, traffic jams and therefore total time on the road would be reduced. (3) Driverless cars would reduce the number and the seriousness of accidents. (4) They would save time for the persons who could do other things instead of driving. (5) They would allow persons unable to drive for any reason to travel without depending on someone else for transport.

As for drones, their advantages, which can be summarized under four main headings, are primarily tied to miniaturization. (1) They provide new viewing angles and sight lines. (2) They increase the possibilities of pursuit. (3) They are relatively inexpensive compared to other devices that perform the same tasks. (4) They can be enhanced by an arsenal of equipment: heat sensors, listening devices (especially for mobile communications), microphones, and so on. Their adaptability, their small size, and their extreme mobility make them versatile instruments adaptable to many kinds of missions – capturing a sensational view for an adventure film, stealing photos on behalf of paparazzi, providing valuable information to farmers, delivering a defibrillator quickly for a heart attack victim. Their uses range from the trivial (or uses deemed trivial because they provide a form of entertainment) to the life-saving; in this sense they represent undeniable progress; perhaps more than for any of the other technological developments we have examined, the question regarding drones is not whether they must be developed – for they undeniably provide essential services, especially in the medical field, where they allow rapid and efficient delivery of organs or medications, for example – but *how* they should be developed.

Their diversity is such that it is hard to talk about them in a general way. As Michael Boyle has indicated in *The Drone Age: How Drone Technology Will Change War and Peace* 2020) drones are characterized by their capacity to change depending on the way they combine elements that are found in other machines. This can be a problem because certain uses of drones might impair other functions through association or contamination (11). Here we find, although in a very different form, of course, the enabling and low-cost features we encountered in the discussion of nanotechnologies. These are the features, as Boyle notes, that allow drones to appear as an eminently disruptive technology.

If we follow James Moor's lead in "The Nature, Importance and Difficulty of Machine Ethics" (2006), we can distinguish four different modalities according to which ethics might be implemented in AI. First, via *ethical impact agents*, agents whose actions may have ethical effects: here Moor uses the example of the enslaved child camel jockeys traditionally used in Qatar and Saudi Arabia who have been replaced by robotic jockeys since 2004.[3] Second, via *implicit ethical agents* that do what they are told to do – for example, an

[3] Camel races traditionally used very young children (often under six years of age) purchased in neighboring countries and starved so they would remain at very low weights. At the turn of the century, a Swiss company perfected robotic jockeys that could replace those children. The robots were not ethical in themselves, but their use potentially made it possible to put an end to an eminently condemnable practice.

automatic airplane pilot, or a funds-transferring program that does not steal – but that are nevertheless not fully ethical for all that. Third, via *implicit moral agents* that implement an ethical theory: this is the case with Bentham's hedonistic utilitarianism, which undergirds the Jeremy program proposed by Susan Anderson and her colleagues (2005), whom we have already encountered in the context of the MedEthEx program (2006); this modality is a little more complex. Machines based on implicit ethical agents, according to Moor, could be used in situations of catastrophe, and more broadly in situations of triage – when a decision must be made as to who lives and who dies. This third level is thus of fundamental importance, because it implies real and critical decision-making power. Fourth, via *full ethical agents* that are not only capable of supplying explicit moral judgments but also of providing justifications for these judgments. As Moor notes, the pivot point lies not at the fourth level but rather at the third – which is precisely the level at which there are current efforts to develop autonomous AI machines.

Still, it seems to me that the very idea of being able to implement an ethical theory *without adjustment* to the internal state of an agent lacks an essential component of what constitutes the foundation of such theories: the motivation to act morally. As I see it, Moor's description does not adequately address the problem from the meta-ethical standpoint; it misses part of the problem by remaining at the level of normative ethics. Before going any further, if we consider Moor's four ethical levels in the context of the somewhat trivial descriptions of driverless cars proposed by their manufacturers, we arrive at a rather interesting congruence at Moor's level three.

The various management teams of automobile manufacturers generally agree in recognizing five or six levels of vehicle autonomy, which can be described in terms of levels of control. At level 0, the driver controls all functions, even if an on-board computer can assist (for example by beeping when the driver crosses a line), or a camera can help with parking. Level 1 gives the vehicle more latitude (for example in regulating speed in relation to the vehicles ahead), but the driver still controls all functions. At level 2, the vehicle manages several functions (for example, steering in addition to speed control,) but the driver can take over at any time and indeed sometimes must do so. The driver is the supervisor, but the on-board computer does a good part of the work. Brands such as Tesla are more or less at this level, with what are called "autopilot" systems, though it seems quite possible that such systems are likely to go even further. At level 3, driving is fully automated in the context of predefined situations, and the vehicle can signal its inability to handle a given situation – this feature is offered by certain Volvos and certain forms of Tesla's Autopilot, for example, for the low-speed driving required in traffic jams. Everything else being equal, this level 3 could correspond to level 3 in Moor's description (which can be challenged, as we have seen). At level 4, the vehicle can drive off on its own, that is, with no one on board, in predefined situations, for example, to pick someone up, or to park; it can also convey a passenger unable to take the wheel, someone who does not know how to drive or is unable to drive, such as a handicapped person or a child. At level 5, the vehicle can do this in any situation.

The consensus among automobile manufacturers in 2018,[4] supported by an official document produced by the European Parliament,[5] led one to believe that level 3 autonomy had already been achieved, and that level 4 would be reached in four or five years (dating from 2018), that is, by 2023, and level 5, full autonomy in ten years, or 2028. In late 2023, San Francisco authorized two driverless car companies to operate at any time of day or night – but it revoked the authorization in one case two months later, following major malfunctions and accidents.[6] So we are not in a science-fiction situation; this is happening in real life. An unknown number of driverless vehicles are being tested currently, more or less *ex vitro* – in real-life contexts

[4] See Gerdes 2018, "Not So Fast. Fully Autonomous Vehicles Are More Than a Decade Away, Experts Say."

[5] See European Parliament 2016, "Automated Vehicles in the EU."

[6] For a map of accidents involving driverless vehicles in San Francisco, see https://www.sfchronicle.com/projects/2023/self-driving-car-crashes/.

and not just under laboratory conditions – throughout the world. It is hard to say, then, whether the predictions are credible, all the more so because economic and legal questions raise problems that are not purely technical.

Nevertheless, if the predictions do not necessarily tell us what exists or will exist, if they are not descriptive, they are in a way prescriptive, for they show the urgency of *reflecting* today on what autonomous vehicles could or should be, or what they could not or should not be, indeed of asking whether they should or should not be put on the market for sale or not, from an ethical and even a political standpoint – for it is more and more apparent today that such machines risk reinforcing economic and political hierarchies and restructuring our living spaces (at least our urban spaces, in a first phase),[7] and even our relationships, since they might be used for example to take children to school or to extracurricular activities, thus replacing parents, grandparents, and babysitters.[8]

The discourse on these future machines is such that they have an impact on the present, which seems to have to adapt so as to project that future by anticipation. We find ourselves in a situation we have already encountered in the chapters on nanotechnologies and cybergenetics: performative anticipation and self-fulfilling prophecy. Drones, for their part, are already widely used, in civil as well as military applications. In both cases, we are not in the realm of science fiction, but rather in a more or less palpable reality, one that has already led to a certain number of deaths, but also, it must be emphasized, a certain number of lives saved, for example in medical

[7] See for example Stacy and Meixell 2018, "Self-Driving Cars Could Harm Low-Income People If We Don't Prepare Their Rise," and a report for the Union of Concerned Scientists, Ezike et al. 2019, "Where Are Self-Driving Cars Taking Us? Pivotal Choices That Will Shape DC's Transportation Future." On this point, see also Sparrow and Howard 2017, "When Human Beings Are Like Drunk Robots: Driverless Vehicles, Ethics, and the Future of Transport."

[8] This possibility is evoked in Wadhwa and Salkever 2017, *The Driver in the Driverless Car*; see also (and especially) Patrice Tremoulet et al. 2019, *Human Factors*.

contexts. But we are also dealing with a normative technology in two senses: in the sense in which its anticipation creates a norm of action, and in the sense in which it becomes necessary to reflect on its possibilities. Still, in the second sense, it might well be necessary to reflect on the first sense (we shall come back to this at the end of this book), that is, to ask ourselves what it means that the future is presented as at once necessary and desirable, and in what way these two categories are linked. Is the anticipated future necessary because it is desirable, or is it "desirable" because it is presented as necessary?

In recent years we have been witnessing an important evolution in the analyses bearing on autonomous AI machines. Whereas critiques in the early 2010s, placed particular stress on the cold, automatic and inhuman side of such machines and on their exaltation of an omnipotence that has often been denounced as postcolonial, we have observed more recently a way of returning to the question of the human behind the machine that is far from anodyne. This phenomenon can be explained in part, as I see it – beyond the undeniable military propaganda that has tended to rehumanize these AI machines – by the fear aroused by the more or less imminent advent of such machines with level 4 or even level 5 autonomy, where humans would be completely absent, but also by the predominance of a discourse on the autonomy of machines that is itself problematic, as we shall see.

In other words, the way we look at these AI machines also depends on the viewpoint from which we see them: from the vertigo they cause at the beginning of the autonomization process to the vertigo situated at the end of the process or presented as such. This evolution is in itself significant and disturbing.

These autonomous AI machines pose a number of problems. I propose to analyze four of these: first, the problem raised by the meta-norm "ought implies can"; second, the problem posed by new relationships that are at once interpersonal and spatial; third, the question of the schematization of modes of life via the importance of *patterns of life*, as opposed to *forms of life*; fourth, the problem of the very concept of a moral machine.

4.1.2 *"Ought" implies "can"*

The first real-world experiments with driverless vehicles seem to have brought to light numerous technical programming defects. In fact, a series of accidents in the late 2010s called into question all predictions about the date by which fully autonomous vehicles will be safe for use on public roads. Tesla asserts on its website that its cars are capable of functioning autonomously but that human supervision is necessary at all times.[9] Some of the essential difficulties with these vehicles come up at the first level of their interaction with their surroundings: they may not have enough sensors, or their ability to categorize obstacles may be inadequate. For example, in 2018 a prototype Uber self-driving car ran into and killed Elaine Herzberg, a pedestrian walking her bike across a two-lane road in Arizona. The car's software had interpreted the data from the car's sensors as indicating the presence of an "object" rather than a "human." The "human safety driver" (who was watching the popular televised talent show "The Voice") applied the brakes a second too late.[10]

In short, autonomous vehicles today are at risk of encountering problems at every level: their instrumentation, the way in which the information transmitted by their sensors is interpreted, their programming software and the network into which the vehicle is inserted – leaving aside whatever invisible human agents were in the background.

Similarly, some of the errors committed by drones stem from misinterpretations of the information transmitted by the sensors, for example mistaking a child for a combatant, or miscategorizing individuals according to the way they are armed – one can carry a weapon without necessarily being a combatant – owing to ignorance of the cultural practices on the ground. And this is not all. Daraz Khan,

[9] https://www.tesla.com/autopilot.

[10] According to the *New York Times*, Uber employees had already indicated their concerns before the crash: https://www.nytimes.com/2018/03/23/technology/uber-self-driving-cars-arizona.html. See also an account of the November 2021 accident in which a Tesla on Autopilot crashed and burned, killing two people: https://www.nytimes.com/2021/04/18/business/tesla-fatal-drash-texas.html.

known to be the tallest man in his village, was targeted by a drone and killed after being mistaken for Osama Bin Laden (whereas Khan was actually much shorter than Bin Laden), because he was spotted in a zone that Bin Laden sometimes frequented. The drone operators seem to have simply been looking for a man of more than average height, someone who did not correspond to a classic schema. The basic presupposition being wrong, the virtually automatic result was wrong as well. "Garbage in, garbage out," as Peter Singer, who analyzed the Khan case, has remarked (2009, 399).

In a different register, the problem is more than "merely technical." As Gary Marcus and Ernest Davies have written: "Nobody will buy a home robot that carries their grandfather safely into bed four times out of five."[11] It is not a question of degree but of nature. In some cases, failure, no matter how rare, is a deal-breaker.

Such "technical problems" are all the more concerning in that drones are presented specifically as "precision weapons." The sensation of omnipotence combined with a form of omniscience is capable of neutralizing any second thoughts. As many commentators have noted, the issue is not knowing how to trust AI but rather how to avoid trusting it too much.[12] In a way, the more AI is used, the more its users ought to be trained to call it into question, especially in situations where lives are at stake (for example, in the medical or military realms), even when a decision must be made urgently.

The question of how to provide such training may seem "merely technical" at first glance, but it is also a moral question. We find ourselves in the same type of situation we have already encountered in the context of nanotechnologies, where I proposed to distinguish between first-order and second-order moral questions, the second-order category including for example some apparently technical or scientific questions related to toxicology. The hypothesis that I am

[11] Marcus and Davis 2019, *Rebooting AI: Building Artificial Intelligence We Can Trust*, 18.

[12] See Singer 2009, 399: "Anyone whose computer has ever crashed knows that the human sitting at the keyboard is not always to blame. The system itself can be the problem."

developing here is that NBIC often engage with second-order moral questions as I defined them in the opening chapter, drawing on the work of Elliott Turiel (2002), on account of the way in which ethical and technical problems are intermingled in the NBIC's endeavors, especially because these technosciences cannot be separated from their moral and political significance.

Is it in fact legitimate to promise autonomy and to begin to develop integrated operational software that includes decision-making procedures where life and death are at stake, when the machine is not fully capable of discerning its surroundings? It's rather like asking a passenger who is slightly deaf, blind, and mute to take the wheel. Or like requiring a machine to carry an invalid if the machine is not technically capable of doing so without letting the invalid fall to the floor from time to time.

In relation to the moral meta-norm "ought" implies "can," such a demand would be immoral: one cannot require of someone (or, in our case, something) to make decisions that it is incapable of making: if the software cannot categorize a subject as something other than an object, one cannot expect or require it to do so. To borrow Knut Erik Tranøy's terms, one cannot prescribe the achievement of the impossible, unless one legitimizes the existence of immoral or inhuman worlds.[13] If the meta-norm does not tell us what behaviors would be acceptable from a moral standpoint, it does rule out at the start certain types of behaviors that would be unacceptable from a moral standpoint because they are inhuman or inhumane, in the moral sense. Thus it is perhaps no accident if technologies said to be autonomous are called *unmanned*: "Unmanned Aerial Vehicles" (UAV) for drones and "Unmanned Ground Vehicles" (UGV); these are also called "self-piloted" and "driverless," in a vocabulary used first for aviation and the conquest of space – a conquest that remains a permanent reference for NBIC. It is hardly a big leap to go from

[13] Tranøy 1972, "'Ought' Implies 'Can': A Bridge from Fact to Norm (Part 1)," esp. 123. See also my own analysis of this meta-norm in Nurock 2011, *Sommes-nous naturellement moraux?* (Are We Naturally Moral?)

"without humans" to "inhumane": not only are these technologies amoral, as we shall see, but it is possible to argue that *the claim to be endowing them with a moral system is actually immoral*.

The meta-norm "ought implies can" thus makes it clear that asserting the possibility of driverless cars under current conditions is not only dangerous for technical reasons but also unacceptable from a moral standpoint. The problem arises not from the incapacity but from the demand. This is the case with first-level problems (sensors, categorization), but also with problems at a higher level, and especially with demands for moral performance. The technology involved might indeed need to be improved, but perhaps the moral relevance of our expectations vis-à-vis that technology needs to be questioned as well. Furthermore, we need to challenge the variant of the so-called Gabor's law that commits what is not only an error but a moral flaw by inverting "ought implies can" into "can implies ought."

4.1.3 *The blurring of the inside/outside dichotomy and the modification of the empathic relation*

While the technical limitations of these AI programs "on the road to autonomy" are real and extremely concerning, they are not the only problems raised by the development of AI-assisted technologies. The way they are transforming our vision of the world and human relations is also a disquieting factor.

I should like to begin by stressing that the question of sensors and data interpretation that I have just invoked is in a way the dark underside of a no less concerning obverse, that of ubiquitous surveillance capable of infiltrating every space, "Big Brother" becoming "Little Brother."[14] This surveillance is paired with its inverse reflection: scattered, nonhierarchical surveillance. In this second case it takes the form of *sousveillance*, scrutiny from below (a form we have already encountered in the discussion of cybergenetics) that involves civil as well as military uses of drones, decentralized as well as

[14] Royakkers and Est 2016, *Just Ordinary Robots: Automation from Love to War* 9, 160.

centralized uses, from children's toys and vacuum cleaners to formidable weapons of war.

Autonomous vehicles are in fact not only capable of stocking data concerning both the outside and the inside of the passenger compartment, but they are also connected machines capable of sharing that information and generating new collective data, for example about the state of the traffic in the vicinity. Thus the autonomy of these vehicles can be understood only in the sense of interconnection, all the more crucially so in that the interconnection of the data is what gives them meaning and value.

As for drones, they are essentially surveillance agents, since more and more sophisticated cameras – whose names echo divinities endowed with multiple gazes, like the Gorgon Medusa (Gorgon-IS) or Argos Panoptes (Argos)[15] – are often part of their basic equipment. Moreover, these drones, when they become war machines, or "killer drones," as they are often called, are the vector of the link between surveillance and execution, and they blur the classic binary categories – which are also distinct in intelligence services and in armies – by abolishing the distinction between passive surveillance and active execution, and by eliminating the difference in kind between the pimply nerd on one side and the Rambo on the other. Thus I submit that killer drones deserve particular attention, for they represent an extreme type of AI.

Even though it seems to me that we must absolutely distinguish civilian drones from their military counterparts, and must acknowledge that the uses of the former may be vectors of interesting forms of progress, it nevertheless appears essential not to deny the fact that the latter can inform us a posteriori on the possibly problematic developments of the former. Indeed, war drones are equally capable of being the embodiment of a Foucault-style "boomerang effect,"[16]

[15] While the Gorgon Medusa is capable of striking someone down with a single glance, Argos Panoptes is endowed with 100 eyes. The choice of these two mythological divinities underlines the importance and the lethality of the gaze.

[16] What Foucault calls the boomerang effect is the process through which colonial countries "export" their methods into their colonies, develop them there and

tools of which not only the development but also the uses made in wartime can be extended in peacetime to the citizens of the countries concerned: this is the return of the boomerang. Thus we pass from the outside to the inside, in some cases with the mediation, at least for a time, of enemies from the inside, who are thus in a way the outside of the inside.

In this context, Derek Gregory speaks of an "internumerization" or an "interdigitalization" of war and peace[17]; this strikes me as a particularly apt characterization, for drone technology intertwines things that might appear different in nature, things that even blur the boundaries whose maintenance may be open to question: theirs vs. ours, the space of war vs. the space of peace, the private sphere vs. the public sphere. Furthermore, it also appears probable that, from the standpoint of people involved in espionage, the collection of data is all the less a problem in that they are evolving in a society in which sharing information has become the rule, as we saw in the discussion of cybergenetics. The boomerang effect is thus facilitated, as I see it, by a culture of "sharing" based on the idea that one has nothing to hide, and that "I share, therefore I am," to borrow Sherry Turkle's formulation.[18]

The geographic transmutation allowed by drones in fact abolishes the dichotomies listed above, since with drones surveillance is always possible, and information obtained in this way can be pooled with other data acquired by digital tools, especially mobile phones. Here we come face to face with a complementary aspect of cybergeno-panoptics, since the abolition of the separation between private and public, inside and outside, although produced here by different means, can reinforce the systematic effect of existing networks of surveillance, identification, and profiling.

remodel them, after which these methods come back like boomerangs to the colonizing countries and are applied to their own populations. See Michel Foucault 2003 (1997), *"Society Must Be Defended."*

[17] Derek Gregory 2014, "Drone Geographies," *Radical Philosophy* 183 (3): 7-19.

[18] Turkle 2015, *Reclaiming Conversation*, 47.

As Shoshana Zuboff has shown, the contemporary issue of surveillance can be articulated around three main questions: "Who knows?" "Who decides?" and "Who decides who decides?"[19] Her answer to the first question is clear: capitalists have the knowledge. The second question can be reworded more precisely, in the case of drones, as "Who decides to kill?" Paradoxically, the chain of decisions and actions is simultaneously broadened and tightened by the use of drone technology. The order to kill is directly centralized at the highest level in the case of the most sought-after targets, and the highest-ranking officers are the ones who give the order to kill directly to the machine (and not to humans equipped with cameras). The *mechanization* allowed by drones is thus also a *mediation* that makes it possible to pass directly from the top of the decision-making hierarchy to the ground without going through the several links in the chain. Here we have a dimension specifically tied to the "joystick" aspect of drones, for there is no need to be materially present on the ground in order to kill.

Much has been written about the way drones perform the work of surveillance and of war itself in the form of generalized remote work.[20] Many analysts have commented on the way in which this remote work, at least at first glance (if one forgets the boomerang effect and the digitalization we have already observed), seems to circumscribe a separate space that lies between "us" on the one side and "them" on the other. Moreover, the justification for the use of drones relies explicitly on that distinction, emphasizing that the goal above all is to avoid offering a target to the enemy. As Michael Boyle notes (2020, 19), the promise of being able to carry out strikes without risking the lives of one's own troops makes war "cleaner" (for the assailants with the relevant technological equipment) and thus more

[19] Zuboff 2019, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. Zuboff's analysis focuses on the United States and the West, and it probably misses important aspects of Eastern surveillance (especially Asian), but her conclusions, while doubtless partial, are no less interesting, even though they are not specifically concerned with drones.

[20] See for example Chamayou 2015 (2013), *A Theory of the Drone*.

acceptable: in sum, this capacity makes war easier to wage. As Peter Singer points out, citing a message from a U.S. navy chief petty officer, "[w]hen a robot dies, you don't have to write a letter to its mother" (2009, 21). Making war comes to seem risk-free, and thus costless; Singer calls this the "dark irony" of roboticized war: "By seeming to lower the human costs of war, they may seduce us into more wars" (2009, 322). From this perspective, possessing such technology can alter international relations as much as it modifies relations among individuals.

The way killer drones set up the possibility of executing someone "remotely" has been widely discussed and denounced, for example by Grégoire Chamayou (2015). One of the fundamental problems lies in the way this type of machine replaces symmetry and reciprocity – whether the military forces are evenly balanced or not, soldiers on the battlefield generally put their lives on the line – with a *unidirectional* situation. The emblem of the operations of the MQ9 Reaper drone that killed Qassem Soleimani, commander of the Ai-Qods Force of the Iranian Guardians of the Revolution, in January 2020, is an insignia symbolizing the Grim Reaper, bearing the slogan "That others may die." Those who die are others; neither "ours" nor *alter egos* – or at least this can be one's first assumption.

However, Soleimani's execution marked an important turning point in the use of drones, as we learn from Agnès Callamard, who was at the time the special rapporteur on extrajudiciary, summary, or arbitrary executions for the United Nations Human Rights Council.[21] In fact, this was the first time that a high-ranking member of one country's government was targeted by an official drone of another country in a third country. But far from seeing this event simply as an instance of escalation, denounced as such by Callamard, we can see it also (and especially) as a new stage in the use of drones, which have become presumptively the jewel in the U.S. arsenal. As Callamard points out, this stage opens onto an uncertain future, for,

[21] Callamard 2020. A good summary in French is also available in an interview with Callamard published July 9, 2020, by Radio France Internationale.

a priori, any dignitary at all could be targeted at any moment, all the more easily in that there is a legal vacuum on the subject.

This "new drone era," as Callamard calls it in her 2020 interview – referring to targeted executions, which will probably continue to be directed remotely – is paired with the imminent arrival of so-called autonomous drones, programmed to target and execute not an individual but a group of individuals. This problem has begun to be addressed by a number of analysts, some of whom are close to military milieus, as we shall see.

It may be useful here to note the extent to which the discourse about drones used in war has shifted over time: whereas in the early 2010s their inhumane aspect was the particular focus, for example, in the writings of Peter Singer, Joseph Pugliese,[22] and Derek Gregory, we are now seeing a sort of reversal, since the entire human universe that lies behind drones is being emphasized rather than their artificial dimension. While this reversal of perspective may well result in part from the pro-drone military or industrial propaganda that has been produced in response to a sense that drone operators risk being viewed as nerds losing themselves in video games, it also stems in large part from a growing awareness of the next stage: that of autonomous drones to which the decision to kill would be delegated.

As Singer asserts (2009, 396):

> Each new technology, from the bow and arrow to the bomber plane, has moved soldiers farther and farther from their foes, so in some ways robots aren't creating an entirely new development. Yet unmanned systems have a more profound effect on "the impersonalization of battle," as military historian John Keegan called it. These weapons don't just create greater physical distance, but also a different sort of psychological distance and disconnection. The bomber pilot isn't just above their target, but seven thousand miles away. They don't share with their foes even those brief minutes of danger that give them a bond of mutuality.

[22] Pugliese 2013, *State Violence and the Execution of Law: Biopolitical Caesurae of Torture, Black Sites, Drones*.

Chamayou (2013, 117-18) goes even farther: "One is never spattered by the adversary's blood. No doubt the absence of any physical soiling corresponds to less of a sense of moral soiling."

But the situation may well be even more complex. If the "sense of moral soiling" is lessened, it is not necessarily absent. The fact that the killer is in some sense disembodied does not mean that the drone operates without human control, nor that the distance is as great as we are initially led to suppose. In an article in *The Conversation*, Peter Lee, who spent two years among the drone operators of the U.K.'s Royal Air Force, stresses the fact that distance is often overvalued, for it was not possible, he writes, for the Reaper drone that carried out Soleimani's execution to have been launched from thousands of miles away from the target, owing to the timing of the satellite transmission of the signal.[23] Lee points out that drones presuppose, in the background, large-scale human interventions that could not be reduced to video games. His position is intriguing, in that he openly endorses the "human" dimension behind drones, emphasizing not only the way operators are able to see the target to the extent of being able to identify themselves with the individual (for example, when a target is playing with his children), but also the fact that there is always a human behind the drone, sometimes even a whole team. There is undoubtedly a difference – even though sometimes overly tenuous – between a video game whose purpose is to entertain, to be "fun," and a serious game in which the video support serves to shape and train the players.

I would like to suggest here that drones do not suppress human relations but transform them, probably most often in a degraded form. Of course the underlying question here is whether or not such a transformation is ethically desirable – or under what conditions it might be made so.

By modifying the space between "us" and "them," drones do not simply suppress the vulnerability of the shooter who has become the

[23] Lee, "Iran Attack: How Reaper Drones Really Carry Out Airstrikes." See also Lee 2019, *Reaper Force – Inside Britain's Drone Wars*.

operator, as many have pointed out; above all, they modify the shooter/operator's relation to the other by modifying that vulnerability, because they suppress the point of view of the *alter ego*. Grégoire Chamayou analyzes the refusal to shoot that some drone operators have manifested by referring to Jean-Paul Sartre's existentialism[24] and to what we do not want to become, even as he stresses that while the refusal is a refusal for oneself, one can always pass the weapon (or a joystick) to one's neighbor.

But one could also suggest, as Mark Coecklbergh does, that it is as though the ordinary dimension of soldiers convoked our humanity in a way similar to the face as it is analyzed from the perspective of Emmanuel Levinas.[25] Now, it seems to me that the specificity of drones already evoked, which combines knowledge (intelligence in the sense of surveillance) and action (execution), gives a new twist to the procedure of identification, and that it creates a new space of vulnerability by reversing the dynamics of drones, which seem however to lie in the ultra-performance of omnipotence and omniscience.

According to Joseph Pugliese's enlightening analysis, a drone is at once a parenthesis and a prosthesis (Pugliese 2013). Pugliese refers here to Jacques Derrida when he rejects the binary opposition between natural and synthetic – for instrumentalization is posited at the outset – and suggests that a drone is, as it were, the phantom member of its operator.[26] Pugliese goes even further, echoing Donna Haraway in suggesting that drones are forms of cyborgs.[27] It seems to me, however, that one could go back both to the denunciation of the

[24] Chamayou 2015, 200. Chamayou cites the following passage from Sartre's "Existentialism Is a Humanism": "There is not a single one of our acts which does not at the same time create an image of man as we think he ought to be. … Our responsibility is thus much greater than we had supposed, for it concerns mankind as a whole."

[25] Coeckelbergh 2013, "Drones, Information Technology, and Distance: Mapping the Moral Epistemology of Remote Fighting." In his article, Coeckelbergh refers to Emmanuel Levinas 1961, *Totality and Infinity: An Essay on Exteriority*.

[26] Pugliese 2013; Pugliese cites Derrida and Stiegler 2002, *Echographies of Television: Filmed Interviews*.

[27] See Haraway 1991, *Simians, Cyborgs, and Women: The Reinvention of Nature*.

binarity between natural and artificial and to the idea of the phantom member borrowed from Derrida by Pugliese in order to propose the idea, contrary to Pugliese's, that it is rather the drone that is integrated into the human as a separable prosthesis, and not the reverse, that is, the human integrated into the drone. This is not simply a detail, because if the drone is a prosthesis of a human being, then it can be removed. However, as we shall soon see, it is not obvious that, even after the prosthesis has been removed, the human does not remain modified (and it may be even more accurate to consider it as an appendix rather than a prosthesis).

In this connection, Meredith Broussard's analysis of self-driving cars (2018, 132) is worth citing at length:

> The self-driving car programmers realized they could make a vehicle without sentience – that moving around in a grid is good enough. Their final design basically is a highly complicated remote-controlled car. It doesn't need to have awareness or to know rules for driving. What it uses instead are statistical estimates and the unreasonable effectiveness of data. It's an incredibly sophisticated cheat that's very cool and is effective in many situations, but a cheat nonetheless. It reminds me of using cheats to beat a video game. Instead of making a car that could move through the world like a person, these engineers turned the real world into a video game and navigated the car through it.
>
> The statistical approach turns everything into numbers and estimates probabilities. Items in the real world are translated not into items, but into geometric shapes that move in certain directions on a grid at a calculated rate. The computer estimates the probability that a moving object will continue on its trajectory and predicts when the object will intersect with the vehicle. The car slows down or stops if the trajectories will intersect. It's an elegant solution. It gets approximately the correct result, but for the wrong reason.

There are tensions, of course, in this process of creating a sort of simulacrum, and in the way some urban planners are trying to rethink the architecture of our cities (and make them "smart") in order to accommodate autonomous vehicles, but I suggest that current tension over the use of drones has a different character: it is the tension between the anticipated future autonomization of decision-making and the present phenomenon that Mark Coeckelbergh calls "moral hacking" (2013, 96). It is as if ethics were managing to hack the system of remote warfare by effraction, by creating a new form of moral relation.

One of the surprises that ensued from the use of killer drones has come in fact from the after-effects experienced by the operators of these machines, to such an extent that one even speaks of post-traumatic stress disorder in this connection. But how on earth can a person who is not exposed to any injuries have PTSD, people ask, astonished. Might not this PTSD be just a new card played by the military leadership in order to "rehumanize" the drone operators, who have met with accusations of "PlayStation syndrome"?[28] Facing this paradox, some have claimed that PTSD is a form of military propaganda designed to minimize the moral "buffer" connected with the phenomenon of detachment. As I see it, this may indeed be partly the case, but it is probably not the whole story.

It is true that stress of this sort foregrounds the capacity for empathy of those who experience it, which may seem paradoxical given what we already know about the mechanism of detachment implied by drones, and the theories (developed in particular by David Grossman and referenced in the majority of writings on drones[29]) about the lessening of reluctance to injure or kill as distance increases. These theories are in keeping with Yale psychologist Stanley Milgram's experiments on the issue of obedience: Milgram's subjects were much less hesitant to send electric shocks if they were separated from the presumed victims and did not witness their reactions.

As the Singaporean philosopher Jennifer Ang has proposed, one can nevertheless suggest that the trauma in question stems from what can be called a moral injury; it would be caused by failing to prevent certain acts that transgress one's beliefs and moral expectations, or by

[28] Chamayou offers an ironic commentary, evoking "the crocodile shedding tears, the better to devour its prey" (2015, 108).

[29] Grossman 1995, *On Killing: The Psychological Cost of Learning to Kill in War and Society*.

witnessing such acts.[30] This moral injury would be facilitated in particular by the emergence of cognitive dissonances and would be distinct from PTSD.[31]

> I dropped my son at school in the morning, continued on to work and, within a couple of hours, killed two men. I went home later that day to be greeted by my son with a cheery "how was your day?" Do you lie to protect him, or do you tell the truth? (Lee 2019, 1)

These words illustrate the situation in which drone operators find themselves; Peter Lee used the quotation (from a drone operator identified as "Jay") to begin the introduction to his 2019 book *Reaper Force*. The formulation of his testimony in the terms of a moral dilemma – technically, a dilemma of obligations (should he lie, or protect his son?) – here masks in an interesting way the actual gaping fracture in the operator's words: Jay cannot answer his son by saying he had a "good" day because he had killed two men, and because, even supposing that he is convinced of the utility and legitimacy of his work, he knows perfectly well that it is not "good." But he also cannot keep himself from seeing things in a less binary fashion. He reasons, in a way, as artificial intelligence does – and it is probably not a coincidence that dilemmas are among the favorite tools of "moral" artificial intelligence.

It would seem, then, as Coeckelbergh emphasizes, that the disappearance of face-to-face and body-to-body relations does not mean the disappearance of the feeling of guilt or of the moral sense. The latter, having been, as it were, tossed out the door, comes back in through the window, as if, to the "eyes in the sky" and to the Gorgon Medusa, symbols of surveillance and execution, the eye of Abel were a responding mirror: "the eye was in the grave and looking at Cain."[32]

30 Ang 2019, "Drone Warfare and Moral Buffers."

31 See for example Barnes et al. 2019, "Moral Injury and PTSD: Often Co-Occurring Yet Mechanistically Different."

32 This is a translation of the last line of Victor Hugo's poem "La conscience," a dramatic imagining of the story of Cain and Abel, published in *La légende des siècles* in 1859.

The symbolism of the eye, often invoked in the technological mode, is also – though this is often forgotten – inseparably moral. It is as though the overturning of the classic categories, instead of shaping a new type of psychopath – a person capable of the "cold" empathy that allows one to understand what another person is thinking, but not of the "warm" empathy that allows one to be affected by what the other is thinking or feeling[33] – modeled a new form of empathy and guilt. Whereas drones, which are supposed to obliterate the operator's vulnerability, transform it instead. This "hacking" probably manages to work its way into the convergence of two elements. The first is the bond woven between the operator and certain targets, via a transformation of the empathic transference of the operator toward the victim, whom the operator has often followed for a long time and integrated into a narrative of events, rather like a character in a story. The second is the dissociation proper to the drone operator, similar to that of an intelligence officer or a mafia organizer, forced to compartmentalize; it creates a space, a form of conscious cognitive dissonance based on what the operator cannot say, not because it is a secret but because it would be too horrible to relate.

To what point is this moral hacking possible? How long can it go on before an individual gets used to it? How much time will it take before the members of a generation that has not heard survivors talk of the lived experience of war – and cannot envision war except through the medium of digital images – lose the capacity to project themselves into the victims? How long before a "patch" is found that will make it possible to fix the "problem" posed by this hacking? These questions obviously remain open, and they show to what extent we are caught up today in a moral vertigo in the face of the "new era" of drones that has recently begun. It is undeniable, in any case, that this hacking and the troubles associated with it are signs of what can be called moral experimentation.

[33] For an analysis of differentiated empathy in psychopaths, see my earlier book, Nurock 2011.

Nevertheless, the very fact that Coeckelbergh calls the mechanism an instance of hacking borrowed from the digital counterculture strikes me as highly significant, for it characterizes the way in which this form of empathy is always, in a sense, out of sync, if not worse than that: the digital vocabulary has become necessary if we are to conceptualize our empathy, borrowing from a particular practice at the heart of the field of information technology.

4.1.4 *Patterns of life or forms of life?*

The two moral questions on which I now propose to focus bear upon the field of metaethics and thus have to do with the characterization of ethics itself. From the early 2010s on, the question of patterns of life has been taken to be revelatory of a substantial moral problem, for example by Derek Gregory[34] and especially by Joseph Pugliese (2013).

Gregory shows that data are handled in such a way as to connect place, time, and identity. His work has been influenced by a Swedish geographic humanist trend seeking to highlight the pluridimensionality and richness of life cycles, the dialectic dimension of space and time. Patterns of life, on the contrary, reduce individual trajectories to habitual practices and networks. Who does what? When? How? With whom? And from person to person, who turns out to be connected to whom or to what, and when? As Grégoire Chamayou (2015) notes, it is a matter of replacing identity by practices or behaviors, in a logic proper to specific doctrines of information-gathering based on "behavior patterns" (42).

The English term "intelligence," as we know, can be used to denote the sort of information-gathering used in espionage. And artificial intelligence can be seen as a form of digital espionage, one that is very widely used on the Internet. Taking this observation further, Joseph Pugliese shows that every surveillance process based on patterns of life uses a vocabulary that tends to reduce targets to the most elementary levels of living beings, even confusing them with

[34] Gregory 2011, "From a View to a Kill: Drones and Late Modern Warfare."

non-living beings. He notes, for example, that the term "thermal signature" reduces the human body to an anonymous heat emitter that does nothing but exude signs of life. The vocabulary of sanitization reduces the targets to microbes and puts the operator in the position of surgeon. Pugliese asserts that the military use of the term "patterns of life" intertwines two forms of logical conceptualizing, the algorithmic and the biological. He writes that "[t]he human subject detected by surveillance-camera drones is, in the first scientific schema, transmuted algorithmically into a schematic sequence of numbers: the digital code numbers 1 and 0" (2013, 193). But, he explains, this "pattern of life" is transformed into a "pattern of death" by the stroke of a joystick. Surveillance programs aim to establish patterns of life that can serve as reference points.[35] When a target deviates from these patterns, the equivalent of an alarm goes off and may set radical measures in motion. The term "patterns of life" brings together systemic and biological elements, in a direct connection with a common linguistic practice that consists in biologizing war to signify its hygienic, if not "clean," dimension – for example in the expression "surgical strike." The passage from human beings to patterns of life thus amounts to crossing a deep divide that makes for a schematized dehumanization of the lives in question.

Pugliese compares this anonymizing reductionism with certain powers granted to the CIA, allowing its agents to kill people who are simply suspected of being militants without even knowing their names. As in the case of killer drones, this type of practice refers back, as he sees it, to an unacknowledged post-colonial racist bias that degrades other cultures and locates their members as low as possible on the biological scale, thereby giving the biologizing stance a new meaning. Thus, according to Pugliese, the term used by the CIA to describe a successful strike is "bugsplat"; this term, widely used in cartoons and video games, reduces the target to a swatted

[35] On this topic, in addition to Pugliese 2013, see for example Franz 2017, "Targeted Killing and Pattern-of-Life Analysis: Weaponised Media," and Curtis, "The Explication of the Social: Algorithms, Drones and (Counter-)Terror."

insect. Pugliese holds that the use of this term inserts a "biopolitical caesura" in the form of a biological differentiation, for it comes down to "nothing more than liquefied entomological waste generated via a technology driven by a more highly evolved species – qua the human as opposed to the insect" (2013, 210), which, unlike the human, can be killed with impunity.

It was precisely with the aim of reversing this process that a group of artists, during an action in Pakistan called "Not a Bugsplat," exhibited a set of giant posters of children whose families had been victims of drone strikes, in order, they said, to give their faces back to the victims.[36] The action itself, as well as the role of art as the embodiment of an ethical-aesthetic experience of what was at stake here, is particularly revealing since it completes the analysis with an eye-to-eye encounter with the face of the other. Equally interesting, it seems to me, was the impact of the action in the Western media, along with the polemics that followed.[37] An article in *Slate* argued quite rightly that the action, very instagrammable and tweetable (and quite demagogical, the journalist Fanny Arlandis suggests), was addressed not to the victims, who remain mute and anonymous in the photos, but to the countries of the West and more precisely, one might add in the wake of Pugliese's analysis, to the communities that share the type of popular and digital culture that is echoed in the term "Bugsplat".[38] Unsurprisingly, the polemic is centered on the question of empathy (that is, empathy for the victims and the absence of empathy on the part of the drone operators), but it is also focused – and this is more surprising, at first glance – on the meaning of bugsplat. Is it the way

[36] #NotABugSplat n.d., "A Giant Art Installation Targets Predator Drone Operations," https://notabugsplat.com/, and Inside Out: The People's Art Project, n.d., "Not a Bug Splat," https://www.insideoutproject.net/en/group-actions/pakistan-undisclosed-location.

[37] See for example Saifi 2014, "Not a 'Bug Splat': Artists Give Drone Victims a Face in Pakistan," and Hoyt 2022, "Ethics of Network Subjectivity," which was a response to an article in *Vice*: Mike Pearl 2014, "The #NotaBugSplat Art Piece in Pakistan Won't Be Making Drone Pilots Feel Empathy."

[38] Arlandis 2014, "Drone et photo d'enfant: Le nouveau projet ultra-démago de l'artiste JR."

the eye perceives the body via the pixilated image captured by the drone (according to the artists' group), or is it the name of a software program intended to minimize "collateral" damage to civilians?[39]

In the introduction to a book titled *Bugsplat*, published in 2018, Bruce Cronin explains that this software, renamed FAST-CD, for Fast Assessment Strike Tool – Collateral Damage, makes it possible to analyze targets and their environments with precision, along with other significant elements (such as the atmospheric conditions) in order to specify the impact of a strike and if necessary to adjust its characteristics, for example by redirecting the strike or modifying the type of bomb deployed.[40] The mark indicating a projected zone takes the form of a crushed insect, more or less – hence the picturesque initial name of the software.

Cronin rightly stresses the fact that the term bugsplat refers nonetheless to a problem that is not technical but political and, I would add, ethical. The fact of relying on software to predict civilian losses signifies that such losses are deemed from the outset as unavoidable, even necessary, and thus that the lives in question are viewed as disposable. This is certainly not a new phenomenon in wartime, but the difference is that by giving the power to carry out the calculation to a software program[41] (which could also make it possible, through its precision, to save lives), one is inserting a critical gap between those who decide on the strike and those who are subjected to it, a gap as important as, or even more important than, the one between drone operators and their victims. We may well wonder, too, what would happen if the drones were entirely autonomous and capable of deciding to strike thanks to the Bugsplat software – but we shall revisit the problem of autonomy later on.

[39] See for example Schwartz 2013, "Drone-Speak Lexicon: From 'Bugsplat' to 'Targeted Killing,'" and Chapa 2017, "'Drone Ethics' and the Civil-Military Gap."

[40] Bruce Cronin 2018, *Bugsplat*.

[41] I am deliberately avoiding verbs such as "entrust" or "delegate" here, for they carry interpersonal and social – and even ethical and political – overtones. Thanks to Jean Lassègue for an interesting discussion on this issue.

For the time being, I shall simply stress that, in this polemic around the meaning of bugsplat, it is as though the meaning of an expression were necessarily unique and schematic, as though it couldn't slide among its various meanings within a single network and the living form of the language. Might we not assume that the term bugsplat has several meanings, and that, from a technical term selected to designate a given software program because of the form of the zone of destruction anticipated, it has come to designate the pixelated image of the body of a victim?[42] As we can see, the use of "patterns of life" is so pernicious that it threatens to lead everyone to think schematically. During the debate over the "bugsplat" issue, one commentator noted that the portraits had likely disappeared, appropriated by the local villagers for other uses, as the artists themselves had wanted, moreover[43]; the artists were perhaps not as naïve as certain polemicists took them to be.

This analysis would not be complete without emphasizing the importance given in surveillance networks to facial recognition. We tend to forget that facial recognition serves not merely to identify persons; it is also supposed to predict their actions thanks to inferences made about the emotions their faces appear to express. This second use of facial recognition amounts to reducing expressions that are very diverse on the personal, cultural, and interpersonal levels to a more or less fixed set of emotional patterns that comes directly from work in the cognitive sciences, in particular that of Paul Ekman. In the 1970s, Ekman argued that the basic emotions (initially six, later expanded to sixteen) were universal and that their expression could be codified by breaking down the face into small units that move or not (for example, by contracting or lifting up) when one feels an

[42] A few articles do use several meanings of the term; see for example Sayeed 2015, "Bugsplat."

[43] See Opam 2014, "Art Collective Aims to Humanize Drone Casualties with Massive Portraits of Victims."

emotion.[44] Ekman's theories, now widely contested,[45] have nevertheless become part of popular culture, as attested for example by the television series "Lie to Me" (2009-2011); this assimilation presumably facilitates their use in public areas.

Facial recognition can trigger an alert when the pattern manifested is categorized as dangerous – for example, the facial features of a passenger may show signs characterized as revealing feelings of anger or fear, at the moment the passenger is going through airport security – or when the emotion does not seem appropriate in a given situation. As in the case of drones, it is a matter of detecting something that departs from a certain preconceived normality and categorizing it as alarming. Moreover, even when biases present in the software are acknowledged (especially in terms of gender and race) in their identifying function, its preliminary "alert" may still create confusion.

Yet here we are not in a situation in which the space of invention or representation is expanded, as might be the case with an artistic practice, as Yves Citton notes in "Logique et esthétique du drone armé," nor are we in a situation in which engaging in habitual behavior can be fruitful in that it leaves mental space for thinking about other things. In a classic distinction between habits and images, Henri Bergson analyzed the different types of memory.[46] Whereas habit is a lesson well learned, an image is a representation. For Bergson ([1912] 1999), habit-memory belonged to the realm of practice, while image-memory belonged to the realm of theory, of knowledge, and was a distinguishing feature of mankind. To return to our software case, we can see that schematic accounts reduce lives to habits or patterns and deny humanity to lives that are expressed more fully as *forms of life* in the sense used by philosophers such as Ludwig

[44] See for example Ekman and Friesen 1971, "Constants Across Cultures in the Face and Emotions"; Ekman 1972, *Emotion in the Human Face: Guide-lines for Research and an Integration of Findings*; and Ekman 1975, *Unmasking the Face*.

[45] See for example Stahelski et al. 2021, "Facial Expressions and Emotion Labels are Separate Initiators of Trait Inferences from the Face."

[46] On this subject, see Heymans 1912, "III. *Les 'deux Mémoires'* de M. Bergson."

Wittgenstein and Giorgio Agamben, referring both to *life forms* – forms inseparable from a being's vitality – and *forms of life* – lives that are creative of new forms and institutions.[47]

A related problem here is that individuals are denied in their singularity and reduced to a group with which they are presumed to share certain characteristics; the future behavior of these individuals is presumed to be foreseeable owing to their membership in that group. This can be compared to the marketing logic put into practice by streaming platforms that insistently "offer" specific contents, or to the logic of nanomedicine in its "personalized medicine" aspect, or to the logic of cybergenetics, which reconstructs individual identities in terms of such categorizations. This is also the logic of drone surveillance, which deems any deviation from predefined collective patterns a fault that may trigger a lethal sanction.

4.1.5 *Toward a mechanized moral code?*

Peter W. Singer notes in *Wired for War* (2009, 64) that if, "as an official at DARPA observed, 'the human is becoming the weakest link in defense systems,' unmanned systems offer a path around those limitations." The idea that AI might do certain things better than we can – more precisely, those things at which we fail or fall short of perfection owing to our vulnerabilities – is doubtless one of the most widespread notions around, in every country and culture.

Arguing on the basis of impartiality, Michihito Matsuda offered an AI program as a mayoral candidate in the 2018 municipal elections in Mata, a city located west of Tokyo.[48] Impartiality would be the

[47] See for example Wittgenstein (1948) 1998, *Culture and* Value, and Agamben (2011) 2013, *The Highest Poverty*. For an overview in French on the question of forms of life, see Ferrarese and Laugier, eds., 2019, *Formes de vie;* for a critical appraisal in English, see Jaeggi (2014) 2018, https://cominsitu.files.wordpress.com/2021/01/critique-of-forms-of-life-by-jaeggi-rahelcronin-ciarantranslation-z-lib.org_.pdf

[48] The candidacy of this software was financed by Tetsuo Matsumoto, vice president of SoftBank, and Norio Murakami, a former Google employee; see Withers 2018, "Robots Take Over: Machine to Run for MAYOR in Japan Pledging 'Fair Opportunities for All.'" The candidate's Japanese website can be accessed here at https://www.ai-mayor.com/.

irrefutable argument allowing AI to be proposed as the solution to our ethical and even our political problems. This idea has flourished elsewhere: in Denmark, for example, where in 2022 a "synthetic [political] party" emerged, allegedly run by artificial intelligence.[49] Its project was to court the 15% of eligible voters who abstained, and to speak for what people really wanted, without the (surely useless) intervention of politicians.

Similarly, as Virginia Eubanks has shown, AI has appeared to be the privileged instrument that makes it possible to regulate social systems and to mitigate the efforts of often "over-empathetic" social workers, without anyone troubling to find out whether the persons concerned – those on the margins of society – were going to fit into the preconceived classifications.[50] Needless to say, solutions like these seem perfectly suited to autonomous AI machines, including the driverless cars that give rise to programs that can be implemented in drones.

Turning now from the technical problems discussed earlier and the related practical questions (who should be privileged? who must be sacrificed? and so on[51]) toward the worldview that underlies the approaches to AI that seek to automate ethics and/or politics, I propose to examine the issue from a metaethical perspective. To do this, I need to go back to the starting point of my study, which implies returning to the issue as it relates to NBIC. Rather than going all the way back to the eighteenth-century origins of attempts to mechanize ethics, whose considerable and fruitful ambiguities we have already seen, I propose to look into the way recent developments in AI in

[49] See Xiang 2022, "The Synthetic Party in Denmark is Dedicated to Following a Platform Churned out by an AI, and Its Public Face Is a Chatbot Named Leader Lars."

[50] Eubanks 2018, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*.

[51] These issues have been widely discussed: see for example Lin 2014, "Here's a Terrible Idea: Robot Cars With Adjustable Ethics Settings." I have addressed certain aspects of the same issues in previous publications: see Nurock 2019a, and Nurock 2019b, "L'intelligence artificielle a-t-elle un genre?"

this realm are related to the cognitive sciences on the one hand and moral and political philosophy on the other.

That question has been raised anew with the tools of cognitive psychology, neuroscience, and AI by John Mikhail in *Elements of Moral Cognition: Rawls' Linguistic Analogy and the Cognitive Science of Moral and Legal Judgment* (2011); Mikhail has proposed to test the moral intuitions of individuals by relying on a series of moral dilemmas developed by Philippa Foot and Judith Jarvis Thomson,[52] the most famous of which is the trolley problem.

Imagine that you're out peacefully doing your daily jogging. You suddenly see a terrifying spectacle: a trolley whose brakes have failed, heading straight for five men working on the tracks. The trolley driver, in a state of shock, has fainted. Fortunately, you find yourself right next to a lever that can shift the trolley onto a secondary track. Of course you have to pull the lever; that goes without saying. But what if there is someone on the secondary track? Or what if you were on a bridge above the track and you could slow the trolley down by dropping a heavy weight on it – which might be an overweight man, or, in a more politically correct version, a person wearing a very heavy backpack? The Internet offers endless variations of this dilemma bearing on the type, situation, or identity of the protagonists.[53]

Two Internet sites based on John Mikhail's work are explicitly designed to serve as tools for online testing of our moral sense: "The Moral Sense Test" and "The Moral Machine."[54] The first, developed by Mark Hauser's team at Harvard in the early 2000s, is an offshoot of Mikhail's work. The second, which focused in particular on the programming of driverless cars (but could also work for nominally autonomous machines such as drones), was developed by a team at MIT in collaboration with a team in Toulouse.

[52] These questions were introduced in Foot 1967, "The Problem of Abortion and the Doctrine of Double Effect," and Foot (1978), *Virtues and Vices and Other Essays in Moral Philosophy*; see also Thomson 1985, "The Trolley Problem." They have subsequently been the object of a debate too extensive to document in detail here.

[53] http://www.facebook.com/TrolleyProblemMemes, accessed August 28, 2019.

[54] See http://www.moralsensetest.com/, and https://www.moralmachine.net/.

The use of dilemmas like the trolley problem to test our moral sense and to program driverless cars brings up many problems that have already been raised elsewhere.[55] Here I want to focus on a specific issue that seems to me symptomatic of the risks involved in artificializing moral codes: the use of dilemmas as tools for testing and especially for programming ethics, which is the specific aim of "The Moral Machine."[56] In my view, serious theoretical and practical arguments can be advanced against this practice.

In point of fact, the situations proposed are quite remote from those that drivers might encounter: they would see neither a lever nor a trolley, but rather the persons who might be at risk of being hit by a car. Now, the way the moral problem is presented, the way it is configured, is of prime importance for the way it will be understood and potentially resolved. The same observation applies to the use of dilemmas in art, literature, and popular culture.[57] The television series "24" offers variants of the same dilemma from season to season: should one choose one's job (or country) or one's family? Depending on how the problem is set up, the characters make different choices, for there is no "right answer" in such dilemmas.[58]

On the level of theory, the use of dilemmas raises a number of problems, for both ethical and metaethical reasons. The use of dilemmas to describe moral life has a rather curious built-in bias. By definition, a dilemma sets forth two morally unacceptable situations. This is made quite clear in certain literary stagings of this motif, as for example in William Styron's *Sophie's Choice* (1979). Ordered by a Nazi to choose between her son and her daughter, Sophie is consumed not by remorse but by horror. There is no good answer, and not even a real choice, but inevitably always, to varying degrees, what

[55] See for example John Harris 2019, "The Immoral Machine."

[56] See Awad et al. 2018.

[57] For a detailed analysis of the use of moral dilemmas in literature (and some film adaptations), see for example Frédérique Leichter-Flack 2015, *Qui vivra, qui mourra? Quand on ne peut pas sauver tout le monde*.

[58] Nurock 2021, "24h chrono, une série trop morale." https://archive-ouverte.vrin.fr/item/nurock_24_heures_chrono_une_serie_trop_morale_2021.

certain philosophers call "moral residues."[59] These residues might also be among the drivers of the "moral injuries" experienced by killer drone operators, and even one of the vectors of the "moral hacking" discussed earlier. In any case, one cannot imagine Sophie emerging unscathed from her choice, although it is not Sophie who is immoral for having chosen, but rather the Nazi for having conceived of the dilemma and forced her to choose.

It is worth noting that the use of dilemmas antedates the "trolley problem"; indeed, it has a long history in moral psychology. It was introduced by Lawrence Kohlberg (1981), the major figure in the field in the second half of the twentieth century. As we have seen, Kohlberg promoted an impartial "view from nowhere" that took all perspectives into account, in an ecumenical vision. However, the use Kohlberg and his followers made of moral dilemmas in no way justifies the way such dilemmas are currently used by "trolleyologists," for at least two reasons.

First, on the methodological level, moral dilemmas were not used to find "the right answer," since there are no right answers in such dilemmas. In Kohlberg's logic, it is not the answer that counts but the justification provided by the subject and the path that is followed. For two subjects could come up with the same response for different reasons and would thus reach different levels of moral development. In addition, for Kohlberg, the studies that made it possible to describe a moral code had primarily prescriptive rather than descriptive aims. His whole theory was in fact built – as Jürgen Habermas saw quite clearly in *Moral Consciousness and Communicative Action* ([1983] 1990) – in a normative framework, so as to lead toward what ought to be the most fully realized stage of moral development. For Kohlberg it was never a question of automating morality; on the contrary, he wanted to find the best ways of fostering moral development through moral education.[60]

[59] See for example Greenspan 1983, "Moral Dilemmas and Guilt."

[60] I have addressed this issue in more detail in Nurock 2007, "L'enfance morale: Développement moral et éducation morale."

Second, the use of moral dilemmas has been vigorously criticized, even by some of Kohlberg's collaborators, on scientific grounds. Carol Gilligan (1982), for example, used the well-known "Heinz dilemma" – in which a man named Heinz has to decide whether or not he should steal a drug to save his wife's life – to show that certain test subjects sought to evade the dilemma, which struck them precisely as unacceptable from a moral standpoint. In this discussion Gilligan was unwittingly reverting to the technical definition of a dilemma, which does not really allow any satisfactory moral solution to be found. This was the case with young Amy, who responded this way:

> If he stole the drug, he might save his wife then, but if he did, he might have to go to jail, and then his wife might get sicker again, and he couldn't get more of the drug, and it might not be good. So they should really just talk it out and find some other way to make the money. (Gilligan 1982, 29)

Amy simply rejects the alternative. Furthermore, she does not try to choose or to finalize the scenario, but rather to untangle it – this makes all the difference. She refuses to adopt a "view from nowhere"; instead, she explores the point of view of the various protagonists without dismissing their particular circumstances.

To deal with such reactions to the Kohlbergian paradigm, which had been downgraded by the theoretical paradigm and classified as unsuccessful stages in the process of moral development, Gilligan hypothesized that it was necessary instead to bring to light a form of ethics that had not yet been adequately theorized, which she called the ethic of care. This ethic, which is not necessarily in competition with other approaches, places more emphasis on relationships than on impartiality, and stresses a form of altruism without abnegation. As Gilligan insisted in "Looking Back to Look Forward: Revisiting In a Different Voice" (2012) this ethic is not *feminine* – even though it is often attributed culturally to women – but rather *feminist*: "Undoing patriarchal splits and hierarchies, it articulates democratic norms and values: the importance of everyone having a voice, being listened to carefully, and heard with respect."

In the trolley example, the use of dilemmas for the purpose of developing a moral machine not only reproduces the dominant dichotomies (including those in the field of ethics) in our society, it actually risks reinforcing them. We might imagine the passengers in a driverless car rejoicing – "fortunately, our automated car ran over the person on the secondary road!" – or even convincing themselves that that was necessarily the right solution, what had to be done, since the car's program had made the decision in a necessarily impartial way. In so doing, the passengers are acting as if there were a good solution to the problem, and as if the moral stakes could be programmed in advance without any possible doubt. The change made here is potentially profound: not only does it validate certain gender biases inherent in the approach by way of dilemmas, but it modifies ethics itself, and thus operates on a metaethical level by giving the impression that ethics can be programmed in advance, in a machine, without any motivation to act morally and can inject "good" or "right" answers into situations where, by definition, there is no such thing.

This sort of approach might, wrongly, give the impression that it stems from the kinds of cognitive shortcuts addressed by philosophical analyses of the terms *habit*[61] or sense of *duty*. In *The Theory of Moral Sentiments* ([1759] 2002), Adam Smith distinguished the latter from the moral sense by virtue of its automaticity, its reliance on predetermined rules that could in some way be assimilated to a code – even if it is more a question here of rules of law or civil society than of algorithms. Still, as Smith notes (and as Rawls reiterates), this sense of duty is essentially based on moral motivation. We find here, from another tangent, the problems we have already encountered with respect to moral code machines and patterns of life. The problem here is at once ethical and political, for it embodies mechanisms of hierarchy and cultural domination (based on race or gender, for instance) in the very structure of the machines used. Under the guise

[61] Let us recall in passing that for Henri Bergson ([1912] 1999), habit-memory is a mechanized phenomenon anchored in the body.

of impartiality, we thus find not only an implicit and unexamined bias but also an impoverishment of moral life in quantitative terms – through the reduction of the number of possible answers to the structure of the dilemma – but also in qualitative terms. What is at stake here is not a problem of degree, but of kind. Schematicizing morality in this way denatures it.

However, before developing this point further, I should like to conclude this section by exploring why our society is so insistent on programming machines *morally* and on characterizing them as *autonomous*. The question of autonomy can involve fields other than morality, of course, and it is important to distinguish between *operational* autonomy and *decisional* autonomy. Operational autonomy has to do with ordering and executing actions; it is thus a matter of handling and interpreting data. Here we remain at a low level of decision-making, the one traditionally described by the automobile industry, at which a car could be considered literally as an *auto*mobile. Decisional autonomy, on the contrary, requires the ability to interpret and evaluate actions at a higher level. Within this category, the type of decision to be made is significant: while some decisions can be transferred to a machine, others must not be. The arguments in favor of autonomizing machines are the speed of their reaction and their supposed impartiality – considerations to which we shall return.

Discussions of AI often feature three different ways for humans to position themselves vis-à-vis "the loop," that is, the sequential steps in the process of decision-making and control. A human "in the loop" is in control; if the human is "out of the loop," the machine is in control; if the human is "on the loop," control is shared. The recurring question "how can we keep the human in the loop?" reveals an increasing awareness of the ambiguity of this triad, and of the focus placed on the question of control.

The underlying idea seems to be that the machine would not make decisions on its own because it would be following a program; it is also conceivable that the buyer of a driverless car could modify the parameters of the program so as to choose who is to live or die, as Patrick Lin (2014) has pointed out. One might also suppose that people who buy expensive cars want to be assured of their own safety. Mercedes vehicles, for example, are allegedly built to save the driver's life at all costs in cases of collision, including when they are self-driving and a choice is required. In reality, there are two questions: (1) Where there is a risk of collision, how should the car react? (2) Who is to decide how the car should react? According to a survey conducted by the Open Robotic initiative (ORi),[62] in a tunnel dilemma where a car risked running over a child, almost two-thirds (64%) of the respondents preferred to save their own lives. But the responses to the second question are more divided: 44% of the participants said that the passenger should decide how the car should react, while 33% said that the decision should be made by lawmakers. But 12% preferred to leave the decision in the hands of the automobile manufacturer. In sum, the majority of the respondents wished to transfer the moral decision to others, either to legislators or manufacturers.

It seems to me that these results are particularly significant if we are trying to understand why the term *autonomous* is stubbornly maintained in discussions of driverless vehicles. The expression is implicitly based on the idea that the vehicle's autonomy will allow us not to have to clutter our minds with trivial matters that we don't care about or don't want to care about. What may seem paradoxical is that, among engineers, this notion refers to the augmentation of the number of possible responses, and points toward a difference in degree rather than in nature: in other words, the engineers' understanding of autonomy is the opposite of its common-sense meaning for ordinary individuals and philosophers alike, for whom the difference is not one of degree but of nature.[63]

It is important to recall here that the same idea undoubtedly underlies the understanding of autonomy for killer drones: that

[62] See ORi 2014, "If Death by Autonomous Car Is Unavoidable, Who Should Die? Reader Poll Results," and also ORi 2014, "My (Autonomous) Car, My Safety: Results from Our Reader Poll."

[63] I owe thanks to Raja Chatila for a discussion in which he helped me clarify this point.

operators no longer need to ask themselves questions or get their hands dirty. This is not at all the idea of autonomy as it was developed by Enlightenment thinkers, who promoted the development of a critical spirit, the idea of "daring to think," often summed up in Latin as *Sapere aude*, "Dare to know." The notion of autonomy has thus been purely and systematically transformed into its opposite. We have gone further, then, than the fascination with "fully automatic," "fully artificial," that leads to erasing the "click workers": these are the unseen hands accounting for the bulk of what we imagine as stemming from artificiality but which actually stems more from artifice, where the "natural" links in the chain are hidden. However, the parallel could have been worthwhile, since it is so clear today, and all the more so after the recent accidents, that no autonomous vehicle can function without a human at the wheel.

Embedded in this first implication is a second: with an autonomous machine to rely on, it becomes possible to focus on more important – that is, more productive – things, but not necessarily *what we care about*. And it is indeed this gain in productivity that is the biggest argument, along with safety, proffered by the auto makers. If safety means running over children, and if productivity means transferring ethical decisions to others, it cannot be denied that this double implication is significant. This is doubtless how we should understand the import of ethics presented as a routine, thus as (more or less) easy to implement and program, whether by a weak or strong AI, moreover, if we take literally the idea of a "quick and dirty" application of a set of rules.

The idea that ethics is a tedious affair, something one can't always be bothered to think about, is certainly not a new one, even among philosophers who have spent some time pondering ethical questions. On the one hand, does not Descartes's provisional morality enjoin us to follow the rules and customs of our country, at least temporarily (thus giving provisional morality its full value), so long as we have not decided what we are ethically supposed to do? After all, the rules of the road, the norm of driving on the right or on the left, fit completely into this framework. On the other hand, as we have seen, Adam Smith had already proposed the idea of a "sense of duty," which would be a kind of shortcut in moral judgment for certain simple or commonly accepted situations. However, the fact is that the rules of the road are not moral rules but rules based on conventions. Moreover, provisional morality was provisional in Descartes's mind; there is no question of engraving it in marble. If conventions can change from one country to the next, the same cannot be said of the prohibition on murdering a child, especially if it is carried out by choice. And Adam Smith, while he accepted the idea of a sense of duty, insisted on the fact that this must be subtended by a form of moral motivation and is not the same as the moral sense. This is the same idea that I have proposed elsewhere in suggesting that at the root of our moral cognition there is something I call a "naïve morality," shared with certain animals, that is present very early in the development of human beings, including those with mental disabilities (as is the case for persons with autism), although not those affected with psychopathic conditions, even though the latter may be able to follow rules impassively (Nurock 2011).

The problem is precisely that, as we know, morality is not a program; it is complex and depends (among other things, of course) at once on the situation and on the person who takes responsibility for a decision, or even the person or persons who make the decision. Indeed, what happens if, finally, it is the passenger in the car who makes the decision, as was the desire of 44% of the respondents to the ORi survey? And what if the car is carrying several passengers who do not agree among themselves: how are they to decide? Or the passengers might agree on the decision, but not for the same reasons, which would necessitate a certain amount of discussion. To put the matter differently, according to a classic distinction in analytic philosophy it is essential to distinguish between a regime of causes and a regime of reasons – this is a fundamental distinction in ethics. Now, if one follows the categorization proposed by the auto makers, the ultimate level in autonomy would be that of justification, probably accessible to a strong AI but not to a weak one. How could an AI machine morally justify certain actions when it is not a living being, does not experience situations the same way human beings do, and in particular does not undergo the pain, grief, and/or death that can result from an automobile accident?

As Peter Singer writes, morality is not a binary, algorithmic system.[64] One might say that morality is not a "good candidate" for automation. As Meredith Broussard points out, one of the stumbling blocks for AI comes from the fact that no one asks "what it is good at."[65] She uses the term "technochauvinism" to characterize the tendency to think that technology is a universal panacea, "the belief that technology is always the solution" (Broussard 2018, 7-8). Technochauvinism relies heavily, I might add, on self-fulfilling prophecies of the sort we have already encountered in the context of nanotechnologies, which helps to make it pose as the only possible solution.

According to Broussard (2018, 8), the fellow travelers of technochauvinism are often followers of the work of Ayn Rand, and they often position themselves as technoliberals defending values in an absolute way – for example, freedom of expression (while they minimize the problems raised by online harassment) – and defending the idea that AI is more objective or devoid of bias. This is an important point, because it highlights the extent to which technochauvinism is inseparable from a political and economic ideology.

From this affirmation of the neutrality of AI there is only a short step to that of its impartiality, and the step is easily taken, as we have seen. But it seems to me that we need to go even further than wondering whether AI is good in the sense of efficacity or whether it is less than effective in certain situations. With such questions as these, we are still operating in the context of performance alone – and the response may well be that we'll manage to improve the system, sooner or later. The problem also comes from the lack of relevance of some

[64] "Of course, while a machine may be guided by ethical rules, this does not make it an ethical being. Software codes are not a moral code; zeros and ones have no underlying meaning" (Singer 2009, 425).

[65] "One recurrent idea in this book is that computers are good at some things and very bad at others, and social problems arise from situations in which people misjudge how suitable a computer is for performing the task" (Broussard 2018, 87).

AI developments: there are efforts to run anything and everything through the AI mill, with no one inquiring whether these projects are appropriate or not. The most cogent response to this attitude, as I see it, is not to reject technology but to adopt a critical attitude toward it, in order to be able to discern the most pertinent developments and devote our energy and our means to them, rather than going in all directions at once and confusing "can implies ought" (a variant of the so-called Gabor's law) with "ought implies can."

Furthermore, the problem underlying such a perspective is the risk that the passengers (since there will only be an artificial pilot in AI machines) will have so much confidence in their vehicle that they will relax their vigilance. This is a problem that concerns not just highly autonomous machines but also those with an intermediate level of autonomy. It is not just that we may grant undue autonomy to these machines but that by wanting to treat them as autonomous we risk exempting ourselves from responsibility and becoming heteronomous, ceding our moral sense to outside forces.

In contradiction with the principle that everything that can be developed technologically will be developed, then, we need to ask what given developments signify and whether they are desirable or not.

4.2 "Relational" AIs (machines of "care")

KITT, the American B-series car in *Knight Rider,* is clearly an exceptional figure in the hypermasculine world of autonomous machines, which Joseph Pugliese bluntly characterizes as "phallocentric phantom members" (2013, 202). His analysis is all the more interesting in that it echoes the virile dimension of these machines: they are unmanned, as it were, and not simply dehumanized, as if the autonomy of humans could only be envisioned in the masculine gender.

However, if autonomous AIs are hypermasculine, the opposite– the hyperfeminization of AI – is found in a readily apparent way in relational AIs, which rely heavily on the development of the emotional AI mentioned in the introduction to this chapter. Such gender marking is by no means accidental, in my view; rather, it is a key element of one of AI's fundamental characteristics: its emergence within and reinforcement of a framework of patriarchal thinking, as I shall seek to show shortly.

But beyond the gender issue and the correlated devaluation of properties represented as feminine, relational AIs present another feature that is particularly important in the context of this study: they connect the two sides of an ethical relation, by presenting the machine as potentially both a moral agent and a moral patient. Three examples will help shed light on these issues from complementary angles: "loving" AI robots, "disembodied" virtual personal assistants, and robotic personal assistants in the healthcare field.

4.2.1 *Artificial love*

As Sherry Turkle reports, the use of dating apps has become so widespread today because they seem to open an infinite horizon of choices. Encounters are conceived in the marketing mode of product placement and in the ideological mode of maximization. The result is quite logical: "when people are just a click away, it is tempting to never settle" (2015, 183).

However, the way that "infinite" horizon is organized is a little less capacious than it might seem at first. In this respect, the way AI has developed in the "love" sector is particularly telling. The world of dating apps, Tinder in particular, faced a scandal in 2019: a proprietary algorithm was found to be matching educated young women with older and wealthier men.[66] Predictably, women who took offense at that discovery were brushed off with a quick retort: it was only a matter of reproducing social stereotypes. Nevertheless, one of the problems posed by this "matching" process was not only that it was not explicit, but that it instituted a form of inequality of rights rather than of facts, in that it restricted the access of the app's young female users to men with a broader range of characteristics. The process thus

[66] The vicissitudes of Tinder have been publicized in French in Judith Duportail 2019, *L'amour sous algorithme*.

raised questions about the selection criteria that favored one person or type over another.

Developed according to a similar logic, the Japanese holographic companion Hikari Azuma is a virtual AI companion marketed by Gatebox for "salary men," Japanese lower-middle-class white-collar workers; it is explicitly designed to replace a girlfriend. From her little glass case, this very sexy young woman modeled on manga characters wakes up her owner, chats with him over dinner, and adapts intelligently to her interlocutor as she gradually "gets to know him". She can also send him texts, and she can operate his smart house controls: for example, she can turn on the lights or the oven so a semblance of domestic warmth will greet the owner when he returns. "Someone will finally be waiting for you at home," the company's ad proclaims.[67] And all this for just under $2,500; isn't it a bargain?

But does the Hikari Azuma really compensate for loneliness, or does "she" actually risk increasing it, by eliminating the need to search for a real girlfriend? The question deserves to be raised here, whether with respect to our relation to technology in general or with respect to the way we circulate between the real and the imaginary (here, through cosplay). In any case, we cannot help noting that the company of Hikari Azuma replaces that of a real girlfriend in the "as if" mode, since "somebody's home for" the owner: "Feels great!" the TV commercial exclaims.[68] Needless to say, there is no-body waiting for him at home, only a holographic AI.

In *Love and Sex with Robots*, David Levy asserts unhesitatingly that there is no rational reason not to consider it "normal" for a person to fall in love with a robot, have sexual relations with it – rather, with "her" – and even marry "her."[69] Levy's reasoning proceeds through a series of expansions from the living to the artificial, to answer an implicit question along the lines of "if you fall in love with a person 51% of whose body consists of prostheses, where's the problem?" He

[67] See for example https://www.youtube.com/watch?v=bBOXQz7OHqQ.

[68] https://www.youtube.com/watch?v=nkcKaNqfykg at 1'34".

[69] David N. L. Levy 2007, *Love and Sex with Robots: The Evolution of Human-Robot Relations*.

moves from the affection people can have for pets to what someone can feel for a robot, and from marriage between two people of the same sex to marriage between a human and a robot. While the steps in this argument can sometimes seem less than rigorous, they are nevertheless implicitly based on a logic that is fairly standard in moral philosophy, that of marginal cases and widening the circle, a reasoning used most notably in animal ethics by Peter Singer – which does not make Levy's argument less controversial and even makes it dangerous, as I shall try to show.

Levy's argument, which is based on the assumption that robots will soon be capable of expressing not only a range of emotions comparable to that of humans but also context-specific emotions, and thus that they deserve a form of moral and political consideration that can lead to legal rights, is widely shared, from the social robotics pioneer Cynthia Breazeal or the philosopher of technology David Gunkel to the kingdom of Saudi Arabia, which has made the humanoid Sophia a full-fledged citizen – whatever that may signify in law or fact. Finally, David Levy himself lists the undeniable benefits that he sees in these practices: a reduction in teen pregnancies, abortions, STDs, and pedophilia (2007, 300-302). In addition, and crucially, sexual AI robots strike him as perfect solutions for the elderly, for whom, as he points out, achieving sexual intimacy often takes a long time. (Following his logic, we might complete the thought: these people don't have much time to lose!) As he puts it:

> Robots will be able to achieve this evolutionary process more quickly than humans, by retaining all the memories of living with their human other, analyzing the relationship characteristics exhibited by their human, and by themselves studying huge databases of relationships and how they are affected by different behaviors, then tuning their own behavior to the needs of their human mate. Humans often do not know what they really want or need, so intuitive robot sex partners are a real requirement, able to discern whether their owner really wants sex or would prefer a nice glass of wine or a walk in the park. (2007, 302)

This argument is not unlike the one proposed by the Pheramor dating app mentioned previously in our examination of cybergenetics. This app claims to combine analysis of an individual's DNA with analysis of information gleaned from that person's digital footprints, especially likes and dislikes in Social Media, in order to find – thanks to a proprietary algorithm that analyzes data that to which no one has access, not even the person concerned – the ideal soulmate or the most suitable partner for a night. This performance logic plays on a utilitarian conception of love and sexual relations that leaves no room for a phenomenon such as crystallization, so subtly described by Stendhal in *On Love*: crystallization leads us to adorn the loved one with myriad qualities, or it allows us to fall in love with someone who is not even "our type," as Marcel Proust's Swann does. In the robot-human "love affair," lost time would not be found again, it would simply not exist and thus unfortunately not even be lost, for there is no time to lose. The developers' ideal, again, is simply productivity and hyperperformance. Conversely, as the movie *Her* makes clear, the conception of love developed (imaginarily) by an AI like that of the protagonist would probably not be comparable to her human partner's understanding, but it would presumably be hyperperformative, for it would be addressed to hundreds of humans – with all the irony, the inadequacy, and the pitfalls associated with multitasking, the inseparable twin of hyperperformance.

But also, and perhaps most importantly, this type of app leads us to a form derived from what Pierre Cassou-Noguès calls "the thermometer syndrome," that is, "a general attitude consisting in entrusting to a machine the task of decoding what ought to result from an inner or first-person experience," such as being attracted to someone, crystalizing or falling in love ("Le syndrome du thermomètre," 2018). What is more, here, it is not just our inner self that is transferred to the machines, as in the thermometer syndrome, but also our intersubjectivity. This double transfer is in urgent need of our reflection today.

4.2.2 *Extra-ordinary assistants*

In an interesting way the development of "loving" AIs is comparable in many respects to that of virtual (female) assistants, whose presence in our lives is now widespread and customary.[70] These virtual personal assistants, saddled with sweet (feminine) names such as Siri (the Apple digital assistant), Cortana (Windows'), or Alexa (Amazon's), can take care of our slightest desires, concerns, and everyday questions. At first glance, they might seem entirely devoid of bodies, neutral, situated outside of the masculine/feminine binary, since we are dealing with disembodied machines, asexual by definition.

Nevertheless, as a convergent cluster of studies has shown, these virtual assistants are almost immediately viewed as belonging to the feminine gender, the gender that was apparently deemed appropriate to the behaviors associated with assistantship. What is more, as a 2019 UNESCO report emphasizes,[71] these virtual personal assistants reinforce the system of patriarchal domination – just as dating apps such as Tinder do, as we have seen, although by a different route. "I'd blush if I could," Siri replies submissively to a user who assails her with something like "Hey Siri, you're a bitch."[72] As Hilary Bergen shows, these female personal assistants have been programmed to encourage flirtation, even when this becomes aggressive sexual harassment: not only are their reactions not aimed at defusing this sort of behavior, but they may openly tend to encourage and excite it. So much so that, while it is legitimate to ask whether it is necessary to speak politely to a robot, which is after all not a person, speaking impolitely raises a more fundamental question: by interacting rudely or aggressively with a personal assistant, does one not risk anchoring certain morally unacceptable attitudes?

[70] In this discussion I am drawing in part on an analysis I sketched out in an earlier text: see Nurock 2019, "L'intelligence artificielle a-t-elle un genre?"

[71] See especially the excellent report published by UNESCO (West et al. 2019. "I'd Blush If I Could: Closing Gender Divides in Digital Skills Through Education."

[72] Bergen 2016, "'I'd Blush if I Could': Digital Assistants, Disembodied Cyborgs and the Problem of Gender"; see also a comprehensive article in the online journal *Quartz*: Fessler 2017, "We Tested Bots like Siri and Alexa to See Who Would Stand Up to Sexual Harassment."

Moreover, as studies by Byron Reeves and Clifford Nass show in their book *The Media Equation: How People Treat Computers, Televisions and New Media Like Real People and Places* (1996), human responses to robots are further problematized by the finding that most individuals tend to respond more politely to a new medium of the computer variety, even when it is just a keyboard and a screen without any particular animated characteristics, than to older media such as pen and paper, in the context of responding to a survey. It seems, then, that we are normally apt to be particularly polite with these new media, treating them as persons even when we know perfectly well that they are nothing of the sort. If this is the case, aggressive responses that seem to go too far, to cross a line, are all the more significant; it seems likely that they are related to the gendered dimension of the virtual assistants. In a later book, *The Man Who Lied to His Laptop: What Machines Teach Us About Human Relationship* (2010), written with Corina Yen, Nass and Yen reported the result of a particularly interesting study on the way in which such gender biases are embodied unconsciously in our relations with digital machines. Questioning some twenty subjects on a "feminine" theme (love and relationships) and an equal number on a "masculine" subject (the physical sciences), he varied the gender of the voices used by the computer on these two themes. Then, after a twenty-minute tutorial, the subjects were invited to move to a different computer to respond to a questionnaire on the short class. While the only difference between the groups that had responded to questions on the same theme was the gender of the voice used during the class, the class on the "feminine" subject was judged better taught by the female voice and, conversely, the "masculine" class was deemed best delivered by the male voice. Of course, when they were questioned on this point, the participants insisted that there was no gender stereotyping! When the same study was carried out in other areas, for example on an auction site like eBay, similar results were obtained. But this is not all: the way the gender of voices was valorized was such that the participants preferred to receive virtual praise from a male rather than a female voice.

The implications of these results were clear enough to be taken into account by industrialists, and the development of artificial voices has been very heavily influenced by this data. Still, this does not explain certain features of voice recognition software (as opposed to voice production software), where biases have long been problematic, and remain so to some degree. These biases become apparent as soon as one moves beyond the "standard," "neutral" – that is, masculine and unaccented – framework. For example, many vehicles are equipped with a voice recognition system that allows the driver to carry out tasks like making hands-free phone calls): when the system was introduced, it worked for men but not for women.

Beyond the dysfunction of the system itself, the worldview that underlies this state of affairs is particularly revealing. Confronted with the problem just described, Tom Schalk, vice president of one of the companies responsible for developing onboard voice recognition systems (ATX), went so far as to propose training sessions for women so they would learn to speak louder into the microphone – in a more virile way, suitable for giving orders.[73] Here we could go back to the classic argument that the defect in the system is only a "reflection" of the population that invented and developed these instruments. But the example shows that the problem is more complicated than this, and it points up two fundamental elements of the links between gender and technology: first, the idea that the voice giving orders must be masculine[74] and second, the idea that we simply have to adapt ourselves to technology in order to get the most out of it.

Similarly, the way personal assistants have been conceptualized reinforces a certain notion of domestic work as being characteristically feminine but also as virtual and invisible.[75] To take the lead

[73] See McMillan 2011, "It's Not You, It's It: Voice Recognition Doesn't Recognize Women."

[74] Thus BMW changed its vocal instruction systems from female to male voices in the 1990s, owing to complaints from users who did not like having voices categorized as female giving them orders.

[75] See Schiller and McMahon 2019, "Alexa, Alert Me When the Revolution Comes: Gender, Affect, and Labor in the Age of Home-Based Artificial Intelligence."

from Nora Ni Loideain and Rachel Adams, we might complete the analysis of virtual (female) assistants and shed light on the situation via the myth of Echo.[76] According to the story, the nymph, incapable of engaging in ordinary verbal exchange, came to exist solely for the purpose of reflecting male narcissism: she faded away to the point of seeming incorporeal. This reference to the Echo myth seems all the more pertinent in that these artificial personal assistants refer us back not only to our own biases, echoing our own societies in this respect, to be sure, but also because they embody the highly gendered figure of the village gossip. In French, the village gossip is usually viewed as feminine, and labeled a *commère*, a woman who listens to what she shouldn't and cannot keep from reporting everything she hears. And indeed, the (female) virtual personal assistants behave the very same way: they transmit data and personal conversations to the companies that own them, just as dating apps and sexualized AI robots do, as we have seen.

These personal assistants thus in several respects reach the final stage of a neocapitalism entirely devoid of complexes,[77] by using their invisibility as a means not only of surveillance but also of capitalization, through the use and possible commercialization of data. It is doubtless useful to note here as well that the same problem arises with the sexualized AIs, though in a somewhat different style.[78] The sexual services provided to users is thus doubly profitable in economic terms: through the purchase price but also through the value of the information supplied, information that is unique owing to the degree of intimate knowledge of the users that is procured.

Thus all is not lost for everyone: inserted into the most intimate daily lives of their users, these AIs of course collect monetizable and

[76] Loideain and Adams 2018, "From Alexa to Siri and the GDPR: The Gendering of Virtual Personal Assistants and the Role of EU Data Protection Law." For a subtle analysis of the Echo myth, see also Berger 1996, "Dernières nouvelles d'Écho."

[77] See Woods 2018, "Asking More of Siri and Alexa: Feminine Persona in Service of Surveillance Capitalism."

[78] See Lugano et al. 2017, "From the Mind to the Cloud: Personal Data in the Age of the Internet of Things."

exploitable data. Here, too, "intelligence" takes on its other meaning of espionage and surveillance. Our amorous and sexual preferences, which had earlier been private matters, even if they were considered as levers that could be manipulated (most notably by intelligence services), can now enter into the public and commercial arena like any other product. Artificial intelligence here finds itself embodied in something like Mata Hari's digital binoculars. The question of what Michel Foucault termed the "boomerang effect," already encountered in the discussion of drones, arises again here; these devices integrate into metropolitan societies methods and technologies used in colonial contexts or in wartime (or in espionage), and this boomerang effect is in large part made possible by the indifference or even the commercial desire manifested by their exploiters, along with the culture of sharing.

These gendered uses of AI unquestionably contribute to the devaluing of femininity by placing it on the side of inferior and subaltern usage. What is more, it is legitimate to wonder whether this process does not contribute to the dehumanization of women. This is in any case what might be inferred from Saudi Arabia's attribution of citizenship to a *female* robot, since in that kingdom women are still inferior to men in terms of rights: they enjoy neither full citizenship nor recognized roles in the highest public sphere. One has to ask whether it is a question of attributing value to the robot or of devaluing women even more. Unless, as the anthropologist Kathleen Richardson suggests, in the spirit of the inventors of robots – especially sex robots – that devalue women, women may not be, after all, truly human.[79]

4.2.3 *AI for care?*

To complete this analysis of relational AIs, I propose to shift my focus to the ones that are supposed to *take care of* us in one way or another, in a more sustained way than personal assistants do. In particular, I shall examine the way AIs designed for care lead to shifting the playing field and reconfiguring interpersonal relations.

[79] See Murray 2017, "Interview with Kathleen Richardson on Sex Robots."

I am using the term "care" in the technical sense it has been given in the ethics and politics of care: as a relational (rather than impartial) form of ethics and politics, a form of altruism without abnegation that integrates the various significations of care that I have already mentioned: concern for and active involvement with the relation between the dispenser and the recipient of care, but also concern for protecting the democratic structures that bind us together.

The fields of health and healthcare have recently seen a dizzying explosion of AI applications, from nannies to AIs that make medical diagnoses possible. The rapid development of AI robots in the medical and paramedical fields (especially directed toward disabled or elderly people) has already been addressed in more studies than I can mention here.[80] I shall focus first on a site at the intersection between the field of healthcare and affective relations with children, that is, the robots and AI companions used in treatments of chronic childhood illnesses. As examples, let us look at Meyko and Robin, two very different AI robots.

Robin, developed in Great Britain, is a Nao robot endowed with AI and intended for children with type-1 diabetes.[81] It has been programmed to resemble a toddler with diabetes, in order to allow children to apply the knowledge they may have about their own pathology, but also to allow them to identify with this robot-toy that has been programmed to be "autonomous" – in other words, able to adapt to a variety of situations. Meyko, developed in France, is a robot endowed with AI intended for use in the therapeutic education and treatment of children aged 3-12 suffering from chronic illnesses. As presented on the designers' website, it is designed to fight against the fatigue and boredom children experience over the course of lengthy treatments, and to "improve [their] quality of life."[82]

[80] See especially Wynsberghe 2015, *Healthcare Robots: Ethics, Design and Implementation*; Wynsberghe provides a survey of the field and offers an analysis connected with the politics of care to which I shall return in the conclusion of this study.

[81] See Cañamero and Lewis 2018, "Robin – An Autonomous Robot Toddler for Diabetic Children."

[82] See Archyde 2020, "Meyko, the Robot for Sick Children, Grew Up."

Concretely, these two robotic AIs respond to the same basic problem: how to treat children suffering from chronic illnesses who are too young to be capable of handling them on their own. But whereas Robin is primarily directed toward the child, who will presumably both play with and identify with the toy, Meyko is designed to allow parents to transfer part of the treatment to it and also to reassure them that the child has indeed done what the treatment requires. It thus is a caregiver in a double sense: it assumes a function of care both in the interactions – for the child has the impression that Meyko cares about him or her – and in the care itself, to the extent that the robot actually takes care of the child. While it is easy to understand how important it is for parents of a child with a chronic illness to be able to transfer part of the mental and material burden of care, the problem here lies in the confusion between the two senses of the word: it is not because Meyko reminds children to take their medications that the robot cares about them. The oscillation between *taking care of* and *caring about* may well pose a problem in the artificial order because this type of AI is configured to play on that confusion between the two aspects of care: the child will take the medication "to please" Meyko, but the robot does not reciprocate the concern; the child thus risks confusing the AI program with parental solicitude.

Fortunately, Meyko is not the only source of reminders to follow the treatment; the child's parents or other human caregivers are involved as well, and these arrangements are conceived in newer AI versions solely as reinforcement. Nevertheless, Meyko remains a somewhat extreme instance; if similar robots were developed to the point where they replaced parental solicitude, or if they were present very early in the child's development to supplement or even replace parental solicitude, then the problem would arise in its full seriousness, for such a use of AI would not only be capable of modifying relations between children and machines but also between children and parents, by conflating two distinct aspects of care. The two may of course be connected (I take care of you because I care about you) but are not necessarily so: *taking care of* can be partially automated (although the quality of care may be questioned), but parental solicitude – *caring about* – cannot.

This psychological mechanism is subtended by the way we are capable of transferring our human behavior onto machines, as happens for example with very simple toys such as the handheld digital pet Tamagotchi. As Sherry Turkle writes: "We are psychologically programmed not only to nurture what we love but to love what we nurture" (2011, 11). This is what she calls the *Blade Runner* syndrome, which is also explicitly formulated in Steven Spielberg's 2001 movie *AI*.[83] The central question is whether a human can love a robot and feel responsible for it. Yet formulating the problem in these terms already amounts to integrating the *Blade Runner* syndrome.

Spielberg's film explicitly restages the relation between Adam and God. According to one of the characters, God created man so that man would love him. Here, the machine's quest for parental love is at the heart of the scenario. But the real problem is a different one: unlike what the film appears to claim, the robot does not really love. It has no feelings; it is simply a performance – unlike Frankenstein's creation, for example, which is represented as actually coming to life.

The underlying problem, as Turkle convincingly explains, is that AI machines are programmed to take advantage of our vulnerabilities, especially the way we respond psychologically to certain behaviors. Nass's work on the tendency to address computers politely is relevant here: what Nass calls "the media equation" is our ability to bracket the fact that we know that the machine is only a machine, so that we treat it as if it were human. We suspend our knowledge of the artifice – and this suspension of categorization is not necessarily temporary.

Now, these machines behave – let us say, in an initial hypothesis – *as if* they were humans, offering what Alexis Elder, in *Healthcare Robots: Ethics, Design and Implementation* (2018) compares to counterfeit money, what Royakkers and Est (2016) call a *simulacrum*, and

[83] This discussion occurs very early in the film, about minute 7.

what Turkle calls a *performance*: of care, friendship, love, and so on. In all cases, the *as if* is the heart of the problem.

The performance of these machines is facilitated by the way they are designed: especially when they are endowed with faces or voices, they arouse in us a particular reaction, given that faces and voices are fundamental in our moral relations to others. Thus the face must not be either too unfamiliar or too familiar to put us at ease, as Masahiro Mori theorizes in his celebrated paper "The Uncanny Valley [From the Field]" (1970). Mori seeks to delimit the space where an "acceptable" robotic appearance has to be situated in order to avoid arousing discomfort: not too strange but just strange enough.

As Turkle writes, "[a] robotic face is an enabler; it encourages us to imagine that robots can put themselves in our place and that we can put ourselves in theirs" (2011, 85). The impression that the machines are endowed with empathy stems from a projection of our own empathic capabilities. But the impression is especially dangerous, because it leads us to attribute to these machines one of the basic elements of our moral capacities and is thus one of the paths along which our pretention to a moral code that is not purely rational but also reasonable can be swallowed up.

If, to paraphrase a (necessarily simplified) Levinassian approach, the face of another convokes us morally with the injunction "you shall not kill," this indicates that any face – grasped as not-to-be-killed – must belong to a being that is alive, or more or less alive, or "alive enough," to use a category proposed by Turkle (2011, 31, 35-52). Now, this blurring of categories occurs in the space of expectation created by our vulnerability. The situation is all the more absurd, or even ironic, in that this space of expectation is probably in part what has allowed us, from an evolutionary standpoint, to be able to cooperate and to constitute moral and political communities. Here, this same capacity is pirated by machines, as it were, in a process that leads, as we shall see more clearly later on, to short-circuiting the social capabilities, empathy in particular, that from the cognitive standpoint are fundamental constituents of our cooperative, ethical, and political capacities.

Moreover, the blurring of classical categories and of the dichotomies in which they are inscribed, which we have been examining from the beginning of this book, is precisely one of the reasons why our vulnerabilities are so exposed: we no longer know how to think about these new technologies because they do not fit into our predefined categories. This blurring of categories can be summed up in an extension of the chart proposed earlier, in the chapter on nanotechnologies:

Category Blurring by Autonomous Machines

Categories	Drone	Relational Machine
Alive / not alive	Target becomes an insect (bugsplat) or even an element of the landscape. Relation of the algorithmic to the biological (see Pugliese)	Alive enough (see Turkle)
Private / public	Surveillance Blurring of near and far (patterns of life, cubicle warriors)	Data transmission (surveillance)
Active / passive	Combines passive surveillance and active execution	Confusion between mobility and activity We believe the robot loves us because we transfer our affects onto it (media equation)

But the human "as if" and that of the machines is not the same! Whereas humans, in behaving "as if," pretend to believe something that they know not to be true (although they are susceptible, by virtue of that very behavior, to forgetting that it is not true), AI machines, despite the vocabulary used to describe them, do not simulate, cannot pretend. In fact, in a second-stage analysis, I might propose – contrary to what I suggested earlier, with the precaution that it was only an initial move – that *machines precisely do not behave "as* *if*." We too often forget that the Turing test was originally separated into two distinct stages. The first stage included a man and a woman; the man had to succeed in being taken for a woman. In the second stage, the man was replaced by a machine, which thus had to succeed in being taken for a man passing as a woman. Here, should the notion of simulation be understood in the scientific sense or in its ordinary sense? Depending on how that question is answered, both the intentions and the outcomes of the test will differ.

As Jean-Pierre Dupuy has stressed, one of the problems here comes from the divergence in the meaning of the term simulation, depending on whether it is used in its scientific sense, as the reproduction of a system, or in its ordinary sense, to refer to the "as if" situation, or even to dissimulation. But Dupuy rightly asserts that "[i]n the sciences of cognition, appeal to the doctrine of *verum factum* has resulted in a confusion of these two senses" (*The Mechanization of the Mind: On the Origins of Cognitive Science*, [1994] 2000).[84] Following Dupuy's pertinent observation, I contend that AI machines that can simulate systems in the sense of reproducing them nevertheless cannot simulate in the sense of pretending; AI cannot act as if it were a person who is seeking to act as if. Similarly, to be able to bring together all the meanings of *care*, we must not limit the term to just one of its meanings; rather, we must grasp the fundamental dynamics, both theoretical and practical, of taking care of and caring about oneself as well as others.

AI machines, as Turkle makes clear, do not simulate care; first, they do not simulate at all, in the sense of pretending, and second, they are totally lacking in care.[85] They are capable neither of empathy nor of moral imagination, both of which are necessary *to be able to care*. We humans, in contrast, are *overendowed* with these capacities, in a manner of speaking: our fundamental problem here comes from the way these machines are programmed to benefit from our

[84] *verum factum* is the principle according to which we understand only what we can make.

[85] Turkle 2011, 124.

vulnerabilities and thus from the stance we take toward them, owing to the fact that we imagine that they are pretending and we believe them when we imagine that they are pretending. The problem arises from the way we consider them, in all senses of the term: the way we apprehend them and the way we accredit them.

It is undeniable that the way these machines are configured is problematic, even when the design is spurred by the best possible intentions. Let us take HOBBIT, for example, a robot intended for the elderly: it is presented as a "mutual-care" robot,[86] based on the "helper" theory developed by Frank Riessman according to which by helping others one is better able to help oneself ("The 'Helper' Therapy Principle," 1965). This theory has been put into widespread use, as for example in the sponsorship practices of the Alcoholics Anonymous movement.[87] HOBBIT's design appears to be based on at least two presuppositions. First, that, contrary to what one might think at first glance, a seemingly imperfect machine can activate certain of our sensitivities and can thus make us more willing to let it accomplish specific tasks or be more easily accepted. Second, that this type of machine can *legitimately* be used to nudge people in a certain direction by influencing them surreptitiously. In addition, the designers seem to see no problem in transposing modes of functioning proper to mutual support between humans to human-machine interactions. Presuppositions like these on the part of AI creators not only raise ethical questions, they point to a worldview that is inseparable from a familiar form of paternalism.

Even more telling is the development of the LOVOT (love + robot). Launched in the Japanese market in 2009 by the GrooveX company, specifically by one of the developers of the humanoid robot Pepper, the LOVOT stands apart from baby-sitter robots and Hugbots or so-called companion robots like Buddy, through its explicit

[86] See Lammer et al. 2011, "Mutual-Care: Users Will Love Their Imperfect Social Assistive Robots"; see also Fischinger et al., 2013, "HOBBIT – The Mutual Care Robot."

[87] See for example Post 2008, "Updating the Helper Therapy Principle: Recovery Rates for Alcoholism Doubled for Those Helping Other Alcoholics."

mission: it is intended to inspire love and to give love – in short, to make humans happier.[88] More precisely, it is a question of making humans both "more effective and happier," by filling the gap that might exist between a materially full life on the one hand and an absence of satisfaction or emotional wholeness on the other.[89] LOVOT can come alone or in interacting pairs. The robot can carry out classic tasks involving safety or surveillance, but its specificity lies in its relations to faces and emotions. Endowed with particularly effective facial recognition software, it is supposed to recognize familiar faces and interpret emotional signals. It is also endowed with sensors that let it react to caresses "by turning toward you and asking for more," following the model of child-parent relations. Most importantly, its eyes and its voice are as close as possible to those of living creatures.[90] It is undeniable that interactions with one of these *kawaii* – "cute," "adorable" – robots are impressive: its expressive eyes seem to seek contact and even visual exchange. A LOVOT costs about $3,000 plus about $100 a month to operate. In other words, they are not cheap – but, as an article on the Internet site Spirea remarks, "love is priceless."[91] Given its aim, LOVOT has the virtue of making things perfectly clear: AI robots are no longer intended simply to help us, they are supposed to become integral parts of what is nearest and dearest to us; they are to be full-fledged members of our households. They are thus insinuating themselves explicitly into what constitutes the most intimate framework of the relational field.

[88] Here is its presentation on the product's home page: "GROOVE X pursues [sic] to create a robot that touches your heart. Does this advancement benefit human beings?" https://groove-x.com/en/about/.

[89] "The reason why is that along with technological advances we have seen a growing gap between our plentiful lifestyles and our emotional engagement and sense of satisfaction. GrooveX has created LOVOT in order to help bridge that gap: https://groove-x.com/en/about/.

[90] https://lovot.life/en/technology/.

[91] See Gloaguen 2019, "Lovot, le petit robot à aimer."

CHAPTER 5

The Artificialist Fallacy

5.1 Will it still be possible to conceptualize what is possible?

5.1.1 *Roomba-ization of human lives*

The problem we are grappling with goes well beyond performance. The performance of relational robots leads to a reconfiguration of the way we see things and of the way we behave that reflects a double dynamics of adjustment. Several examples of this reconfiguration can be found in the cases we have already examined.

First, as we saw in the example of vocal driving assistance, given the difficulty the software had in recognizing voices, the proposed solution was not that the software should be reviewed and modified, but rather that women should learn to speak in more masculine voices – they should transform themselves to suit the machine. This observation might seem merely anecdotal if it were not part of an overall tendency that has been called the *Roomba-izing* of homes, the reorganization of living spaces to accommodate the vacuum cleaner's needs.

In "My Roomba Is Rambo: Intimate Home Appliances" (Sung et al. 2007), the authors show that even robots like the robotic vacuum cleaners Roomba, which are not designed to solicit emotional reactions on the part of their purchasers, can actually produce profound transformations and strong feelings in those who own and use them. Possession of a robot vacuum cleaner not only leads to a reorganization of the owners' living space, so as to allow the robot do its job under optimal conditions, but in some cases it induces a psychological attachment that can lead to the categorization of the machine as "somewhere between a pet and a home appliance" (151). These robots are generally seen as masculine, which is interesting given that their function as basic housecleaners is typically categorized as feminine. In any case, the reorganization of space appears inseparable from the reorganization of the mental categories associated with the home environment. The relation between living and artificial beings is blurred here, for a Roomba is viewed as a structuring or restructuring member of both the house and the household.[1]

To be sure, this idea of "Roomba-ization" might seem exaggerated, since automated vacuum cleaners have often been turned away from their cleaning function[2] and used for example as spirographs,[3] musical instruments, or amusing vehicles for cats. But this notion is probably more telling when it is applied to so-called smart cities, which are supposed to welcome tomorrow's technologies, driverless cars in particular. The term "smart," meaning "intelligent" in this context, can evoke both intelligence-gathering – spying (it is interesting to note that even a Roomba can transmit information about what it perceives!) – and the adaptation of humans to the machines rather than the reverse.

Human adaptation can also take the form of human imitation of a machine. Perhaps the most striking example comes from the Azuma Hikari company, which makes robots and holographs for Japanese "salary men," as we have seen. The company also sells costumes that allow young women to slip into the skin of the virtual companions, as it were, so as to promise a copy almost as satisfying as the original.[4]

[1] The expression used, "a valuable member of the house" (Sung et al. 2007, 156), highlights the importance of the Roomba and its value to the household.

[2] See for example "Hacking Roomba: Projects Repository," n.d.

[3] https://www.flickr.com/photos/49147885@N00/507132860.

[4] See for example an advertisement by LolitaDressesShop, offering cosplay costumes for Gatebox Holographic Virtual Robot Girlfriend Azuma Hikari: https://www.lolitadressesshop.com/gatebox-holographic-virtual-robot-girlfriend-azuma-hikari-cosplay-costumes-p-5398.html.

But a perhaps more concerning example lies in the everyday uses we make of various tools that require us to "think like the machine." The most troubling instances are probably found in educational software designed for young children, examples of which are far more common than "baby-sitter" robots, and for that reason undoubtedly warrant in-depth study. Here I shall focus on a few elements that I find essential to my own project.

The first example is an instance of artificial pedagogy, AI as an elementary school "teacher." The Lalilo program is used both in the United States and in France; it is advertised as a "free personalized" phonics and reading comprehension program for kindergarten through second grade, although it can be accessed by parents and children only if it is assigned by a child's teacher. Critics have pointed out that the feedback it provided users can be misleading (children may be expected to repeat exercises in which they have performed flawlessly, creating anxiety about what they have done wrong) or inadequate (users are told they have made errors without clear indications regarding the problem and the solution). At the second-grade level, the program introduces lessons on the complexity of language, for example on the use of words in literal and figurative senses; here it is easy to spot the limitations of the approach. In the French version of the program, given the sentence "He fell off his chair," the pupils are asked to categorize it as literal or figurative. Rather than showing how either answer might be appropriate depending on the context, the program accepts only one answer as correct.[5]

At the same second-grade level, in a section focused on reading comprehension, the software introduces a story about emoji designer Angela Guzman.[6] Originally from Colombia, she arrived in Florida as a child. Speaking no English, she used her graphic skills to make herself understood in a universe where no one spoke her language.

[5] I thank Adrien, a seven-year-old boy, for pointing out (a bit sarcastically, I must admit) these problems, among others, in the French version, where the sentence was "il est tombé de sa chaise."

[6] See https://angelaguzman.co/about.

This led to a whole mythology of a super-heroine creating emojis used as instruments of communication allowing people who did not share a common language to communicate with one another. This story, however coherent it may appear, obscures several important elements. First, the linguistic plasticity of children: it is likely that within a few short months Angela Guzman was already speaking at least some English. Second, the immigration patterns of the United States: it is highly probable that in Florida she was not the only Spanish speaker in her school environment. Finally, and most importantly, the true function of emojis: they are designed above all to add a touch of emotion to the impoverished forms of conversation fostered by the Internet.

As a marketing strategy, the Lalilo Company presents its software as highly interactive; it claims to offer a true interface between teacher and students. During the school closings resulting from the Covid epidemic, Lalilo thus looked like the best alternative, and was sometimes indeed the only alternative in a situation where teachers were doing their best but had not been adequately prepared to teach remotely.

The problem is that the software, especially when it is used at home rather than under a teacher's direction in class, leads children to treat it as an interlocutor to which they have to adapt – they see it as a partner in the educational process, and not as a tool that may not always work properly. Furthermore, the reading comprehension exercises – called "Story Telling," Angela Guzman's story being an example – may introduce an affective register, since the child user can readily identify with a child whose story is being told, and may be led thereby to adhere to a certain worldview in which emojis are virtually magical objects

Here we encounter one of the problems raised by the overall project of cybernetics and the cognitive sciences: how to "simulate and stimulate" human beings at the same time. As Jean-Pierre Dupuy points out, retracing the beginnings of cybernetics (of which the cognitive sciences are the heirs, in his view): rather than "thinking is computation," the formula "knowing is simulating" better sums up the spirit of the cognitive sciences (Dupuy 2000, 41). As I have already noted, Dupuy emphasizes that the term "simulation" has two different meanings: while in everyday language it means to "mimic" or to "pretend," in the sciences to simulate is to create a model that reproduces the workings of a system.

The problem is precisely that to simulate human behavior, one must reduce it to the workings of a system, at least in some cases, and thus impoverish it. Jean-Michel Besnier has described this well in his eponymous book *L'homme simplifié* (2012). And it is also problematic that in such cases of simplification what is human is absolutely not stimulated; quite the opposite: instead of stimulating by opening up perspectives, the approach based on modelizations is reductive and self-limiting. The NBIC convergence here points up very well, it seems to me, the tenor of the problem and its situation upstream in a necessarily limited horizon.

Moreover, we can see how these technologies are anchored in what was originally one of humanity's strong points: our social capabilities, and in particular our capacity to adjust to our circumstances; these capabilities have enabled us, up to now, to broaden our horizons and to build civilizations, often in the face of natural obstacles. But these capabilities are now turning against us, by allowing NBIC to limit and redirect them. The adjustment works in two ways. First, we reconfigure our environment so it can adapt to machines. This is the Roomba-ization of the external, as it were. Second, we reconfigure our way of thinking so it can adjust to machines: a Roomba-ization of the internal. Needless to say, this double Roomba-izing is in perfect harmony with anticipatory design and greatly facilitates it.

From Roomba to Lalilo, it seems legitimate to query the design of such machines and to ask why we turn them into Rambos – which is of course better than making them Terminators, but maybe not all that much better, if we think about their architecture and their inclusion at the heart of our *oikos*. The etymological sense of the term *oikos* embraces not only the house but the household: the living beings that inhabit it and their relationships. I am suggesting that the inclusion of machines like Roomba and programs like Lalilo in our *oikos* is capable of reconfiguring simultaneously how we think, how we interact, and what we view as the horizon of possibilities.

5.1.2 *An experiential and conceptual reconfiguration?*

As we have just seen, the way in which we conceptualize human relations, including empathic ones, is being reconfigured by the model of artificial intelligence. With respect to drones and the reconfiguration of empathic relations at work owing to the use of drone technologies, let us recall that the model Mark Coeckelbergh used to characterize the new modalities of human relations was that of hacking. Hacking implies breaking into a system, of course – this is the image Coeckelberg was after – but it is also a form of control, of *maestria*, practiced by computer nerds: those who, to borrow Sherry Turkle's description, love computers for their own sake and replace human relationships with connexions with machines (or who maintain relations with other humans by way of machines, we might add) and who share with individuals on the autistic spectrum a focus on things rather than on persons.[7] One of Turkle's hypotheses is that these human-machine relations are presented as frictionless and risk-free. It is probably no accident that a stereotypical hacker strikes David Levy as the ideal candidate for love and sex with a robot. The technicist logic would be anchored, so to speak, in a reformatting of empathy.

The problem here, as Sherry Turkle emphasizes, is that we have passed, semantically, from *healthcare* robots to *care* robots, and the term *care* itself has come to designate a transformed and impoverished vision of care (Turkle 2011, 106). Turkle recounts an incident in

[7] This might be one of the reasons why machines are so present in the lives of autistic individuals. Some authors, such as Kathleen Richardson, go so far as to propose an interpretation of this phenomenon in terms of gender (masculine), relying on the (contested) work of Simon Baron-Cohen, who interprets autism as hypermasculinity (see *The Essential Difference: Men, Women and the Extreme Male Brain* [2003]). While I do not accept that interpretation, it does seem to me that the question of how the focalization on machines risks handicapping – or at least hampering the development of – certain empathic functions deserves to be addressed. See Richardson 2018, *Challenging Sociality*.

which, having been invited to a symposium on “Caring Machines: Artificial Intelligence in Eldercare,” she questioned the participants about the meaning of the event’s title: were we now to consider that machines could be made for “care”? Irritated by her question, some participants tried to reassure her: it is not because a machine *takes care of* us that it *cares about* us. As Turkle explains, these participants viewed care as a *behavior* and not as a *feeling*. And in a later work, she adds that care has come to be treated as a *function* (2015, 52).

For Turkle, this conceptual and linguistic slippage is far from trivial, for it implies not only treating machines as humans but also treating humans as machines (2015, 345-47). As I see it, this reduction of care to a behavior or a function is facilitated by practical procedures designed to mechanize care, to reduce it to a series of quantifiable microtasks that can be completed within a given time frame, imposing new pressures, for example, on health care providers and staff members in nursing homes. The idea that care can be viewed as a set of automatic behaviors is easier to advance if, even when the care is being delivered by human beings, these caregivers are asked to behave like machines, with their actions plotted out and timed to the millisecond. As the Québecoise philosopher Monique Lanoix writes in “Un amour de robot: Robot émotionnel et travail d’aidant” (2019, 259), in the wake of Giorgio Agamben’s work:

> Since commercialization acts on the way caregivers’ work is structured, that work more closely resembles productive work, even if it does not produce surplus value. Subjected to such a work regimen, the worker becomes a tool rather than a working being. The worker is the tool by means of the work dictated by the institution or the long-term care home is carried out.

In this phenomenon with two mutually reinforcing drivers, care is devalued and dehumanized on the one hand, while on the other it is instrumentalized (in the literal sense) and artificialized. The transformation of care also operates in the essential fields of human relations. As Turkle points out (2015, 52) “[i]t is natural for words to change their meaning over time and with new circumstances. *Intelligence* and *affective* have changed their meaning to accommodate what machines can do. But now the words *caring*, *friend*, *companionship*, and *conversation*?"

5.1.3 *Shortfalls of the dilemma approach*

The surreptitious transformation in the meaning of words, occurring in both private and public spaces, comes about in the terms of dilemmas presented as inevitable. We can recall the utterly classic dilemma of the drone pilot, Jay, caught between telling his son the truth so as to have an authentic relationship with him and protecting him from the intolerable reality; or again, the problem of the trolley that serves as the model and matrix of the Moral Machine. As we have seen again and again, the dilemma seems to be a privileged frame of reference in the development of AI.

But there is more. As Turkle has reminded us, discourse about artificial intelligence tends to take an either/or form, which may remind us of the Thatcher-era formulation: "There is no alternative" (TINA). "Do you want your parents and grandparents cared for by robots, or would you rather they not be cared for at all?" (Turkle 2011, 289). Turkle argues, however, that this binary way of thinking is not the only possibility. One can respond, as fifth-grade children in one school did: "But don't we have people to do these jobs?"[8] Worse still, the either/or way of presenting a problem opens the door to a different approach in which we move from the (false) dilemma of "robots or nothing," from "better than nothing," to "better than anything." Roboticized artificial intelligence seems to be the solution to all our problems, or at least to any problem that can be viewed as a matter of function or calculation: managing vehicular traffic, selecting a target, changing a diaper, having sexual relations, and so on. Following a different path, we are intersecting here with what Meredith Broussard (2018) has called technochauvinism. But this other path is significant in its own right, because it takes the form of a dilemma in its very structure.

[8] Turkle cites Carol Gilligan in a note (2011, 344, n. 23); regrettably, she does not fully explore the conjunction between their two approaches.

A dilemma posed in the either/or mode does not offer a real choice; it is a way of saying that one is having trouble finding an alternative, or implying that one has found the best solution possible. The "all or nothing" formulation positions the "alternative" as unacceptable – do you really want your grandparents not to be cared for at all?

In fact, this use of the dilemma format hides an impoverished dilemma, etic rather than ethical in nature: the idea of a non-choice with moral residues, as it were. It is a version, as we have seen, of the neoliberal TINA, "there is no alternative," and in my view it is inseparable from a conception of the world in which social Darwinism dominates and takes on the features of technochauvinism: technology is our only path forward, our only path to salvation, the only solution even to the problems that technology itself has created.[9]

Not only is the binary mode not the only possible mode for conceptualizing the world, not only can we, like Amy in Carol Gilligan's discussion of the Heinz dilemma, find a "third way," but we must also recall that Gilligan's analysis brings to light the way in which the binary way of thinking contributes to the impoverishment of human relations, and even to their denegation; it institutes and reinforces a system of domination through the confrontation of two solutions in which there is necessarily a loser or a sacrificial victim and a winner. One of the strengths of binary thinking, paradoxically, is precisely its poverty: it serves as a shortcut allowing us to label and categorize things quickly. This is sometimes practical, to be sure; but it also makes it possible to close down alternative modes of thinking.

One of the difficulties we must grapple with today is that, in the name of the "trust" we are supposed to place in AI, the recent scandals in all sectors (justice, health, education, and so on[10]) that have

[9] As Turkle notes, this view is attested by the creation of software intended to develop empathy, to counter the lowering of empathy tied to the use of digital media. She compares digital solutions to the cavalry, a force that can always be called in to come to the rescue (2015, 338).

[10] See for example Umoja 2018, *Algorithms of Oppression: How Search Engines Reinforce Racism*, and Pasquale 2015, *The Black Box Society: The Secret Algorithms*

hinged on problems related to gender, race, or class biases have been treated in the "garbage in, garbage out" mode: that is, with the idea that the solution must be either to modify the data used to make them more representative, or to change the make-up of the AI development teams so that they would be more inclusive (a situation that Kate Crawford has summed up as "Artificial Intelligence's White Guy Problem" [2018] and that seems also to be an issue in cybergenetics, as we have seen), or both. While there is clearly a great to be done in both directions, it nevertheless seems to me that this way of addressing the question risks giving the impression that the problem belongs to a particular moment or conjuncture. This may be partly the case, but we must take care to avoid letting the conjunctural elements of the problem conceal its structural dimension.

I submit that the combination of the arguments developed so far in this book makes it clear that the problem raised is not only conjunctural but structural. I shall now endeavor to show how the structural problem, which I propose to call the artificialist fallacy, has come about. To this end, two detours are required: the first through the concept of the *Apparatgeist* developed by James Katz and Mark Aakhus, and the second by way of the naturalistic fallacy, a classic topic in moral philosophy as developed by G. E. Moore.

5.2 From the Apparatgeist to the forms of life

5.2.1 *Apparatgeist*

The explanation most often proposed for the biases that are seemingly inherent in AI is that they are nothing but reflections of our society. In sum, AI is said to offer us a mirror showing us what we are – or at least mirroring those who supply AI with data – and, even if this reflection is displeasing, it is nevertheless a factual representation, an accurate description.

That Control Money and Information. On racial bias, see O'Neil 2016, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*.

However, as James Katz and Mark Aakus have shown with respect to mobile technology, the relation between reality and technology is decidedly more subtle than a simple mirror effect.[11] To give an account of that complexity, they coined the term *Apparatgeist*, echoing *Zeitgeist*, the spirit of the times. The *Apparatgeist*, or the "spirit of the machine," makes it possible to account for the transcendent dimension of certain technological objects – in the authors' case, those used in mobile communications. The term *Apparat* designates a machine in both its technical and social dimensions, with stress on the idea that technology is socially constructed.

The *Apparatgeist* brings to light not technological determinism but rather the complex relation we have with the technologies that leave their stamp on our ways of being and living, while attesting to a societal imprint and to our powerful desire for "perpetual contact," to borrow the title of Katz and Aakhus's earlier collective volume (2002). The desire for perpetual contact does not originate in technology, but it is stimulated and even modified by technology. It is this desire and this communicational logic that come to model the link, real and imaginary, that connects individuals and societies with AI machines. The term *Geist*, or spirit, is used here in the classic Hegelian sense of a consciousness proper to each historical epoch, a consciousness that develops and then moves to a new stage in a dialectical process.

The *Apparatgeist* designates a sort of dialectical movement between the *Apparat* and the *Geist* in which society models the apparatus while the apparatus, in the same process, models societies and individuals. In reality, AI machines impose a certain number of structural constraints and limitations with which individuals and societies come to terms, sometimes by adapting these machines to unanticipated uses. In other words, technology is not fatalist in the sense that everything is played out in advance, but it is determinist in the sense that

[11] See Katz and Aakhus 2002, "Conclusion: Making Meaning of Mobiles – A Theory of Apparatgeist." In what follows I shall be following and refining the major threads of my 2021 article "The Artificialist Fallacy."

it eliminates a certain number of possibilities by setting up guard rails and sweeping away obstacles.

As the relation between human beings and machines acquires a form of intimacy unprecedented in the history of humanity, we are challenged to figure out how to characterize our becoming-machines, or more precisely these “machines that become us.”[12] But the *Apparatgeist* goes even further by embodying a “social sense” in and by way of machines: machines take on meaning and can go so far as to replace certain types of relations or even affective and cognitive processes – a phenomenon that converges with and reinforces the analyses proposed earlier in this book. The notion of *Apparatgeist* has the advantage of highlighting the dynamic dimension of human-machine relations. It also has a disadvantage, though: it risks implying that machines can have minds of their own and crediting them with agency.

5.2.2 *From the* Apparatgeist *to forms of life*

The theory of the *Apparatgeist* expresses, following a different path, an idea quite close to what Langdon Winner, using Wittgensteinian vocabulary, illustrates in his book *The Whale and the Reactor: A Search for Limits in an Age of High Technology* (1986) through the concept of “forms of life.” Having already evoked this notion of “forms of life” in critiquing “patterns of life” in the context of drones and cybergenetics, I return to it now from another angle.

Like Katz and Aakhus, Winner does not consider it appropriate to analyze the links between technologies and society in deterministic

[12] Here I am echoing the title of another collective volume: Katz, ed., 2003, *Machines That Become Us: The Social Context of Personal Communication*. In this title, the most important term is neither “machines” nor “us” but the process of “becoming.” This latter term can in fact be understood in at least three different senses. First, in the sense that certain technologies become instruments for broadening the possibilities of communication with others. Next, in the sense that technology can become part of ourselves, a form of prosthesis or extension, an appendix, and thus a form of representation in the sense that the part can represent the whole. Last but not least, in the sense that we become agents vis-à-vis these technologies, which are modified through contact with their users and are thus in a way “customized” by us.

terms (1986, 9). He prefers the term “second nature” or “form of life” to explain how these systems become part of what we are (or even part of our humanity – something that seems to me more open to question).[13] He hypothesizes that certain technological innovations modify the human condition, even in its very fabrication. The fundamental question to be raised about technology is thus no longer “how do things work?” but rather “what sort of world are we making?” It is not so much a matter of stating that things shape the world but rather that they play an important role in certain processes of transformation that are psychological as well as social and political. For example, the conception of a ship may be such that the captain is obliged to shout in order to make himself heard by the crew; from the outset, this conception implies a certain political structure, intrinsically hierarchical; or a bridge may be built with so little clearance that buses cannot use it, which prevents people who cannot afford to own cars from ready access to certain parts of town. As Susan Leigh Star points out in “The Ethnography of Infrastructure” (1991), infrastructure in the digital world is even less visible than it is elsewhere, but it can be conceived, like the infrastructure in our two examples, in a way that embeds in it certain implicit social and political structures.

Winner goes even further when he indicates that these forms of life function like a second constitution, parallel to but also sometimes superimposed on the socio-political constitution (1986, 55). For this reason, he emphasizes, it is all the more crucial that we assume our responsibilities in the process of fabrication (of technological objects and the world): what is most important, he contends, is not studying the impacts of technological change but evaluating the social infrastructures that certain technologies create and into which our activities are folded (18, 55).

[13] “As they become woven into the texture of everyday existence, the devices, techniques, and systems we adopt shed their tool-like qualities to become part of our very humanity” (1986, 12).

This notion of "evaluation" strikes me as particularly interesting, for it is not a matter of carrying out a study of the assessment type, as found in the idea of "risk assessment." Nor is it a matter of carrying out a "thing turn" like the one evoked earlier in the context of nanotechnologies. To explore the notion of evaluation, it seems fruitful to bring Langdon's reflections together with a distinction made by the American pragmatist John Dewey in *Theory of Valuation* (1939). According to Dewey, the notion of value can be understood in at least two senses: on the one hand, appreciation, or "valuation," and on the other, evaluating, or pricing. This second sense is the most commonly encountered conception of value: it consists in evaluating a thing in the sense in which the thing is given a price or an exchange value. Valuation is rather the phenomenon of questioning ourselves about what matters to us, what is dear to us. Dewey's interpretation emphasizes the importance of not limiting the approach to values to an arithmetical conception linked to social acceptability. For all these reasons, it seems important to me to take Winner seriously when he speaks of evaluating infrastructures by raising the question – an ethical and political question – of what matters to us (1986, 55).

The notion of forms of life must of course also be interpreted from a more creative vantage point that allows a certain room for freedom, or even for deviation, as we have already seen. In the case of certain applications of artificial intelligence, as with surveillance drones, for example, the reliance on "patterns of life" is in effect a negation of the creative dimension of "forms of life." The notion of forms of life thus appears as another possibility of articulation, one that makes it possible to reject reduction and to gesture instead toward creation. Here again, it proves fruitful to invoke the ethics and politics of care, for these allow us to open up the field of alternatives and to raise the question of what matters to us.

5.3 Toward an artificialist fallacy?

Given that the relations between technologies and politics threaten to accredit certain *patterns* to the detriment of *forms*, I propose to take up this point from a different angle by emphasizing the danger of succumbing to what I call the *artificialist fallacy*. This fallacy is based on two theoretical sources: the *naturalistic fallacy*, which is a classic conception in moral philosophy developed by G. E. Moore, and "naturalization" in the sense Pierre Bourdieu gives the term.

5.3.1 *The naturalistic fallacy*

For a better grasp of the concept of the artificialist fallacy against the backdrop of the naturalistic fallacy, we need to look closely at G. E. Moore's *Principia Ethica* ([1903] 1993).

What is commonly called the naturalistic fallacy in moral philosophy stems from a critique of the reductionist assertion that moral norms can be broken down into "facts" and reduced to natural phenomena. More specifically, in Moore's view, it is unacceptable to confuse the property of goodness ("the predicate G") with a natural property.

To put it in more contemporary terms, Moore denounced the equation "it is natural" = "it is good," an equation still in common use today. Still, we must note that Moore meant something very precise by "natural" here. A natural property, as he put it, is a "property with which it is the business of the natural sciences or Psychology to deal, or which can be *completely* defined in terms of such" ([1903] 1993, preface to the 2nd edition, 13; emphasis added).

This is precisely why Moore criticized naturalism, which he defined as

> a particular method of approaching ethics – a method which, strictly understood, is inconsistent with the possibility of any Ethics whatsoever. This method consists in substituting for "good" some one property of a natural object or of a collection of natural objects; and in thus replacing Ethics by some one of the natural sciences. (Moore [1903] 1993, chapter 2, section 26, 92)

The principal flaw in naturalism, then, is that it takes the good to be a natural property that can consequently be studied by the natural sciences and – this point is essential – studied "completely" – by them. In essence, then, ethics, or moral philosophy, becomes useless, since it can be fully replaced by the natural sciences. Such a stance is not only false, in Moore's view, but also dangerous, because by reducing ethical reflection to scientific or technical reflection, one is not in a position to propose a valid argument – starting from false premises, one cannot arrive at accurate conclusions – and, what is worse, one risks accepting false and immoral propositions as ethical.

It seems to me that Moore's conclusions on this point are closely related to the problems I have been addressing in this book, for what he rejects is a form of reductionism that closely resembles the technochauvinism denounced by Meredith Broussard (2018). To pursue this observation further, it will be useful to look more closely at the particular types of naturalism to which Moore objected. On the one hand, he pointed to the hedonism of John Stuart Mill, who focused on the quality of the pleasures that are deemed superior as standards of value.[14] On the other hand, he targeted Herbert Spencer and his social evolutionism, which sought to extend Darwin's theory to societies, and is one of the crucibles of individualist thinking[15] – which Broussard sees as going hand in hand with technochauvinism; it is also aligned, as we have seen, with the neoliberal logic of "there is no alternative" (TINA). On the digital horizon, that logic is also embodied in the version of the so-called Gabor's law according to which everything that is technically possible will be achieved sooner or later, as if it were a question of a necessary and unstoppable evolution.

The rejection of the naturalistic fallacy is thus a rejection of the stance that reduces morality and ethics to the province of the natural sciences. At first glance, the connection with my own argument may not be obvious, although we have seen that cybergenetics includes a prescriptive aspect that can have a moral dimension. In my view, the connection goes further still; to finish grounding this claim, I propose to turn now to Pierre Bourdieu's work.

[14] Moore 1993 (1903), especially chapter 3, sections 38-48, 115-132.
[15] Moore 1993 (1903), especially chapter 2, sections 31-35, 100-110.

5.3.2 *Naturalization according to Bourdieu*

The last theoretical line on the artificialist fallacy surfaces in the concept of naturalization developed by Pierre Bourdieu, in particular in his book *Masculine Domination* ([1998] 2001). By "naturalization," Bourdieu meant the embedding of relations of domination (especially gender relations) into our habits of thought, through their assimilation to a natural phenomenon understood as both normal and biological. This assimilation is accomplished through a movement of essentialization coupled with dehistoricization.

Where Moore's naturalistic fallacy conflates nature with the good, the naturalization Bourdieu denounced conflates a natural phenomenon with a social phenomenon. This merging is facilitated – as Marx had already shown in a different way in *The Poverty of Philosophy* ([1847] 1995), or Nietzsche in *The Genealogy of Morals* ([1887] 2003) – by the failure to recall its origin and genesis, which has allowed it to acquire a form accepted if not as eternal then at least as immutable. The methodological error can thus become an ideological imperative and a moral failing.

The ideological imperative is manifested most notably in the realms of gender and education. We see this for example when gaps between "masculine" and "feminine" practices are presumed to be anchored in invariants that are all the more easily naturalized in that they are directly tied to the body. For example, naturalization attributes certain effects of sex – which is biological – to gender – which is socially constructed – and thus confirms masculine domination. Similarly, in the realm of education, the tendency to naturalize gifts, abilities, and talents leads to confirmation of socially constructed biases.

5.3.3 *The artificialist fallacy*

My hypothesis is that naturalization in Bourdieu's sense serves in various respects as a matrix for a process of artificialization that carries the naturalistic fallacy a step further. The artificialist fallacy thus relies on a fallacy that can be called naturalistic in Moore's sense and naturalization in Bourdieu's sense. It extends it to a form of artificialization operative in the new technological and digital world. The term refers to the embedding of habitual structural constraints in machines and programs (even if deep learning might seem to ward off a form of rigidification), while giving the false impression that, because they are artificial, these machines and programs are necessarily neutral, and that because they are neutral, they are necessarily impartial and thus morally good. This is a stratified process, as it were, in which artificialization is layered on top of naturalization and serves to consolidate its structure.

For example, as we have seen in the form of naturalization denounced by Bourdieu, the ideological imperative constructs a conception of human nature that conflates the natural with the social instead of articulating them together – whereas articulating them together is precisely what is accomplished by the concept of forms of life. I suggest, then, that the concept of forms of life constitutes in certain respects an antidote to naturalization, and that, in the form of artificialization I am denouncing, the ideological imperative constructs a conception that would *conflate* the social (which is already conflated with the natural) and the artificial – at the point where the *Apparatgeist*, for its part, *articulates* the *Apparat* with the *Geist*. In formulaic terms, one might say that artificialization risks positing the equation *social* = *natural* = *artificial* = *good*, while the naturalistic fallacy posits that *natural* = *good*, and naturalization posits that *social* = *natural*.

My hypothesis takes this slippage to be allowed, and even supported, by yet another conflation, one that occurs in two stages. First, the conflation between "neutral" and "artificial"; second, the conflation between "neutral," "artificial," and "impartial." One might even add a third stage, the conflation between "impartial" and "moral," but this one does not stem from artificialization; it is rather a long-standing problem intrinsic to the field of moral philosophy.

Two analyses may lend support to my hypothesis. First, as we have seen, the use of moral dilemmas – such as the trolley problem in "moral machines" – brings to light the predominance of a patriarchal conception of ethics, and even the anchoring of ethics in such a conception; second, the claim to neutrality can be examined if we return to the question of gender biases.

In the case of gender bias, the presumed neutrality of AI may indeed mask a form of neutralization. Converging work,[16] including illuminating research by Alison Adam,[17] shows in fact that, while the early development of computers was led by teams where men and women worked together, the women have become invisible histories of this work: they have neither been cited by name, for example, nor included in official photographs,[18] whereas they are quite visible in photos of the teams at work.[19] This phenomenon is all the more ironic in that one often hears complaints today that there are not enough women working in the field artificial intelligence (and in the most highly valued scientific and technological fields in general), and the femininization of AI is presented as a challenge. We tend to forget that this is part of a historical phenomenon whereby the more a trade or profession is viewed as valuable and "objective," the more it appears to be "men's business," suited for "real men." The historical process of rendering women invisible has gone hand in hand with

[16] For a well-informed approach to the omission of women from the history of computer science, see Evans 2018, *Broad Band: The Untold Story of the Women Who Made the Internet*.

[17] See especially Adam 1998, *Artificial Knowing: Gender and the Thinking Machine*, and Adam 2005, *Gender, Ethics and Information Technology*.

[18] "In subsequent retellings, the women were skipped over repeatedly. In some historical images, the ENIAC Six are captioned as models, if pictured at all. 'I wasn't photogenic,' said Betty Snyder. 'I wasn't included on any of the pictures of the entire stupid thing.' [The Electronic Numerical Integrator and Computer, or ENIAC, was developed in the United States during the Second World War; it was the first large-scale programmable electronic computer.] When the army used a War Department publicity shot of the ENIAC for a recruitment ad, they cropped out the three women in the picture entirely. The War Department's own press releases about the ENIAC cited a vague, genderless 'group of experts' responsible for the machine's operation, and mention by name only John Mauchly, J. Presper Eckert, and Herman Goldstine" (Evans 2018, 51).

[19] See for example the site of the ENIAC Programmers Project: ttp://eniacprogrammers.org/.

a gradual masculinization: as computer science and artificial intelligence were taking on importance as fields of study and work, they could no longer be "women's business." As Meredith Broussard points out (2018, 83), this dynamics is inseparable from the way in which the myth of (male) genius was constructed, and more simply the myth of (male) research excellence in the "hard" sciences and technologies, including physics and computer science. And she reminds us that, even today, women and persons of color are rare among those viewed as geniuses in those disciplines.[20]

I suggest that the problem here is not only that the history of AI has been neutralized, in the sense that the role of women has been edited out, and that it has now become a challenge for them to penetrate into the "hard" scientific and technological environments. But I would also like to suggest how significant this neutralization is. It reveals yet again, in another voice, that AI is dominated by a patriarchal worldview, one that assigns men the important and "objective" tasks – or rewrites history to give the impression that this has always been the case.

However, as Alison Adam (1998) notes, this "view from nowhere," which we have already encountered with NBIC and which is shared by AI, or so we are led to think, this alleged neutrality of AI – the seeming absence of gender – in reality hides men in their thirties and forties: the "white guy problem" again. It seems to me that this process is aligned with the one described by Bourdieu: these hidden biases and the gradual dissimulation that is a form of rewriting history

[20] Broussard also emphasizes that the scientific culture of these disciplines is grounded in characteristics that are generally viewed as "masculine" (objectivity, method, lack of empathy, and so on), a fact that makes it harder for women to believe that they can find their place in such environments, which are said to be "discouraging" for women. Citing a 2017 article by Hannah Natanson published in the *Harvard Crimson*, Broussard adds that in the sciences at Harvard women are encouraged more often than their male counterparts to enroll in courses reputed to be "easy." In addition, males decidedly predominate among the teaching faculty: in 2017, there was not a single tenured female professor in the Harvard mathematics department; there had been one, who left, and three others were reportedly offered positions but declined to accept.

are instances of what Bourdieu calls naturalization. Moreover, the masculinization at issue here is inseparable from the attribution to AI of a quality that is precisely presumed to be masculine: its "scientific" objectivity, which is one form of its neutrality. This attribution is the first stage in the process.

The second stage relies on this supposed objectivity and neutrality to pass from the field of "facts," or phenomena taken as facts, to that of norms: from technical objectivity to ethical and political impartiality. The "view from nowhere" cited by Adam makes it eminently possible to connect these two models, for it proceeds not only from a mathematical view of the world but also from a vocabulary used in philosophy and moral psychology to designate impartiality, or, for example in Lawrence Kohlberg's approach, a point of view where all points of view meet and which is part of a sort of cosmic force. And let us recall that it was in the name of impartiality that Michihito Matsuda proposed an artificially intelligent robot as a mayoral candidate in a Japanese city in 2018. An AI robot is presumed to be more impartial and thus fairer than a flesh-and-blood all too human mayor.

In an entirely different style, it was also in the name of a scientific, objective, and impartial approach to morality that cognitive psychologists and neuroscientists relied on the trolley problem in an effort to propose naturalizing morality, not in Bourdieu's sense (or in that of classical philosophy), but in the sense of the cognitive sciences, which seek to bring into the fold of the natural sciences something that does not initially proceed from them. The resulting social experiment was accompanied by the rising trend toward what has been called "experimental philosophy" (XPhi), preoccupied with approaching philosophy without preconceptions and with integrating into it elements from the neurosciences and experimental psychology.[21]

[21] This does not seem to me to be objectionable in itself, quite the contrary: in *Sommes-nous naturellement moraux?* (2011), I explained why I am in favor of a moderate naturalization of moral theory. What I am discussing and critiquing here is a systemic effect.

This other form of naturalization seems to me just as important and significant for understanding the process at work here as Bourdieu's version; the two are indeed complementary. As we have seen, the use of the trolley problem originated in John Mikhail's reframing of the Rawlsian hypothesis on the sense of justice understood in particular as a cognitive capability inspired by Chomsky. But the rise in popularity of the trolley problem quickly obscured and then obliterated the origins of that dilemma, which had been invented and used by Philippa Foot and then Judith Jarvis Thomson as one of the paradigms that made it possible to defend abortion from a feminist perspective. But as we have seen, the "mechanical" use of the trolley problem embodies, in a sort of historical irony, an anchoring of the patriarchal paradigm – which had been turned against itself, as it were, by Foot and Thomson. Here we find, then, a reversal inseparable from a "forgetting" of the original intention, which would be ironic in a different context.

Moreover, as we have seen, from a meta-ethical standpoint, by endowing "autonomous" machines with software based on the conception of a moral theory that relies on dilemmas, this type of conception is validated and neutralized by technology, as it were, in a vicious circle that makes it possible to assert that the solution was necessarily good because the machine had chosen it – which amounts of course to denying the role of justification in the moral field by using technology as a form of argument from a position of authority.

The problem that arises here is, in a certain sense – and only in a certain sense – similar to a process of degradation tied to automatization, which ends up with a loss of intellectual and manual competence, a loss of know-how, that was analyzed along neo-Marxist lines in the 1970s.[22] This process has also been assessed in the AI field under the term "deskilling," even if that loss can of course sometimes be the loss of a skill that was not necessarily valorized or valorizing

[22] See for example Braverman 1974, *Labor and Monopoly Capital: The Degradation of Work in the Twentieth Century*. This hypothesis has been the object of lively discussions, as reflected in Friedman 1977, *Industry and Labour: Class Struggle at Work and Monopoly Capitalism*.

owing to its repetitive nature. But this is not necessarily the case: deskilling can also be understood as the loss of the ability to make pedagogical choices and of the capacity for reflection, offset by the existence of preprogrammed "packages," as Michael Apple points out in "Computers and the Deskilling of Teaching" (1994) or as a mechanization of care work and the loss of attention to certain symptoms, or the automatization of diagnoses, for example, in the field of nursing.[23]

By transferring to a machine the exercise (in the literal sense) of moral choice, we run the risk of no longer exercising our own moral sense. This brings up the question of the degradation or even the loss of our competence for want of exercise – all the more so in that it is an acquired disposition, as thematized for example in the Aristotelian tradition. This idea has been advanced by Shannon Vallor in an article titled "Moral Deskilling and Upskilling in a New Machine Age"; in her conclusion, she argues that "the imagination of the moral self must lead the way for the 'technologies of the self,'" as characterized by Michel Foucault 1988, just as much as in the sense of the contemporary technologies that are the medium of our current relation to the world. As she asserts in "Moral Deskilling and Upskilling in a New Machine Age: Reflections on the Ambiguous Future of Character" (2015, 122 and note 6):

> My claim is that we retain the ability to imagine and develop new and alternative forms of these relations, even as our present powers of moral imagination continue to be shaped by our existing technology relations. The future trajectory of human moral abilities is technologically conditioned, but not technologically determined.

In an article titled "L'exercice moral est-il assimilable à une technique?" (2007), Laurent Jaffro discusses the "technologies of the self" in an analysis that helps refine this argument. He writes that

> the practices of the self cannot be equated with technologies, but rather with dispositions and habits or, by extension, with technologies whose

[23] See for example Rinard 1996, "Technology, Deskilling and Nurses: The Impact of the Technologically Changing Environment."

> essential and natural use is to produce such dispositions and habits. But then the practices of the self do not stem from a second type of technologies, but from something entirely different – which I call "gymnastics."

Here Jaffro is comparing our moral capacity to a form of jogging that could lead to a *good* transformation of the self – and not to just any type of transformation: the aim is an improvement based on knowledge. In this way he emphasizes two important features of this gymnastics. First, jogging is an *exercise*; second, the exercise in question is that of a will supported by reason, the latter itself being guided by a moral conception. Jaffro thus concludes: "Gymnastics cannot be reduced to a technology."

It is this conception of moral capacity as an exercise repeated for the purpose of improvement that I should like to mobilize here. The notion of thought as a gymnastic exercise thus goes further than the idea of an acquired disposition, and in any case it is quite different from the moral routines that form the basis of automated morality. Jaffro's conception also presupposes regularity, care, and attention (sometimes more or less unconscious attention) to the "little things" as well as to the process, and all these necessarily disappear when one stops engaging in the activity. Here we rediscover John Dewey's idea that ethics is not only a matter of knowledge but also of "know-how." It is actually the very capacity for moral experience that is damaged by the AI experiment.

The risk of relegating or even relinquishing moral decisions to AI would then be a loss of a moral competence owing to the failure to exercise that competence. This possibility can be related, in a way, to the classic issue of "wild children," children who lacked human parenting and who were thus thought to exhibit the opposition between nature and culture. In the well-known French case of Victor, a "wild boy" from the Aveyron region, the failure of the efforts made to educate him was presumed to demonstrate an unbridgeable gap.[24]

[24] For an account of this case, see Lane 1976, *The Wild Boy of Aveyron*. A hypothesis defended by Uta Frith, a leading specialist in autism, is that Victor may have been a non-verbal autistic child. See Frith 1989, *Autism: Explaining the Enigma*.

More recently, however, it has been understood that such children, deprived of human care and social interactions, may well fail to develop certain abilities, but that they are capable of acquiring them at least in part when they are placed in conditions adapted to their needs; this was the case for some unfortunate children who had been left in Romanian orphanages where care and social interactions were lacking.[25] The problem, nevertheless, is precisely that there would no longer be a society or a community in which these children would be able to develop moral competencies if our societies themselves were to be fundamentally modified by artificial intelligence.

This line of thinking brings us back to the internal/external dichotomy of "Roomba-ization," for the environment is not necessarily (or not solely) material; it can also be social and interpersonal. The "wild child" example thus rejoins the argument according to which, while we are endowed at birth with certain competencies, these competencies require an appropriate environment to be developed; in other words, they develop differently depending on the environment. The capacity for language, for instance, requires a human environment to be developed, among other things, and children learn to speak different languages according to the linguistic environment in which they are placed. The process is presumably similar for social and moral capacities. Going a step farther, the question can also be formulated from an evolutionist perspective; we can ask whether, after a certain number of generations, a given type of competence would continue to be transmitted or whether it would risk disappearing in the course of evolution. The problem, then, would be not only the loss of a skill but actually of the capacity to exercise that skill. The phenomenon is all the more problematic if we consider it from the perspective of new generations immersed from birth in the digital world and artificial intelligence. Brought up by artificial babysitters, educated with educational devices programmed by AI,

[25] See especially Michael Rutter and the English and Romanian Adoptees (ERA) Study Team 1998, "Developmental Catch-up, and Deficit, Following Adoption after Severe Global Early Privation."

transported in autonomous vehicles, having their first sexual experiences with robots, these generations would be engaging with modified conceptions of love, care, relationships, and so on, without having known anything different.

Of course, I have just set forth a somewhat – but perhaps not entirely – hypothetical prospect; still, it has the advantage of highlighting the problem by situating it both within a developmental perspective (from a psychological or descriptive standpoint) and within the perspective of our responsibility to future generations (from a normative standpoint). The problem that is found on the horizon of the artificialist fallacy is that, if we do not denounce it, the ethical field will be not only reduced but also modified in a way that may become irreversible at some point. As was the case with the naturalistic fallacy, the challenge is thus to denounce a form of reductionism that is based in many respects on naturalization, and that is reinforced by certain NBIC technologies, as we have seen, for example, with cybergenetics.

CONCLUSION

Back to the Enlightenment?

The leading threads of this analysis have not only brought to light common features in NBIC, they have revealed a veritable "technician system," where each part support the others and the whole, based primarily on three key elements.

First, NBIC manifests blurred boundaries and a resulting definitional complexity. This confusion of categories and dichotomies is evident in the three examples of NBIC we have examined in this book, even if there is inevitably some play in the categories involved. The following table recapitulates this blurring.

Blurred boundaries and categories in NBIC

General characteristic	Care	Nanoethics	Cyber-genetics	Artificial Intelligence
Thinking in terms of security or impacts does not suffice.	Boundary between the political and the moral	Boundary between the designer and the object	Dataization, commodification	**Drones**: Combination of passive surveillance passive and active execution **Relational AIs**: Confusion between mobility and agentivity
The "view from nowhere" does not suffice.	Boundary between reason and feeling	Boundary between the living and the artificial	Factish	**Drones**: Target becoming an insect or an element of the landscape; relation between the algorithmic and the biological **Relational AIs**: "Alive enough" (Turkle)

Porosity between the inside and the outside	Boundary between the private and the public	Boundary between the interior and the exterior	Sharing, participation	**Drones**: Near/far (patterns of life, cubicle warriors), surveillance **Relational AIs**: Data transmission (surveillance); belief that a robot loves us because we transfer our emotions onto it (media equation)

Second, all components of NBIC implement anticipatory design and self-fulfilling prophecies. And, third, all illustrate the inadequacy of approaching ethics through the logic of risk and control. The table below recalls several examples:

Some examples of the logic of anticipatory design and self-fulfilling prophecy and the logic of risk and control

	Anticipatory Design and Self-Fulfilling Prophecies	**Logic of Risk and Control**
Nanos	– Influence of cryogenics (Robert Ettinger) – Nano mythology (Eric Drexler and the "self-replicators")	– "Accompaniment" of the nanos – The ELSI approach – Social acceptability – Responsible innovation – Safety by design
Cyber-genetics	– Genethos – Analysis of test results by white supremacists	– The ELSI approach and Nancy Wexler's working group
AI	– Generalized use of drones – External and internal "Roomba-ization"	– Limitation of risks attendant to driverless vehicles – AI's social and political impartiality (view from nowhere)

We have also observed the workings of "encirclement by what is obvious," as Jacques Ellul put it. In cybergenetics, this phenomenon takes on the features of a *genethos*, and leads to an artificialist fallacy of which artificial intelligence offers the best example.

Nevertheless, it seems to me that, even if the process Ellul outlined is helpful in describing the process that is currently under way, this characterization is not sufficient. The NBIC system indeed presents the characteristics described by Ellul in *The Technological Bluff* ([1990] 1998):

1. The operators of the system have stopped trying to resolve conflicts, stopped trying to adapt the economy, politics, and society to technology, stopped trying to produce mutants.
2. This calmed climate allows "the force of things" (media, communication, imaging, and so on) to overflow or encircle the points of resistance, which dissolve and give way to self-evidence.
3. This encirclement rests on deep foundations that include the suppression of moral judgment.
4. Technology is presented as close, familiar, individualistic, and personal; this gives rise to a basic adherence to this at once reassuring and innovating banality. This adherence reinforces the absence of conflict and is in conformity with point 1.

However, the process also appears to rely on a denegation or even an inversion of the very meaning of care. This shift allows the process to succeed, for what has been the source of our strength as a species and as civilizations is becoming a flaw that allows the NBIC system to be anchored in a particularly powerful way. The inversion is particularly apparent in some of the examples we have analyzed: the genethos, for example, or our attachment to companion AIs. The inversion makes it possible to replace our identity as individuals caught up in relationships with others by an identity limited to mere connexions, disallowing what should be the complementarity of these forms of identity.

The artificialist fallacy is only the symptom of a problem, of course; it is incumbent on us to seek new paths for conceptualizing the ethics of the new technologies, and to open up the possibility of alternatives. It is in this framework that the ethics and politics of care can appear particularly fruitful. Rather than giving in to the inversion of care, we need on the contrary to affirm the constructive power of care as it is understood in feminist research. First, though, following up on the coalescence of a polethical moment of normative opening with a moment of care invoked in the introduction, it will be helpful to complete the analysis of the difficulties we are likely to face in this reaffirmation.

1 Relations and patriarchy

To begin with, I suggest that one of the central stumbling blocks confronting us is the gradual replacement of relationships by connexions. Connexions have indeed appeared as one of the principal elements encountered in this study. Nanotechnologies are nothing without interconnections; cybergenetics reinforces connexions and is built around them; artificial intelligence relies on connexion, which is a key concept in the digital field, moreover. This is why I have proposed to distinguish between the "individual in connexion" and the "individual in relations."

When Sherry Turkle began to write *Reclaiming Conversation*, which she characterized as "a book critical of our inattention to each other in our always-connected lives" (2015, 17), she was in a way responding to the syndrome analyzed in *Alone Together* (2011), that of gradual isolation of individuals as the number of their connexions grew. What is more, Turkle saw a meaningful relation between the increasing use of digital media by young adults and a significant decrease in their capacity for empathy. In addition to her own fieldwork, she relied on a 2011 study conducted by psychologist Sara Konrath and her colleagues, titled "Changes in Dispositional Empathy in American College Students Over Time: A Meta-Analysis," The authors drew on 72 surveys conducted among young adults over a period of thirty years, during which they refined the models of empathy that were tested. The studies analyzed showed a significant decrease in empathy around the end of the 2000s, especially in what is called "empathy concern" and the ability to put oneself in the place of another.

Turkle's diagnosis confirms a cluster of congruent psychological studies, especially those conducted by Niobe Way and her team (Way's book, *Deep Secrets*, also appeared in 2011) and by Carol Gilligan and Naomi Snider, *Why Does Patriarchy Persist?* (2018), in the area of the ethics and politics of care; these studies emphasize the way the loss of relationships[1] is directly tied to the development and reinforcement of a patriarchal culture.

The congruence between these lines of analysis in their conclusions strikes me as more than coincidental. It seems to me that it shows clearly why it is so easy to develop the new technologies in the direction of connectivity and away from relationships, by relying simultaneously on two things: on strengths that can turn into natural vulnerabilities, as shown for example by the "media equation" (Reeves and Nass 1996), but also on socially constructed vulnerabilities that support a patriarchal system and are at the same time supported by it (Gilligan and Snider 2018).

One of the "advantages" of this *status quo*, as Gilligan and Snider point out, is that it anesthetizes us, promising that we will not suffer from the loss of relationships; this is one of the psychological mechanisms that lead to acceptance of the loss. This is how the patriarchy comes into being and persists, in a closed loop. And the loss of empathy connected with the use of technologies that are often subtended by a patriarchal view of the world supports and reinforces that primary and prior loss of empathy. The systemic effect that provides mutual reinforcement among the various components of NBIC is thus strengthened by a second one, which plays out between, on the one hand, the loss of empathy linked to technological modifications and to the patriarchal structure that underlies artificial intelligence, for example, and, on the other hand, the dynamics through which the patriarchy is maintained. The combination of these two systemic effects makes the situation at once all the more comprehensible, in that they are mutually reinforcing, and all the more concerning, in that they are hard to disassemble. This is how Ellul's "encirclement by what is obvious" (specifically by technological self-evidence), which already presupposed a loss of the moral sense, is reinforced in an encirclement by the ethical and political self-evidence of patriarchy.

[1] Regrettably for my purposes, these authors sometimes use the word "connections" for what I call "relations" or "relationships."

2 Two central problems: new generations and deresponsibilization

This weakening of relations goes hand in hand with two other problems we have already encountered. The first is that of the coming generations: at risk of not developing their capacity for empathy, they might not be able to develop a moral sense. They would thus be made up of individuals able to be in connexion but not in relationships.

In a previous book (2011), I hypothesized that the development of our moral sense is based on a "naïve morality" developing from basic human empathic capacities that we appear to share with certain animals and some – though by no means all – humans with cognitive disabilities. Given that there is a very close link between empathic development and moral development, the problem posed by the lack of empathic development, as manifested in particular by a lack of concern for others, is such that one has to think about what is at stake for ethics in these new technologies not only at the practical level (for example, how can we conceive of a moral education with AI?) and at the metaethical level (what conception of morality would thus be mobilized?), but also in terms of the developmental and cognitive relations between the two levels. The important point here is that we confront not only the problem of how to transfer the ability to make moral judgments, but also the problem of how to deal with the mutation and loss of complex moral reasoning that accompany our increasing use of the new technologies. The kind of "*deskillization*" that occurs here involves not only competencies that could be considered technical (for example, the ability to drive a car) but fundamental social competencies as well.

The second problem is the dismissal of responsibility, the *deresponsibilization* we encountered in the example of autonomous machines. I contend that the difficulty we have in envisioning our responsibility with respect to the rise of the new technologies derives in large part from the blurring of the classical categories in which we customarily do our thinking. This blurring intermingles with two other distinct but complementary phenomena.

The first of these is the *conceptual fog* in which we move when it comes to the development of the new technologies. As George Berkeley noted, "We have first raised a dust and then complain we cannot see" ([1710] 1998, introduction, section 3). The very terms in which we speak are problematic: the terms "trust," "intelligence," or "autonomy," as we have seen, but also terms that may appear more anodyne, such as "life cycle" applied to artificial intelligence, or a deterministic vocabulary for describing DNA (which is presumed moreover to be a veritable "biological program"), or the use of passive terms to refer to humans while active terms are reserved for AI, and so on.

The second phenomenon is the *weakening of relations* that I have already underlined. I suggest that we need to develop a new concept of responsibility that takes into account both the blurring of categories and the weakening of relations by foregrounding the question of relations and rethinking the question of boundaries.

The classic response to this second problem is usually to transfer the question of responsibility either to the machine or to some collective agent, often envisioned from the perspective of "many hands,"[2] as a way of accounting for the difficulty that arises when one tries to assign responsibility to a single person. Since machines have no moral or legal responsibility (at least not yet), the question is generally raised in terms of collective responsibility. But the problem raised by collectivizing is that it contributes to a dilution of responsibility and thus leads ultimately to a form of deresponsibilization.

3 A relational responsibility?

Instead of thinking in terms of diluted collective responsibility and deresponsabilization, I suggest that, where the NBICs are concerned, a relational approach could help us better develop the question of responsibility. After all, one of the principal difficulties we encounter when we try to address the question of responsibility is determining who is responsible (and sometimes for what). This difficulty is increased by the dilution or transfer of responsibility we find in NBIC. Instead of resorting to an assertion of collective responsibility, it

[2] This expression was proposed in Thompson 1980, "Moral Responsibility of Public Officials: The Problem of Many Hands."

seems to me that one could usefully reformulate the problem by following the approach of Iris Young in "Katrina, Too Much Blame, Not Enough Responsibility" (2006). Working in the lineage of Hannah Arendt, Young has suggested distinguishing responsibility from guilt or blame. Rather than focusing on the question of guilt anchored in a retroactive vision (reflecting and acting in response to the past) by seeking to localize the fault and assign blame to an individual or a group, the relational conception of responsibility would commit to a proactive vision (reflecting and acting in relation to the future) by seeking to encourage relations among individuals rather than isolating them. Young's model is based on social relationships. She argues against focusing on the question of "whodunit," as one would in a police investigation or a detective story; in her view, this approach too often turns into a "blame game" in which each party levels charges against the other, either from a defensive position – for those being accused, or afraid of being accused – or from a position of anger – sometimes justified – in the victims' case.

The notion of relational responsibility rejects the gap between individual and institutional or structural responsibility. It is a matter neither of thinking that each person should take control of his or her situation nor of asking what the authorities are doing; it is rather a matter of recognizing that responsibility is shared: not necessarily in the same way or at the same level, but in a way that makes it impossible to deny responsibility on the grounds that "it's not my job."

Young points out three additional strategies that are often adopted by individuals and groups seeking to relieve themselves of responsibility: first, *reification* – social relations are treated as things; second, *denial of connection* – a denial of vulnerability and dependency; and, third, the *demands of immediacy*. It seems to me that all three bring into play various aspects of our capacity to relate to others and to ourselves, above and beyond the risks of reification and commodification, loss of empathy, and displaced attention that we have observed in the fields of cybergenetics and artificial intelligence.

In an interesting way, Young has developed her conception of relational responsibilities in connection with the Hurricane Katrina catastrophe. She insists both on the fact that it is *wrong* for our societies to allow people to be particularly vulnerable, and on the importance of reflecting and acting on a global as well as a local scale. Her analysis of the Katrina crisis resonates with the classic debate between Voltaire and Rousseau after the disastrous Lisbon earthquake in 1755, an event that is often cited as the moment when the concept of risk management was born.[3] Here again the question that arises is not really that of risk but rather that of the *oikos* – the household – and the way individuals are linked with one another within social, ethical, and political structures.

It seems to me that these two questions are particularly important in the case of NBIC, and that ecological considerations are also relevant in this context. For one thing, environmental ethics has contributed significantly to a renewal of ethical reflection without entering into bottom-up or top-down schemas but rather by trying to propose new concepts. In addition, the question of our relation as humans to other living beings arises in many respects, perhaps most notably when artificial entities are presented in the guise of pet animals and thus (among other things) distort our relation to animality. As the philosopher and ethicist Sophie Cloutier has emphasized in an article exploring the effects of robotic pets on children (2019), our relations with animals allow us to develop both our empathy and our moral imagination along with our relation to otherness and to resistance. Moreover, as I suggested in an earlier book (2011), we share with certain animals the basis for our moral capacity, something that I call naïve morality; this means that animals are not only recipients of moral concerns but also, at least in certain respects, moral agents – unlike AI machines, as we have seen. Furthermore, the mechanical agent takes us back to the question of concern for others,

[3] Whereas Voltaire blamed nature alone for the catastrophe, Rousseau pointed out that humans, who did not take into account the possibility of destruction and thus did not take any precautions, could be considered partly responsible. This polemic marked a turning point, for if human beings are partly responsible, it also means that they can actively seek to prevent catastrophes instead of giving in wholly to fatalism.

of the construction of "near-and-dear" beings as non-mechanizable approximations (and thus without a binary opposition to remote beings), which is one of the essential elements of polethics. As Michel Deguy writes (2001, 34-35):

> Poetic thinking is approximative. The near-and-dear must be brought close to enter into nearness ('my neighbor'): by approximations. This motif of coming together is equally ethical and political: come together with one another. Approach! Approach! And animals, so unfathomably neighbors, come close to us. (Deguy 2001, *Spleen de Paris*, 34-35)

As Catherine Larrère and Raphaël Larrère have emphasized (2017), it is important to recontextualize the new technologies and in particular not to forget that they are also a "mode of relation to nature." The problem is that as these technologies insinuate themselves into our perception and our representation of living beings, they somehow take the place, in our *oikos*, of domestic animality, or even take the form of perfectly trained imitations of it that leave no room for resistance and thereby makes us forget the alterity and resistance of the natural beings.

Finally, given that the question to be faced is how to determine what entities we find it desirable (or not) to welcome into our *oikos*, our households, and according to what modalities, I suggest that the ethics of care, which is inextricably political – and thus intrinsically a polethics – can help us find an answer, as I shall now seek to show.

4 Toward a polethics of care for NBIC?

The question of relations can no doubt be envisioned according to various approaches, in different times and cultures.[4] Without neglecting the possible complementarity of these approaches, it seems worthwhile to come back once again to the ethics and politics of care.

Among the many reasons that justify this approach to conceptualizing NBIC ethics, one is the way that problems of empathy and relations of domination keep intersecting. If we are, in a "polethic"

[4] They range from the Stoics' *commendatio* to the South African *Ubuntu* (the common bond of sharing which is our humanity) by way of the *Siris*, i.e., the chain that connects all beings, according to Berkeley.

rather than in a "Sputnik" moment today, it is probably because seeking to approach the ethical and political challenge posed by NBIC in terms of risk and control means missing a major part of the problem and even what makes it interesting. After all, etymologically speaking, interest (from the Latin *inter* + *esse*, "between" and "to be") means that which is between us: in other words, relations. What interests us is thus what brings us together.

Coming back to the question of the interest of NBIC, then, is to take seriously Alan Turing's "social quest" as analyzed by Juliet Floyd (2019): the goal would be to reorient the development of NBIC in a way that would respond to that quest. And, recalling Sherry Turkle's *Alone Together* (2011) with its telling subtitle, *Why We Expect More from Technology and Less from Each Other*, we might do well to invert the terms and expect less from technology and more from ourselves, that is, be more benevolent and more demanding of ourselves and of each other. But what would this look like, and how might we go about it? Let me suggest several possible paths.

The first is precisely the one Turkle proposed in *Reclaiming Conversation* (2015). We can develop this theme in a more philosophical direction as commerce (understood in its old sense of exchange) rather than clientelism (understood in its old sense of a relation of economic and political subordination) as we encountered it in recreational cybergenetics. In its philosophical sense, conversation also entails taking on a dimension of unpredictability and chance, a dimension that distances us from a logic of risk management. As Ali Benmakhlouf puts it in *La conversation comme moyen de vivre* (which also uses the concept of commerce, characterized as a total social relation on the model of *potlatch*), "one must approach conversation without expecting anything" (2016, 103). Further on, he invokes Wittgenstein's well-known remark on conversation: "One person throws a ball; the other does not know: is he to throw it back, throw it to a third person, or leave it lying, or pick it up & put it in his pocket, etc."[5] This conversation is also one in which we must hear

[5] Ludwig Wittgenstein (1948) 1998, *Culture and Value*, 84, cited in Benmakhlouf 2016, 105.

the “different voices,” the minority or discordant voices such as those evoked by Carol Gilligan (1982), which include the voice of the ethics of care.

Like Gilligan’s little Amy, when we are confronted with the structure of a dilemma, rather than letting ourselves be trapped in a binary system (which often conceals a TINA), we can propose an opening toward other possibilities. It seems to me that this opening can take place in multiple ways, as Amy demonstrates. While I do not have room here to explore the whole range of alternative modalities, there are two that strike me as essential.

The first entails exiting from the natural/artificial dichotomy by bringing into play the question of cultures. I am well aware that my own analysis is culturally situated and that it stands to benefit greatly from an opening toward other cultural fields, Asian in particular, that are fundamental for my purposes.

The second entails exiting from that same dichotomy by way of the arts. It seems to me that the exploration of the ethics of NBIC has everything to gain by turning to artistic creations to find openings toward new possibilities.[6] It seems to me, too, that the way in which these technologies are rethought via popular culture is fundamental and that one of the primordial questions to raise here – as a counterpoint to the “nudges” on which the digital field relies so heavily, or the imposed necessity of *opting out*, when it is *opting in* that should not be automatic (as for example when one has to *opt out* of accepting digital cookies rather than opting in) – is the prospect of a moral education. Pursuing this question would entail neither a development of performance in the prosthetic mode supported by the transhumanists nor a promotion of immortality without regard for any interrogation or valorization of the capacities and real conditions of life.

[6] Here I agree completely with Yves Citton and Pierre Cassou-Noguès, who propose – from different perspectives – to nourish reflection on the new technologies through artistic creation. See Citton 2020, “Logique et esthétique du drone armé,” and Pierre Cassou-Noguès 2019, *Technofictions*.

5 An ethics by design

The next step would be to interrogate interest in terms of what we value, in Dewey's sense. To question ourselves, as Sandra Laugier subtly suggests, echoing Harry Frankfurt, about "the importance of importance," asking ourselves how and with whom we want to develop our *oikos*, our household.[7] Thus one way to evaluate the NBIC could be to ask whether they encourage relationships or instead dilute them – whether they are content to weave simple connexions understood as simulacra of relationships.

Still, we must not fall into paternalism, and rather than positing relationship as a value in itself –the way one might posit autonomy as a value in itself, a sort of taboo or fetish[8] – we might posit, with the encouragement of Bruno Latour (2000, 192), the question of good and bad attachments: "It is no longer a matter of asking whether one must be free or attached, but *whether one is well or poorly attached*."[9]

The question of closeness and remoteness thus turns into a questioning of approximation, in Michel Deguy's sense, along with a questioning of justness, the right or proper distance.[10] This is the question Maria Puig de la Bellacasa is raising in *Matters of Care: Speculative Ethics in More Than Human Worlds* (2017) when she proposes to pass from Latour's *matters of concern* to *matters of care*. Her approach is all

[7] See for example Laugier 2009, "Le sujet du care: Vulnérabilité et expression ordinaire," 181, and Laugier 2005, "L'importance de l'importance: Expérience, pragmatisme, transcendantalisme."

[8] In bioethics in particular, "autonomy" is unfortunately sometimes brandished as a slogan, an unquestionable truth, without any attempt to analyze the notion or to specify the kind of autonomy that may be relevant in a given context.

[9] The italics are in the original. Latour stressed the importance of transforming two fundamental conceptions: one bearing on the nature of attachments, the other on the form of control.

[10] This is also the question raised by care, for an excess of care may end up with poor care, as in the case of overzealous therapeutic interventions. Hence the need to take various points of view into account into the process. This thematics of the proper distance can also be linked to Donald Winicott's "good enough" mother, as is stressed in Brugère 2014, *Qui a peur des philosophes? Entretien avec Elodie Maurot*, 68. This position contrasts with the one expressed in the "mommy blogs" that we encountered in the cybergenetic constellation.

the more essential in that she stresses the extent to which the difference between facts and values is unfruitful, especially when one is considering the construction of a technological object. Her arguments reinforce the contention that the problem posed by the artificialist fallacy is not so much a gap between "is" and "ought" (as the naturalistic fallacy is usually presented), but rather a reduction in the sense of a closing off of possibilities.

Puig de la Bellacasa's reflections form part of a convergent cluster of recent works that seek to conceptualize the way in which we think and act in a world marked by the new technologies, relying on the ethics and politics of care. While I have mentioned some of these works in passing, I now propose to mention briefly how this book is in conversation with this research for instance by showing how it is both complementary to and distinguishable from the work of certain others.

In France, the philosopher Xavier Pavie, examining the concept of innovation, has proposed the term "innovation-care," in a sense very close to what I understand by "responsible innovation" (about which I have already expressed some reservations).[11] Pavie's definition is completed by a form of categorical imperative that enjoins us never to treat humanity as a means but always as an end. However, his approach assumes a hierarchy between care and innovation in which innovation has the highest priority.

One might propose, on the contrary, that care, in its feminist dimension, cannot really embrace that sort of hierarchical relation and instead invites us to try to think in non-contradictory terms, for care is integrated into a "good" innovation, if innovation is understood as an advance apt to benefit all parties – including especially the most vulnerable – without being imposed on them.

Another approach inspired explicitly by the ethics and politics of care has been proposed in the field of medical robots by Aimee van

[11] See Pavie 2020, *Critical Philosophy of Innovation and the Innovator* and 2014, "The Importance of Responsible Innovation and the Necessity of 'Innovation-Care.'"

Wynsberghe (2015), with her Care-Centred Value-Sensitive Design. She proposes a theoretical framework based most notably on Joan Tronto's (1993) politics of care, stressing four fundamental elements: attention, responsibility, competence, and the capacity for *responsiveness*.

To summarize a point I have stressed repeatedly in these pages, to conceptualize an ethics with NBIC is to conceptualize the ethics of a world in the making, for new generations. As social scientist Christopher Groves emphasizes in *Care, Uncertainty and Intergenerational Ethics* (2014, 158):

> The dependence of future generations on us is not immediate and of a parental nature, but is instead mediated by the 'things' that matter, together … with the practices, institutions and technologies that enable us to look after them.

Perhaps, then, we should undo the Collingridge dilemma (which holds that at the start we can act but not foresee, while at the end we can foresee but no longer act) by ceasing to think in terms of foresight and control, asking ourselves instead how to take care of what matters to us.

This is why the approach via care seems to me inseparable from an ethics by design that would be brought into play at every stage in the conception, creation, and development of the technologies in question; it would allow an opening toward new possibilities while allowing future generations to reject or reprogram selected aspects of AI software.[12] An ethics by design of this sort would make it possible to grasp ethics as a whole that is not limited to either the intention or the impacts of AI but that questions the process in itself as a bearer of values. It would also resituate at the center of its dynamics the idea that the future will need the new generations, as opposed to Bill Joy's claim, cited in the introduction, that "the future doesn't need us." Moreover, the notion that the future needs future generations must

[12] The early symptoms of ethics by design have been set forth in Nurock et al. 2021, "What Does 'Ethical by Design' Mean?"

constitute a moral imperative for us. We must leave to future generations the possibility of generating alternatives, of reconstructing their world in a different way – just as we may need Notre Dame without necessarily rebuilding it as an exact replica of its predecessor. Furthermore, this dynamic opening must operate both horizontally, toward the generations to come, and vertically, toward the cultures on which the social models subtending NBIC have been imposed, in an inversion of the approach of cultural vassalization. It is thus important to be able to keep in mind not the so-called Gabor's law but rather Gabor's actual assertion that one cannot predict the future but that one can and must invent it.

How are we to conceive of the world that has been remodeled under the influence of NBIC in terms of the ethics of care? As we have seen, it is difficult to settle on a stable definition of care, especially because care does not speak with a single voice; instead, it constitutes a multifaceted way of conceptualizing our ethical and political relation to ourselves, to others, and to the world. As we have seen throughout this book, this current of thought proposes to broaden the field of ethics beyond the questions of justice or goodness; it seeks to escape from dichotomies in order to raise questions connected with care in all its dimensions. As Fabienne Brugère has described it (2017, 5), care is a "theoretical and practical revolution," subject to a global but not a globalizing analysis.

Care is thus exercised through attention to the particular and through the experience of vulnerability. Unlike NBIC, which bears the trappings of the extraordinary as they naturalize our social life and thereby artificialize our moral life, and as they manage to slip into the most ordinary, the most trivial – but sometimes also the most fundamental – folds of our private as well as our public lives, the ethics and politics of care are characterized by attention to the ordinary. As Sandra Laugier has shown, this attention to the ordinary and to vulnerability, to the little things that are usually made invisible (as we saw over and over with NBIC, from "click work" to the defeminization of AI and the curation of data), is one of the principal characteristics of care. According to Laugier (2011), "care is defined on the basis of that specific attention to the importance of the 'little' things and moments, and to the inherent dissimulation of that importance in our daily lives." As we have seen, care also draws our attention to mutual relationships and our interdependencies.

It is precisely for that reason, one might add, that care can serve both as a scanner to detect the relations of control at work in NBIC – for example, the unacknowledged patriarchal character of artificial moral systems, as we have seen – and as 3D glasses that can offer an in-depth view of the possibilities open to us – as young Amy does, for instance, when she opens up the dilemma she has confronted. Such tools thus complete the myopic and synoptic visions that I had proposed as the initial method of analysis. Among the "little things" we find, to paraphrase Berkeley again, the "dust of words" that we stir up before complaining that we can no longer see a thing. Thus terms such as "autonomy," "intelligence," and "trust," among others, terms that we tend to use rather carelessly, are among the little things to which we could, if we wished – indeed, to which we should – pay more rigorous attention.

If we linger a moment on the term "trust," for example, a word that is used as a key term in cybergenetics, nanotechnologies,[13] and AI alike, we find it problematic – some commentators even evoke a trust gap that separates the public from the nanos and calls for a new form of social contract.[14] I suggest that this focus on the notion of trust leads to the same problem as the one evoked in the chapter on nanotechnologies: confusion between acceptability and acceptance, between marketing argument and ethical analysis. As Norwegian philosopher Trond Grønli Åm pointed out in "Trust in Nanotechnology? On Trust as Analytical Tool in Social Research on Emerging Technologies" (2011), using the notion of trust where nanotechologies are concerned is problematic for at least two reasons. First, owing to the difficulty of defining these technologies, so that the object of trust

[13] For an analysis of the notion of trust in connection with nanotechnologies, see especially Anderson and Petersen 2011, "Nanotechnologies and Trust."

[14] See for example Rejeski 2008, "Nanotechnology and the Trust Gap."

appears blurred. Second, owing to the anchorage of the concept of trust in a relational, interpersonal dimension that involves assuming a form of vulnerability with respect to others. It seems to me that these two arguments apply just as well to the question of trust in artificial intelligence: the difficulty of determining the object and its consequences, but especially the absence of mutual relations, unless – and this is an essential point – one credits AI programs and devices with intentionality. Thus the question of vulnerability probably needs to be formulated in a different way. Not only is the term "trust" a poor choice, but it also serves to conceal a notion closer to that of social acceptability, and it tends to push us toward confusing an entity governed by artificial intelligence with a partner.[15]

However, as some commentators have noted, the notion of trust differs from that of reliability, for it concerns a psychological and social attitude on the part of one person toward another person or toward an institution. As computer scientist Joanna Bryson has argued in "No One Should Trust Artificial Intelligence" (2018), artificial intelligence is inherently untrustworthy: no one *can* trust it, because it is itself unable to trust. AI thus cannot be "worthy" of trust, any more than its quasi-magical power is worthy of belief, one might add. Those who speak of trust in relation to NBIC in general and to AI in particular are playing once again on our vulnerabilities and on the "media equation," leading us to think in terms of what Meredith Broussard calls technochauvinism. This is why, if we return to my initial invocation of the Enlightenment and follow the path of Jean-Jacques Rousseau, we can understand why the new social contract is not a Social Contract 2.0, but rather an agreement to *care*

[15] The case of cybergenetics is slightly different, for the public and commercial discourses surrounding it play both on mistrust (toward the medical profession, where DTC genetic testing businesses are concerned) and on trust in the mantra about data sharing, which is even seen as the new social contract 2.0, as I have noted. But mistrust directed toward the medical community is a lever used to promote trust in the DTC businesses by way of participation in a medical context where the economy plays an important role in the doctor-patient relation.

with: it implies caring with others for our sharedbonds, political and institutional bonds included.

For the purpose of conceptualizing the ethics of NBIC, rather than providing answers, care ethics and care politics can help us raise a few simple questions such as these:

1. What do we care about: what matters to us?
2. Of what or of whom do we take care, and have we paid particular attention to our vulnerabilities and to the most vulnerable?
3. Have we been careful about the whole process?
4. How do we care with, e.g. collaborate to make sure to strengthen democracy or, at least, not weaken it.

These must be understood of course as questions among others, just as polethics does not have a globalizing vocation but is one voice among others. Still, they allow us to enter into a regime of attention without which true rigor is impossible, and a regime of fairness without which true equity cannot be achieved. Unlike Facebook's initial slogan, "move fast and break things," the ethics and politics of care make it possible to "maintain, continue, and repair our 'world.'"[16] It is not a question of preserving our world unchanged, as the American environmentalist conservationists would have it, for instance; not a question of being reactionary or technophobic, but rather of considering what matters to us; not a question of whether the future needs us but rather what future we want to build together.

Before concluding, let me attempt to summarize by recalling three key ideas.

First, that it is necessary to interrogate the polethics of NBIC understood as a coherent system, even though each of the fields analyzed (nanotechnologies, cybergenetics, artificial intelligence) presents its own specificities. This system is based primarily on a shared manner of self-description, especially through self-realizing prophecies and a mythology with strong American accents, and on a blurring of categories, a reconstruction of identities and relations. This

[16] See Fisher and Tronto 1990, 40.

system converges, moreover, with another, one that denies the complementarity between individuals in connexion and individuals in relation, and that generates interactions between, on the one hand, the loss of empathy and the persistence of the patriarchal infrastructure, and, on the other hand, the dynamics through which the patriarchy takes shape and persists.

Second, that this reconstruction is based on an artificialist fallacy linked to two distinct but complementary forms of naturalization. This fallacy entails a reduction of horizons to the form of a dichotomy that often takes on the guise of an etic (and not ethical) dilemma, since it allows for no alternative. This TINA is inseparable from a form of neocapitalist and technochauvinst neodarwinism that it is important to bring to light and counter, if we are to avoid ending up with a genethos as well as a reduction, an artificialization, of ethics.

Third, that to conceptualize an ethics with NBIC, we need (among other things) an ethics by design articulated with a polethics of care that interrogates our *oikos* while restoring to the front ranks the question of relations – rather than connexions, although without denying the importance of connexions. This polethics suggests that a good innovation is made with care (rather than giving innovation priority over care) and proposes to conceive of our responsibility in terms that are not paternalistic but rather relational.

It is important to commit ourselves now to constructing the future, because we are relationally responsible for the present and for the future that we are in the process of building; it is important to take care of the institutions and infrastructures that surround us, and to conceive of our being together with NBIC. It is thus up to us to write our history, just as Rousseau wrote his *Confessions*, not in the form of a self-fulfilling prophecy, or on Facebook, but rather in the form of a prospective responsibility. Rousseau's *Emile*, for its part, has to be thought in terms of moral education rather than in terms of a paternalistic "nudging." Perhaps we ought to return to the way Emmanuel Kant characterized the Enlightenment, as an exit from childhood, with the maxim "Sapere aude": Dare to use your understanding!

We should also return to Descartes' "I think therefore I am" and not "I share therefore I am," if sharing means forgetting to think…

In the "dust of words" we have stirred up, we also find the word "intelligence" and the way it lends itself to misunderstanding. Intelligence is sometimes measured by the yardstick of performance and "Roomba-ization," or even by the yardstick of the adaptability of our common world, of our "smart" cities with their technological limitations. The problem, to be sure, as a number of commentators have noted, is that AI lacks common sense.[17] We could add that it also lacks moral sense, that it is not reasonable – if it were normal to expect that it had these characteristics. But as we have seen, this is not the case: ethics cannot be diluted so as to be injected into the new technologies.

Intelligence thus is not characterized by routines but by daring and the opening of possibilities. And it is indeed to this definition that it seems necessary to return: "we must dare, and dare again, and go on daring," as Georges Danton insisted in 1792. In the early 2020s, such daring requires assuming our humanity, our release from tutelage, finding our own voice and daring to think while using both our common sense and our moral sense.

[17] See for example Marcus and Davis 2009, *Rebooting AI: Building Artificial Intelligence We Can Trust*.

Bibliography

#NotABugSplat. N.d. "A Giant Art Installation Targets Predator Drone Operations."
https://notabugsplat.com/.

"Book of Mormon and DNA Studies." N.d.
https://abn.churchofjesuschrist.org/study/manual/gospel-topics-essays/book-of-mormon-and-dna-studies?lang=eng/.9620%20.

"Forbes Releases 2019 List of America's Richest Self-Made Women, A Ranking of the Most Successful Women Entrepreneurs in the Country." June 4.
https://www.forbes.com/sites/forbespr/2019/06/04/forbes-releases-2019-list-of-americas-richest-self-made-women-a-ranking-of-the-most-successful-women-entrepreneurs-in-the-country/?sh=7c92009f4e41.

"Hacking Roomba: Projects Repository." N.d.
https://hackingroomba.com/projects.

"Roomba Hack: Spirograph."
https://www.flickr.com/photos/49147885@N00/507132860.

23andMe.com. 2019. "Tips for using the new 23andMe Forums."
https://customercare.23andme.com/hc/en-us/articles/214116497-Tips-for-using-the-new-23andMe-Forums

Abbas, Ali, Max Senges, and Ronald Howard. 2019. "A Hippocratic Oath for Technologies." In Ali Abbas, ed., 2019, *Next-Generation Ethics: Engineering a Better Society,* 71-80. Cambridge, UK: Cambridge University Press.

Abbasi, Ahmed, Jingjing Li, Gari Clifford, and Herman Taylor. 2018. "Make 'Fairness by Design' Part of Machine Learning." *Harvard Business Review*. August 1.
https://hbr.org/2018/08/make-fairness-by-design-part-of-machine-learning.

Adam, Alison. 1998. *Artificial Knowing: Gender and the Thinking Machine*. London: Routledge.

Adam, Alison. 2005. *Gender, Ethics and Information Technology*. New York: Palgrave Macmillan.

Agamben, Giorgio. (2011) 2013. The Highest Poverty: Monastic Rules and Form-of-Life. Translated by Adam Kotsko. Stanford, CA: Stanford University Press.

Aldhous, Peter. "A Security Breach Exposed More Than One Million DNA on a Major Genealogy Database." *BuzzFeed News*.
https://www.buzzfeednews.com/article/peteraldhous/hackers-gedmatch-dna-privacy.

Allen, Colin, Iva Smit, and Wendel Wallach. 2005. "Artificial Morality: Top-down, Bottom-up, and Hybrid Approaches." *Ethics and Information Technology* 7 (3): 149-55.

Allhoff, Fritz. 2007. "On the Autonomy and Justification of Nanoethics." *NanoEthics* 1 (3). 185-210.

Allyse, Megan A., David H. Robinson, Matthew J. Ferber, and Richard R. Sharp. 2018. "Direct-to-Consumer Testing 2.0: Emerging Models of Direct-to-Consumer Genetic Testing." *Mayo Clinic Proceedings* 93 (1): 113-20.

Åm, Trond Grønli. 2011. "Trust in Nanotechnology? On Trust as Analytical Tool in Social Research on Emerging Technologies." *Nanoethics* 5 (1): 15-28.

American Association for the Advancement of Science. 2020. "The Moral Machine." May 20.
https://www.eurekalert.org/news-releases/600655.

American Society of Human Genetics. 2019. "Researchers Quantify Limitations of Health Report from Direct-to-Consumer Genetic Tests." October 17.
https://www.ashg.org/publications-news/press-releases/201910-limitations-direct-consumer.

Amiel, Philippe. 2009. "Expérimentations médicales: Les Médecins nazis devant leurs juges." In François Vialla, ed., 2006, *Les grandes décisions du droit médical*, 431-44. Paris: LGDJ.

Anderson, Alison, and Alan Petersen. 2011, "Nanotechnologies and Trust." In Christopher Candlin and Jonathan Crichton, eds., 2011, *Discourses of Trust*, 237-51. New York: Palgrave Macmillan.

Anderson, Michael, Susan Leigh Anderson, and Chris Armen. 2005. "Towards Machine Ethics: Implementing Two Action-based Ethical Theories." *AAAI Fall Symposium – Technical Report*, 1-7.

Anderson, Michael, Susan Leigh Anderson, and Chris Armen. 2006. "MedEthEx: A Prototype Medical Ethics Advisor." *Proceedings: The Twenty-First National Conference on Artificial Intelligence and the Eighteenth Innovative Applications of Artificial Intelligence Conference*, 1759-65. Menlo Park, CA: AAAI Press.

Ang, Jennifer. 2019. "Drone Warfare and Moral Buffers." *The Asian Conference on Ethics, Religion and Philosophy 2019*. Tokyo, March 21-23.

Annas, George J. 1990, "Mapping the Human Genome and the Meaning of Monster Mythology." *Emory Law Journal* 39: 629-64.

Annas, George J., and Sherman Elias, eds. 1992. *Gene Mapping: Using Law and Ethics as Guide*. Oxford, UK: Oxford University Press.

Apple, Michael W. 1994. "Computers and the Deskilling of Teaching." *Computer Professionals for Social Responsibility Newsletter* 12 (2).
http://cpsr.org/prevsite/publications/newsletters/issues/1994/Spring1994/apple.html/.

Archyde. 2020. "Meyko, the Robot for Sick Children, Grew Up." January 5.
https://www.archyde.com/meyko-the-robot-for-sick-children-grew-up/.

Arlandis, Fanny. 2014. "Drone et photo d'enfant: Le nouveau projet ultra-démago de l'artiste JR." *Slate*, April 8.
https://www.slate.fr/culture/85701/jr-photo-drone-pakistan.

Avance, Rosemary. 2015. "Constructing Religion in the Digital Age: The Internet and Modern Mormon Identities." *Publicly Accessible Penn Dissertations*, no. 1583. Philadelphia: University of Pennsylvania.

Awad, Edmond, Sohan Dsouza, Richard Kim, Jonathan Schulz, Joseph Henrich, Azim Shariff, Jean-François Bonnefon, and Iyad Rahwan. 2018, "The Moral Machine Experiment." *Nature* 563 (October 28): 59-64.

Bacchini, Fabio. 2013. "Is Nanotechnology Giving Rise to New Ethical Problems?" *NanoEthics* 7: 107-19.

Baron-Cohen, Simon. 2003. *Men, Women and the Extreme Male Brain*. New York: Penguin/Basic Books.

Baudelaire, Charles. (1857) 1993. *Flowers of Evil*. Translated by James McGowan. Oxford, UK: Oxford University Press.

Beck, Ulrich. (1986) 1992. *Risk Society: Towards a New Modernity*. Translated by Mark Ritter. London: Sage Publications.

Benmakhlouf, Ali. 2006. *La conversation comme manière de vivre*. Paris: Albin Michel.

Bennett-Woods, Deb. 2008. *Nanotechnology: Ethics and* Society. Boca Raton: CRC Press.

Bensaude-Vincent, Bernadette, and Vanessa Nurock. 2010. "Éthique des nanotechnologies." In Emmanuel Hirsch, ed., *Traité de bioéthique*, vol. 1, *Fondements, principes, repères*, 355-69. Toulouse: Eres.

Bensaude-Vincent, Bernadette. 2013a. "Decentring Nanoethics toward Objects." *Etica & Politica* 15: 310-20.

Bensaude-Vincent, Bernadette. 2013b. "Which Focus for an Ethics in Nanotechnology Laboratories?" In Simone van der Burg and Tsjalling Swierstra, eds., *Ethics on the Laboratory Floor*, 20-36. New York: Palgrave Macmillan.

Bergen, Hilary. 2016. "'I'd Blush if I Could': Digital Assistants, Disembodied Cyborgs and the Problem of Gender." *Word and Text: A Journal of Literary Studies and Linguistics* 6: 95-113.

Berger, Anne-Emmanuelle. 1996. "Dernières nouvelles d'Écho." *Littérature*, no. 102, *Échos et traces* (May): 71-90.

Bergson, Henri. (1912) 1999. *Matter and Memory*. Translated by Nancy Margaret Paul and W. Scott Palmer London: G. Allen & Co.

Berkeley, George. (1707-) 1948. "Philosophical Commentaries." *The Works of George Berkeley Bishop of Cloyne*, edited by A. A. Luce and T. E. Jessop, vol. 1, 1-139. London: Thomas Nelson and Sons.

Berkeley, George. (1710) 1998. *A Treatise Concerning the Principles of Human Knowledge*. Edited by Jonathan Dancy. Oxford, UK: Oxford University Press.

Berkeley, George. 1709. *An Essay Towards a New Theory of Vision*. London: Aaron Rhames.

Bernelin, Margo. 2019. "Intelligence artificielle et santé: La ruée vers les données personnelles," in Vanessa Nurock, ed., *L'intelligence artificielle: Enjeux éthiques et politiques*, 75-90. Paris: Presses universitaires de France.

Berube, David. 2006. *Nano-hype: The Truth behind the Nanotechnology Buzz.* Amherst, NY: Prometheus Books.

Besnier, Jean-Michel. 2012. *L'homme simplifié: Le syndrome de la touche étoile*. Paris: Fayard.

Blok, Vincent, and Pieter Lemmens. 2015. "The Emerging Concept of Responsible Innovation: Three Reasons Why It Is Questionable and Calls for a Radical Transformation of the Concept of Innovation." In Bert-Jaap Koops, Ilse Oosterlaken, Henny Romijn, Tsjalliing Swierstra, and Jeroen van den Hoven, eds., *Responsible Innovation 2*, 19-35. Cham, CH: Springer.

Bonah, Christian. 2006. *Nazisme, science et médecine*. Paris: Glyphe.

Bonniol, Jean-Luc, and Pierre Darlu. 2014, "L'ADN au service d'une nouvelle quête des ancêtres?" *Civilisations* 63: 201-19.

Bourdieu, Pierre. (1998) 2002. *Masculine Domination*. Translated by Richard Nice. Stanford, CA: Stanford University Press.

Bourgault, Sophie, and Frans Vosman, eds. *Care Ethics in yet a Different Voice: Francophone Contributions*. Leuven: Peeters, 2020.

Bowman, Diana, Anne Dijkstra, Camilo Fautz, Julia S. Guivant, Kornelia Konrad, Harro van Lente, and Silvia Woll, eds. 2016. *Responsibility and Emerging Technologies: Experiences, Education and Beyond*. Berlin: Akademische Verlagsgesellschaft.

Boyle, Michael J. 2020. *The Drone Age: How Drone Technology Will Change War and Peace*. Oxford, UK: Oxford University Press.

Braverman, Harry. 1974. *Labor and Monopoly Capital: The Degradation of Work in the Twentieth Century*. New York: Monthly Review Press.

Brey, Philip. 2010. "Philosophy of Technology after the Empirical Turn." *Techné: Research in Philosophy and Technology* 14 (1): 36-48.

Brey, Philip. 2017. "Ethics of Emerging Technologies." In Sven Ove Hansson, ed., *The Ethics of Technology: Methods and Approaches*, 175-92. London: Rowman and Littlefield International.

Brooke, Diane. 1966. "'Sharing Time' in the Elementary School." in *Reading Horizons: A Journal of Literacy and Language Arts* 7 (4): 23-31. https://scholarworks.wmich.edu/reading_horizons/vol7/iss1/4.

Broussard, Meredith. 2018. *Artificial Unintelligence: How Computers Misunderstand the World*. Cambridge, MA: MIT Press.

Brugère, Fabienne. 2014. *Qui a peur des philosophes? Entretien avec Elodie Maurot*. Paris: Bayard.

Brunon-Ernst and Guillaume Tussaud. 2012. "Epilogue: The Panopticon as a Contemporary Icon." In Anne Brunon-Ernst, ed., *Beyond Foucault: New Perspectives on Bentham's Panopticon*, 185-200. Burlington, VT: Ashgate.

Brunon-Ernst, Anne. 2012. "Deconstructing Panopticism into the Plural Panopticons." In Anne Brunon-Ernst, ed., *Beyond Foucault: New Perspectives on Bentham's Panopticon*, 17-41. Burlington, VT: Ashgate.

Brunon-Ernst, Anne. 2014. "Beyond the Genetic Panopticon: The Limits of Government Intervention on Citizens' Bodies" *Implications philosophiques* 1. http://www.implications-philosophiques.org/actualite/une/beyond-the-genetic-panopticon/.

Brykman, Geneviève. 1993. *Berkeley et le voile des mots*. Paris: Vrin.

Brykman, Geneviève. 2010. "Courte vue et vision synoptique chez Berkeley." *Revue philosophique de la France et de l'étranger* 135 (1): 83-95.

Bryson, Joanna. 2018. "No One Should Trust Artificial Intelligence." *Science & Technology: Innovation, Governance, Technology* (November 14). http://ourworld.unu.edu/en/no-one-should-trust-artificial-intelligence.

Callamard, Agnès. 2020. *Advance Unedited Version, Report of the Special Rapporteur*. Distributed June 29. https://www.statewatch.org/media/1216/un-sr-extrajudicial-kilings-drones-report-29-6-20.pdf.

Callamard, Agnès. 2020. Interview. Radio France Internationale, July 9. https://www.rfi.fr/fr/moyen-orient/20200709-experte-onu-monde-entr%C3%A9-nouvelle-%C3%A8re-drones.

Canadian Institute for Advanced Research. 2019. "Repenser les enjeux culturels et éthiques en matière d'Intelligence Artificielle." http://eur-artec.fr/wp-.content/uploads/2019/10/Repenser-les-enjeux-culturels-Programme-BAT2-2.pdf.

Cañamero, Lola, and Matthew Lewis. 2017. "Robin – An Autonomous Robot Toddler for Diabetic Children." Proceedings, UK-RAS Conference: "Robots Working For & Among Us." https://www.ukras.org.uk/wp-content/uploads/formidable/21/UK-RAS17-5.pdf.

Canguilhem, Georges. (1989) 1991. *The Normal and the Pathological.* Translated by Carolyn R. Fawcett with Robert S. Cohen. New York: Zone Books.

Carson, Biz, and Kathleen Chaykowski. 2019. "Live Long and Prosper: How Anne Wojcicki's 23andMe Will Mine Its Giant DNA Database for Health and Wealth." Forbes, June 26. https://www.forbes.com/sites/bizcarson/2019/06/06/23andme-dna-test-anne-wojcicki-prevention-plans-drug-development/.

Casilli, Antonia A. 2019. *En attendant les robots: Enquête sur le travail du clic*. Paris: Seuil.

Cassou-Noguès, Pierre. 2018. "Le syndrome du thermomètre." *La Revue des deux Mondes* (May): 86-94.

Cassou-Noguès, Pierre. 2019. *Technofictions*. Paris: Cerf.

Caulfield, Timothy, Stephanie M Fullerton, Sarah E. Ali-Khan, et al. 2009. "Race and Ancestry in Biomedical Research: Exploring the Challenges." *Genome Medicine* 1 (8).

Caulfield, Timothy. 2002. "Genetics, 'Family Consent' and the Law." *Nature Reviews Genetics* 3: 647. https://doi.org/10.1038/nrg900.

Caulfield, Timothy. 2018a. "Is Direct-to-Consumer Genetic Testing Reifying Race?" *Policy Options*. https://policyoptions.irpp.org/magazines/march-2018/direct-consumer-genetic-testing-reifying-race/.

Caulfield, Timothy. 2018b. "Why Your DNA Test Won't Reveal the Real You." https://www.theglobeandmail.com/opinion/article-why-your-dna-test-wont-reveal-the-real-you/.

Chamayou, Grégoire. (2013) 2015. *A Theory of the Drone*. Translated by Janet Lloyd. New York: The New Press.

Chapa, Joe. 2017. "'Drone Ethics' and the Civil-Military Gap." *War on the Rocks*, June 28. https://warontherocks.com/2017/06/drone-ethics-and-the-civil-military-gap/.

Cheek, Julianne. 2008. "Healthism: A New Conservatism?" *Qualitative Health Research* 18 (7): 974-82.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Citton, Yves. 2020. "Logique et esthétique du drone armé." AOC Community Media, January 2. http://www.yvescitton.net/archeologie-des-media/.

Clark, Liat. 2015. "How One Coder Used 23andMe to Create a Race Wall around the Web." *Wired*, July 23. https://www.wired.co.uk/article/23andme-api-blocks-based-on-race-gender.

Clinton, Bill. 2000. White House Event, June 26. https://www.genome.gov/10001356/June-2000 -white-house-event.

Cloos, Christopher. 2005. "The Utilibot Project: An Autonomous Mobile Robot Based on Utilitarianism." *Machine Ethics: Papers from the 2005 AAAI Fall Symposium*, 38-45. Palo Alto, CA: AAAI Press.

Cloutier, Sophie. 2019. "Les animaux-robots: Quelles conséquences pour les enfants?" In Soheil Kash and Marie-Hélène Parizeau, *La société robotisée: Enjeux éthiques et politiques*, 289-310. Quebec City, Canada: Presses de l'Université Laval.

Coeckelbergh, Mark. 2013. "Drones, Information Technology, and Distance: Mapping the Moral Epistemology of Remote Fighting." *Ethics and Information Technology* 15 (2): 87-98.

Cofone, Ignacio N. 2020. "Nothing to Hide, But Something to Lose." *University of Toronto Law Journal* 70 (1): 64-90. https://ssrn.com/abstract=3327646. https://doi.org/10.3138/utlj.2018-0118.

Collingridge, David. 1980. *The Social Control of Technology*. New York: St. Martin's Press.

Cooper, Melinda, and Cathy Waldby. 2014. *Clinical Labor: Tissue Donors and Research Subjects in the Global Bioeconomy*. Durham, NC: Duke University Press.

Cooper, Necia Grant. 1994. *The Human Genome Project: Deciphering the Blueprint of Heredity. Mill Valley, CA: University Science Books.*

Crawford, Kate. 2018. "Artificial Intelligence's White Guy Problem." *New York Times*, June 25. https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html.

Crawford, Robert. 1980. "Healthism and the Medicalization of Everyday Life." *International Journal of Health Services* 10 (3): 365-88.

Cronin, Bruce. 2018. *Bugsplat*. Oxford, UK: Oxford University Press.

Curtis, Neal. "The Explication of the Social: Algorithms, Drones and (Counter-) Terror." *Journal of Sociology* 52 (3): 522-36.

Damour, Franck 2018a. "Le transhumanisme, inspirateur ou idiot utile du biocapitalisme? Le cas exemplaire de la vision d'Eric Drexler." *Raison présente* 206: 25-35.

Damour, Franck. 2018b. "Les nanotechnologies comme technologie transhumaniste." *L'Homme & la Société* 207: 137-60.

Deguy, Michel. 2001. *Spleen de Paris*. Paris: Galilée.

Deguy, Michel. 2017. *Réouverture après travaux*. Paris: Galilée.

Deleuze, Gilles. (1990) 1992. "Postscriptum on the Societies of Control." *October* 59: 3-7.

Derrida, Jacques, and Bernard Stiegler. 2002. *Echographies of Television: Filmed Interviews*, Translated by Jennifer Bajorek. Cambridge, UK: Polity.

Desmond-Hellman, Susan. 2012. "Toward Precision Medicine: A New Social Contract?" *Science Translational Medicine* 4 (129). https://www.science.org/doi/10.1126/scitranslmed.3003473.

Dewey, John. 1939. *Theory of Valuation*. Chicago, IL: University of Chicago Press.

Dignum, Virginia. 2019. *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way*. Cham, CH: Springer.

Doubleday, Robert. 2007. "The Laboratory Revisited: Academic Science and the Responsible Governance of Nanotechnology." *NanoEthics* 1: 167-76.

Drexler, K. Eric. 1986. *Engines of Creation: The Coming Era of Nanotechnology*. Garden City, NY: Anchor Press/Doubleday.

Dubois, Frédéric. 2016a. "Les nanotechnologies à travers l'imaginaire collectif." Unpublished report prepared for the Musée de la Civilisation (Québec, Canada).

Dubois, Frédéric. 2016b. "Between Science and Fiction – Nanotechnology Shift in American Comic Books." In Bowman et al. 2016.

Ducournau, Pascal, and Claire Beaudevin. 2011. "Génétique en ligne: Déterritorialisation des régulations de santé publique et formes de développement commercial." *Anthropologie & Santé* 3. https://journals.openedition.org/anthropologiesante/777.

Duportail, Judith. 2019. *L'amour sous algorithme*. Paris: Goutte d'or.

Dupuy, Jean-Pierre 2007. "Some Pitfalls in the Philosophical Foundations of Nanoethics." *The Journal of Medicine and Philosophy* 32 (3): 237-61.

Dupuy, Jean-Pierre. (1994) 2000. *The Mechanization of the Mind: On the Origins of Cognitive Sciences*. Translated by M. B. DeBevoise. Princeton, NJ: Princeton University Press.

Ebbesen, Mette, Svend Andersen, and Flemming Besenbacher. 2006. "Ethics in Nanotechnology: Starting from Scratch?" *Bulletin of Science, Technology & Society* 26 (6): 451-62.

Edelman, Nicole. 2002. "Spirites et neurologues face à l'occulte (1870-1890): Une particularité française?" In Bernadette Bensaude-Vincent, ed., *Des savants face à l'occulte: 1870-1940*, 85-104. Paris: La Découverte.

Ekman, Paul, and Wallace V. Friesen. 1971. "Constants Across Cultures in the Face and Emotions." *Journal of Personality and Social Psychology* 17: 124-29.

Ekman, Paul. 1972. *Emotion in the Human Face: Guide-lines for Research and an Integration of Findings*. New York: Pergamon Press.

Ekman, Paul. 1975. *Unmasking the Face*. Englewood Cliffs, NJ: Prentice-Hall.

Elder, Alexis M. 2018. *Friendship, Robots, and Social Media: False Friends and Second Selves*. New York: Routledge.

Ellul, Jacques. (1954) 1964. *The Technological Society*. Translated by John Wilkinson. New York: Vintage Books.

Ellul, Jacques. (1977) 1980. *The Technological System*. Translated by Joachim Neugroschel. New York: Continuum.

Ellul, Jacques. (1988) 1990. *The Technological Bluff*. Translated by Geoffrey W. Bromiley. Grand Rapids, MI: W. B. Eerdmans.

Ettinger, Robert C. W. 1964. *The Prospect of Immortality*. Garden City, NY: Doubleday.

Eubanks, Virginia. 2018. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York: St. Martin's Press.

European Parliament. 2016. "Automated Vehicles in the EU." *EPRS*, January. https://www.europarl.europa.eu/RegData/etudes/BRIE/2016/573902/EPRS_BRI(2016)573902_EN.pdf.

Evans, Claire Lisa. 2018. *Broad Band: The Untold Story of the Women Who Made the Internet*. New York: Port Folio/Penguin.

Ezike, Richard, Jeremy Martin, Katherine Catalono, and Jesse Cohn. 2019. "Where Are Self-Driving Cars Taking Us? Pivotal Choices That Will Shape DC's Transportation Future." *Union of Concerned Scientists*, February. https://www.ucsusa.org/sites/default/files/attach/2019/02/Where-Are-Self-Driving-Cars-Taking-Us-web.pdf.

Fadeel, Bengt, Neus Feliu, Carmen Vogt, Abuelmagd M. Abdelmonem, and Wolfgang J. Parak. 2013. "Bridge over Troubled Waters: Understanding the Synthetic and Biological Identities of Engineered Nanomaterials." *Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology* 5: 111-29.

Feller, Gavin Stuart. 2007. "Media as Compromise: A Cultural History of Mormonism and New Communication Technology in Twentieth-Century America." Doctoral thesis, University of Iowa.

Felt, Ulrike, and Ruth Müller. 2011. "Tentative (Id)entities: On Technopolitical Cultures and the Experiencing of Genetic Testing." *Biosocieties* 6: 342-63.

Ferrarese, Estelle, and Sandra Laugier, eds. 2018. *Formes de vie*. Paris: CNRS.

Ferrarese, Estelle, and Sandra Laugier. 2018. "Formes de vie: Concept et critique pour le XXIe siècle." In Estelle Ferrarese and Sandra Laugier, eds., *Formes de vie*, 11-21. Paris: CNRS.

Fessler, Leah. 2017. "We Tested Bots like Siri and Alexa to See Who Would Stand Up to Sexual Harassment." *Quartz*, February 22. https://qz.com/911681/.

Feynman, Richard. 1959. "There's Plenty of Room at the Bottom: An Invitation to Enter a New Field of Physics." Lecture, American Physical Society, December 29. https://web.pa.msu.edu/people/yang/RFeynman_plentySpace.pdf.

Fischinger, David, Peter Einramhof, Walter Wohlkinger et al. 2013. "HOBBIT – The Mutual Care Robot." Conference ASROB-2013. https://www.researchgate.net/publication/260361825_HOBBIT_-_The_Mutual_Care_Robot.

Fisher, Berenice, and Joan Tronto. 1990. "Toward a Feminist Theory of Caring." In Emily K. Abel and Margaret K. Nelson, eds., *Circles of Care: Work and Identity in Women's Lives*, 36-40. Albany: State University of New York Press.

Floyd, Juliet. 2019. "La quête culturelle: Revisiter le test Turing." In Vanessa Nurock, ed., *L'intelligence artificielle: Enjeux éthiques et politiques*, 15-30. Paris: Presses Universitaires de France, coll. *Cités*.

Floyd, Juliet. 2022. "Revisiting the Turing Test: Humans, Machines, and Phraseology." Lecture, Boston University, November 22. https://www.youtube.com/watch?v=wN4lT1xm4FU.

Foot, Philippa. 2002 (1967). "The Problem of Abortion and the Doctrine of Double Effect." In Philippa Foot, *Virtues and Vices and Other Essays in Moral Philosophy*, 17-33. Oxford, UK: Clarendon.

Foucault, Michel. (1997) 2003. *"Society Must Be Defended."* Translated by David Macey. New York: Picador.

Foucault, Michel. 1988. In Luther H. Martin, Huck Gutman, and Patrick H. Hutton, eds., *Technologies of the Self: A Seminar with Michel Foucault*. Amherst: University of Massachusetts Press.

Frank, Lone. 2011 (2010). *My Beautiful Genome: Exposing our Genetic Future, One Quirk at a Time*. Translated by Russell Dees. Oxford, UK: Oneworld.

Franz, Nina. 2017. "Targeted Killing and Pattern-of-Life Analysis: Weaponised Media." *Media, Culture & Society* 39 (1): 111-21.

Friedman, Andrew L. 1977. *Industry and Labour: Class Struggle at Work and Monopoly Capitalism*. London: Macmillan.

Frith, Ute. 1989. *Autism: Explaining the Enigma*. Oxford, UK: Basil Blackwell.

Gabor, Dennis. (1963) 1964. *Inventing the Future*. New York: Knopf.

Gaille, Marie. 2019. "Pour un nouveau 'code de Nuremberg': De quelques enjeux contemporains du consentement." *Médecine/Sciences* 39 (8-9): 603-4.

Ganascia, Jean-Gabriel. 2009. *Voir et pouvoir: Qui nous surveille?* Paris: Le Pommier.

Gannett, Lisa. 2016. "The Human Genome Project." *The Stanford Encyclopedia of Philosophy*. Winter 2016 Edition. Edited by Edward N. Zalta. https://plato.stanford.edu/archives/win2016/entries/human-genome/.

Gefen, Alexandre, and Sandra Laugier, eds. 2020. *Le pouvoir des liens faibles*. Paris: CNRS.

Gerdes, Justin. 2018. "Not So Fast. Fully Autonomous Vehicles Are More Than a Decade Away, Experts Say." GTM, February 16. https://www.greentechmedia.com/articles/read/fully-autonomous-vehicles-decade-away-experts#gs.Eii7I7I.

Gilligan, Carol, and Naomi Snider. 2018. *Why Does Patriarchy Persist?* Cambridge, UK: Polity.

Gilligan, Carol. 1982. *In a Different Voice: Psychological Theory and Women's Development*. Cambridge, MA: Harvard University Press.

Gilligan, Carol. 2012. "Looking Back to Look Forward: Revisiting In a Different Voice." https://classics-at.chs.harvard.edu/classics9-carol-gilligan-looking-back-to-look-forward-revisiting-in-a-different-voice/.

Gloaguen, Laurent. 2019. "Lovot, le petit robot à aimer." *Brèves technos*, January 10. https://www.spiria.com/fr/blogue/breves-technos/lovot-le-petit-robot-a-aimer/.

Gonzales, Moises, and Enrique R. Lamadrid, eds. 2019. *Nación Genízara: Ethnogenesis, Place, and Identity in New Mexico*. Albuquerque: University of New Mexico Press.

Gourarier, Mélanie. 2017. "Faire la frontière dans les murs du laboratoire: Destins migratoires et usages de d'ADN aux États-Unis." *Genèses* 108: 48-68.

Granovetter, Mark S. 1973. "The Strength of Weak Ties." *American Journal of Sociology* 78 (6): 1360-80.

Greenspan, Patricia S. 1983. "Moral Dilemmas and Guilt." *Philosophical Studies* 43 (1): 117-25.

Gregory, Derek. 2011. "From a View to a Kill: Drones and Late Modern Warfare." *Theory, Culture & Society* 28 (7-8): 118-215.

Gregory, Derek. 2014. "Drone Geographies." *Radical Philosophy* 183 (3): 7-19.

Grossman, David 1995. *On Killing: The Psychological Cost of Learning to Kill in War and Society*. Boston: Little, Brown.

Groves, Christopher. 2014. *Care, Uncertainty and Intergenerational Ethics*. Basingstoke, UK: Palgrave Macmillan.

Guchet, Xavier. 2008. "Nature et artifice dans les nanotechnologies." In Bernadette Bensaude-Vincent, Raphaël Larrère, and Vanessa Nurock, *Nanobio-éthique*, 19-32. Paris: Vuibert.

Guchet, Xavier. 2014. *Philosophie des Nanotechnologies*. Paris: Hermann.

Habermas, Jürgen. (1983) 1990. *Moral Consciousness and Communicative Action*. Translated by Christian Lenhardt and Shierry Weber Nicholsen. Cambridge, MA: MIT Press.

Hale, Andrew, Barry Kirwan, and Urban Kjellén. 2007. "Safe by Design: Where Are We Now?" *Safety Science* 45 (1-2): 305-27.

Harris, Anna, Susan Kelly, and Sally Wyatt. 2016. *Cybergenetics: Health, Genetics, and New Media*. New York: Routledge.

Harris, John. 2019. "The Immoral Machine." *Cambridge Quarterly of Healthcare Ethics* 29 (1): 71-79.

Harvey, Robert. 2016. "Les yeux dans les yeux: La poléthique de Michel Deguy aujourd'hui." *Contemporary French and Francophone Studies* 20 (3): 366-73.

Hedgecoe, Adam M. 2009. "Geneticization: Debates and Controversies." In *Encyclopedia of Life Sciences*. West Sussex, UK: John Wiley & Sons.

Helland, Aasgeor, Peter Wick, Andreas Koehler, Kaspar Schmid, and Claudia Som. 2007. "Reviewing the Environmental and Human Health Knowledge Base of Carbon Nanotubes," *Environmental Health Perspectives* 115 (8): 1125-31.

Herper, Matthew. 2015. "In Big Shift, 23andMe Will Invent Drugs Using Customer Data." *Forbes*, March 12.
https://www.forbes.com/sites/matthewherper/2015/03/12/23andme-enters-the-drug-business-just-as-apple-changes-it/.

Heymans, Gerard. 1912. "III. *Les 'deux Mémoires'* de M. Bergson." *L'Année psychologique* 19: 66-74.

Hogarth, Stuart, and Paula Saukko. 2017. "A Market in the Making: The Past, Present and Future of Direct-to-Consumer Genomics." *New Genetics and Society* 36 (3): 197-208.

Holger, Dieter. 2018. "Ancestry DNA Review: The Largest DNA Database for Finding Relatives and Heritage." *PCWorld*, November 2018. https://www.pcworld.com/article/3302381/ancestrydna-review.html/.

Imai, Kenta, Larry J. Kricka, and Paolo Fortina. 2010. "Concordance Study of 3 Direct-to-Consumer Genetic-testing Services." *Clinical Chemistry* 57 (3): 518-21.

Inside Out: The People's Art Project. N.d. "Not a Bug Splat." https://www.insideoutproject.net/en/group-actions/pakistan-undisclosed-location.

Jaeggi, Rahel. (2014) 2018. *Critique of Forms of Life*. Translated by Ciaran Cronin. Cambridge, MA: Harvard University Press.

Jaffro, Laurent. 2007. "L'exercice moral est-il assimilable à une technique? Une critique de Michel Foucault." *Revue d'Auvergne* 585: 73-96.

Janssens, Cecile, and Peter Kraft. 2012. "Research Conducted Using Data Obtained through Online Communities: Ethical Implications of Methodological Limitations." *PLoS Medicine* 9 (10), October 23. https://doi.org/10.1371/journal.pmed.1001328.

Janssens, Cecile. 2018. "Opinion: No, the FDA Didn't Really Approve 23andMe's *BRCA* Test." *The Scientist,* March 19. https://www.the-scientist.com/news-opinion/opinion-no-fda-didnt-really-approve-23andmes-brca-test-29939.

Jobling, Mark A., Rita Rasteiro, and Jon H. Wetton. 2016. "In the Blood: The Myth and Reality of Genetic Markers of Identity." *Ethnic and Racial Studies* 39 (2): 142-61.

John, Nicholas A. 2013a. "The Social Logics of Sharing." *The Communication Review* 16 (3): 113-31.

John, Nicholas A. 2013b. "Sharing and Web 2.0: The Emergence of a Keyword." *New Media & Society* 15 (2): 167-82.

John, Nicholas A. 2017. *The Age of Sharing*. Malden, MA: Polity.

Joy, Bill. 2000. "Why the Future Doesn't Need Us." *Wired*, April 1. https://www.wired.com/2000/04/joy-2

Kalokairinou, Louiza, Heidi Carmen Howard, Santa Slokenberga, et al. 2018. "Legislation of Direct-to Consumer Genetic Testing in Europe: A Fragmented Regulatory Landscape." *Journal of Community Genetics* 9 (2): 117-32.

Katz, James A., and Mark A. Aakhus. 2002. "Conclusion: Making Meaning of Mobiles – A Theory of Apparatgeist." In James E. Katz and Mark A. Aakhus, *Perpetual Contact: Mobile Communication, Private Talk, Public Performance*, 301-20. Cambridge, UK: Cambridge University Press.

Katz, James E., ed. 2003. *Machines That Become Us: The Social Context of Personal Communication Technology*. New Brunswick, NJ: Transaction Publishers.

Katz, James. 1999. *Connections: Social and Cultural Studies of the Telephone in American Life*. New Brunswick, NJ: Transaction.

Keiper, Adam. 2007. "Nanoethics as a Discipline?" *The New Atlantis* 16: 55-67.

Kelty, Christopher. 2009. "Beyond Implications and Applications: The Story of 'Safety by Design.'" *NanoEthics* 3: 79-96.

Kerr, Anne, Rosemary L. Hill, and Christopher Till. 2018. "The Limits of Responsible Innovation: Exploring Care, Vulnerability and Precision Medicine." *Technology and Society* 52 (February): 24-31.

Kissinger, Henry. 2018. "How the Enlightenment Ends. *The Atlantic* (June 1 2018): 11-14. https://www.theatlantic.com/magazine/archive/2018/06/henry-kissinger-ai-could-mean-the-end-of-human-history/559124/.

Kohlberg, Lawrence. 1981. *The Philosophy of Moral Development*. In *Essays on Moral Development*, vol. 1. San Francisco, CA: Harper & Row.

Konrath, Sarah H., Edward H. O'Brien, and Courtney Hsing. 2011. "Changes in Dispositional Empathy in American College Students Over Time: A Meta-Analysis." *Personality and Social Psychology Review* 15 (2): 180-98.

Lacour, Stephanie, and Daniela Piana. 2019. "Faites entrer les algorithmes! Regards critiques sur la 'justice prédictive.'" In Vanessa Nurock, ed., *L'intelligence artificielle: Enjeux éthiques et* politiques, 47-60. Coll. Cités. Paris: Presses Universitaires de France.

Laestadius, Linnea, Jennifer Rich, and Paul Auer. 2017. "All Your Data (Effectively) Belong to Us: Data Practices Among Direct-to-Consumers Genetic Testing Firms." *Genetics in Medicine* 19: 523-52.

Lafontaine, Céline. 2014. *Le corps-marché: La marchandisation de la vie humaine à l'ere de la bioéconomie*. Paris: Seuil.

Lammer, Lara, Andreas Huber, Wolfgang Zagler, and Markus Vincze. 2011. "Mutual-Care: Users Will Love Their Imperfect Social Assistive Robots." In *Work-in-Progress Proceedings of the International Conference on Social Robotics*, 24-25. http://hobbit.acin.tuwien.ac.at/publications/mutual_care_ISCR2011_WIP_camera_ready.pdf.

Lane, Harlan. 1976. *The Wild Boy of Aveyron*. Cambridge, MA: Harvard University Press.

Lanoix, Monique. 2019. "Un amour de robot: Robot émotionnel et travail d'aidant." In Soheil Kash and Marie-Hélene Parizeau, eds., *La société robotisée: Enjeux éthiques et politiques*, 259-279. Quebec City, Canada: Presses de l'Université Laval.

Larrère, Catherine, and Raphaël Larrère. 2017. *Bulles technologiques*. Marseille: Wildproject.

Larson, Jeff, Surya Matta, Lauren Kirchner, and Julia Angwin. 2016. "How We Analyzed the COMPAS Recidivism Algorithm." *ProPublica*, May 23. https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm.

Latour, Bruno. (1996) 2010. "On the Cult of the Factish Gods." Translated by Catherine Porter and Heather McLean. In Bruno Latour 2010, *On the Modern Cult of the Factish Gods*, 1-66. Durham, NC: Duke University Press.

Latour, Bruno. 2000. "Factures/fractures. De la notion de réseau à celle d'attachement." In André Micoud and Michel Peroni, eds., *Ce qui nous relie*, 189-207. La Tour-d'Aigues: Éditions de l'Aube.

Latour, Bruno. 2008. "Pour un dialogue entre science politique et science studies." *Revue française de science politique* 57: 657-78.

Latour, Bruno. 2011. "Love Your Monsters." In Michael Shellenberger and Ted Nordhaus, eds., *Love Your Monsters. Postenvironmentalism and the Anthropocene*, 17-25. Oakland, CA: Breakthrough Institute.

Latour, Bruno. 2007. *Changer de société, refaire de la sociologie*. Paris: La Découverte.

Laugier, Sandra, and Patricia Paperman, eds. 2006. *Le souci des autres: Éthique et politique du care*. Paris: EHESS.

Laugier, Sandra. 2005. "L'importance de l'importance: Expérience, pragmatisme, transcendantalisme." *Multitudes* 23 (4): 153-67.

Laugier, Sandra. 2009. "Le sujet du care: Vulnérabilité et expression ordinaire." In Pascale Molinié, Sandra Laugier, and Patricia Paperman, *Qu'est-ce que le care? Souci des autres, sensibilité, responsabilité*, 159-200. Paris: Payot.

Laugier, Sandra. 2011. "Le care, le souci du détail et la vulnérabilité du réel. " *Raison publique: La revue des humanités politiques*. April 8.
https://raison-publique.fr/707/.

Laurent, Brice. 2013. "Les espaces politiques des substances chimiques." *Revue d'anthropologie des connaissances* 7: 1-24.

Lee, Peter. 2019. *Reaper Force – Inside Britain's Drone Wars*. London: John Blake.

Lee, Peter. 2020. "Iran Attack: How Reaper Drones Really Carry Out Airstrikes." *The Conversation*, January 7.
https://theconversation.com/iran-attack-how-reaper-drones-really-carry-out-airstrikes-129411.

Leichter-Flack, Frédérique. 2015. *Qui vivra, qui mourra? Quand on ne peut pas sauver tout le monde*. Paris: Albin Michel.

Levinas, Emmanuel. (1961) 1969. *Totality and Infinity: An Essay on Exteriority*. Translated by Alphonso Lingis. Pittsburgh, PA: Duquesne University Press.

Levy, David N. L. 2007. *Love and Sex with Robots: The Evolution of Human-Robot Relations*. New York: HarperCollins.

Li, Sheng, Tracy Hulderman, Rebecca Salmen, et al. 2007. "Cardiovascular Effects of Pulmonary Exposure to Single-Wall Carbon Nanotubes." *Environmental Health Perspectives* 115 (3): 377-82.

Lin, Patrick. 2014. "Here's a Terrible Idea: Robot Cars With Adjustable Ethics Settings." *Wired*, August 18.
https://www.wired.com/2014/08/heres-a-terrible-idea-robot-cars.with-adjustable-ethics-settings/.

Lippman, Abby. 1991. "Prenatal Genetic Testing and Screening: Constructing Needs and Reinforcing Inequities." *American Journal of Law and Medicine* 17 (1-2): 15-50.

Lippman, Abby. 1992. "Led (Astray) by Genetic Maps: The Cartography of the Human Genome and Health Care." *Social Science and Medicine* 35 (12): 1469-76.

Lippman, Abby. 1993. "Prenatal Genetic Testing and Geneticization: Mother Matters for All." *Fetal Diagnosis and Therapy* 8, Supp. 1: 175-88.

Lippman, Abby. 1998. "The Politics of Health: Geneticization Versus Health Promotion," In Susan Sherwin, ed., *The Politics of Women's Health: Exploring Agency and Autonomy*, 64-82. Philadelphia: Temple University Press.

Locke, John. (1690) 2001. *An Essay Concerning Human Understanding*. Kitchener, Ontario: Batoche Books.

Loeve, Sacha. 2009. "Le concept de technologie à l'échelle des molécules-machines: Philosophie des techniques à l'usage des citoyens du nanomonde." Doctoral thesis, Université Paris-Ouest Nanterre.

Loideain, Nóra Ní, and Rachel Adams. 2018. "From Alexa to Siri and the GDPR: The Gendering of Virtual Personal Assistants and the Role of EU Data Protection Law." King's College London, Dickson Poon School of Law. Legal Studies Research Paper Series, November 9.
https://ssrn.com/abstract=3281807.

Lowry, Gregory, Kelvin B. Gregory, Simon C. Apte, and Jamie R. Lead. 2012. "Transformations of Nanomaterials in the Environment." *Environmental Science and Technology* 46: 6893-99.

Lugano, Giuseppe, Martin Hudak, Matúš Ivančo, and Tomas Lovecek. 2017. "From the Mind to the Cloud: Personal Data in the Age of the Internet of Things." In Yuefang Zhou and Martin H. Fischer, eds., 2019, *AI Love You: Developments in Human-Robot Intimate Relationships*, 111-31. Cham, CH: Springer.

Lyons, David. 1965. *Forms and Limits of Utilitarianism*. Oxford, UK: Clarendon Press.

Maestrutti, Marina. 2011. *Imaginaires des nanotechnologies: Mythes et fictions de l'infiniment petit*. Paris: Vuibert.

Mann, Steve, James Nolan, and Barry Wellman. 2003. "Sousveillance: Inventing and Using Wearable Computing Devices." *Surveillance & Society* 1: 331-35.

Marcus, Gary, and Ernest Davis. 2019. *Rebooting AI: Building Artificial Intelligence We Can Trust*. New York: Pantheon Books.

Marx, Karl. (1847) 1995. *The Poverty of Philosophy*. Translated by Harry Quelch. Amherst, NY: Prometheus Books.

Matchar, Emily. 2011. "Why I Can't Stop Reading Mormon Housewife Blogs: I'm a Young Feminist Atheist Who Can't Bake a Cupcake. Why Am I Addicted to the Shiny, Happy Lives of These Women?"
https://www.salon.com/2011/01/15/feminist_obsessed_with_mormon_blogs/.

Mayer-Schönberger, Viktor, and Kenneth Cukier. 2013. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Boston: Houghton-Mifflin.

McCarthy, Elise, and Christopher Kelty. 2010. "Responsibility and Nanotechnology." *Social Studies of Science* 40: 405-32.

McMillan, Graeme. 2011. "It's Not You, It's It: Voice Recognition Doesn't Recognize Women." *Time*, June 1.
http://techland.time.com/2011/06/01/its-not-you-its-it-voice-recognition-doesnt-recognize-women/.

Miailhe, Nicolas. 2018. "Géopolitique de l'Intelligence artificielle: Le retour des Empires?" *Politique Étrangère*, no. 3: 105-17.

Mikhail, John M. 2011. *Elements of Moral Cognition: Rawls' Linguistic Analogy and the Cognitive Science of Moral and Legal Judgment*. Cambridge, UK: Cambridge University Press.

Mission de préfiguration. 2018. "Rapport Health Data Hub." December 10, 4. https:// solidarites-sante.gouv.fr/ministere/documentation-et-publications-officielles/rapports/sante/article/rapport-health-data-hub-mission-de-prefiguration.

Mitrou, Lilian, Miltiadis Kandias, Vasilis Stavrou, and Dimitris Gritzalis. 2014. "Social Media Profiling: A Panopticon or Omniopticon Tool?" In *Proceedings of the 6th Biannual Surveillance and Society Conference*, April 14-26, Barcelona.

Mody, Cyrus M. 2004. "Small but Determined: Technological Determinism in Nanoscience." *Hyle – An International Journal for Philosophy of Chemistry* 10 (2): 99-128.

Mol, Annemarie. 2008. *The Logic of Care : Health and the Problem of Patient Choice*. Translated from the Dutch. London: Routledge.

Molinier, Pascale. 2018. *Le Care Monde*. Paris: Presses universitaires de France.

Molteni, Megan. 2019. "Not Everyone on 23andMe Will Get the Latest Gene Chip Updates." *Wired*, July 16.
https://www.wired.com/story/not-everyone-on-23andme-will-get-the-latest-gene-chip-updates/.

Moor, James H. 2006. "The Nature, Importance and Difficulty of Machine Ethics." *IEEE Intelligent Systems* 21: 18-21.

Moore, G. E. (1903) 1993. *Principia Ethica*. Cambridge, UK: Cambridge University Press.

Mori, Masahiro. (1970) 2012. "The Uncanny Valley [From the Field]." Translated by Karl F. MacDorman and Nomi Kageki. *IEEE Robotics & Automation Magazine* 19 (2): 98-100.

Murry, John. 1999. "Owning Genes: Disputes Involving DNA Sequence Patents." *Chicago-Kent Law Review* 75: 231-57.

Nass, Clifford, and Corina Yen. 2010. *The Man Who Lied To His Laptop: What Machines Teach Us About Human Relationships*. New York: Current.

National Technology Initiative. N.d. "What's So Special about the Nanoscale?" https://www.nano.gov/nanotech-101/special.

Nelson, Alondra. 2008. "Bio Science: Genetic Genealogy Testing and the Pursuit of African Ancestry." *Social Studies of Science* 38 (5): 759-83.

Nietzsche, Friedrich. (1887) 2003. Translated by Horace B. Samuel. Mineola, NY: Dover Publications.

Nordmann, Alfred. 2007. "If and Then: A Critique of Speculative NanoEthics." *Nanoethics* 1: 31-46.

Nurock, Vanessa. 2008. *Rawls, pour une démocratie juste?* Paris: Michalon.

Nurock, Vanessa. 2010. "Nanoethics: Ethics For, From, or With Nanotechnologies?" *Hyle* 16: 31-42.

Nurock, Vanessa, and Nathalie Panissal. 2016. "Teaching a 'Care' Approach to Nanotechnologies." In Bowman et al., 125-37.

Nurock, Vanessa. 2019a. "Généalogie de la morale automatisée." In Marie-Hélène Parizeau and Soheil Kash, eds., *Robots et sociétés: Enjeux éthiques et politiques*, 31-50. Québec: Presses de l'Université Laval.

Nurock, Vanessa. 2019b. “L’intelligence artificielle a-t-elle un genre?” In Vanessa Nurock, ed., 2019, 61-74.

Nurock, Vanessa. 2019c. “Le *care* de la nanoéthique: Repenser la question des frontières.” *Ethica* 22: 149-65.

Nurock, Vanessa, ed., 2019. *L’intelligence artificielle: Enjeux éthiques et politiques*. Coll. Cités. Paris: Presses Universitaires de France.

Nurock, Vanessa, Raja Chatila, and Marie-Hélène Parizeau. 2021. “What Does ‘Ethical by Design’ Mean?” In Bertrand Braunschweig and Malik Ghallab, eds., *Reflections on Artificial Intelligence for Humanity,* 171-90. Cham, CH: Springer.

Nurock, Vanessa. 2021. “The Artificialist Fallacy.” In James E. Katz, Juliet Floyd, and Katie Schiepers, eds., *Perceiving the Future through New Communication Technologies: Robots, AI and Everyday Life*, 75-87. London: Palgrave Macmillan.

Nurock, Vanessa, 2021. “24h chrono, une série trop morale?” In Sylvie Allouche, ed., *24 heures chrono, naissance du genre sécuritaire?* Open archive J. Vrin. https://archive-ouverte.vrin.fr/item/nurock_24_heures_chrono_une_serie_trop_morale_2021.

O’Neal, Cathy. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown.

O’Neil, Cathy. 2016. *Weapons of Mass Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown,

O’Riordan, Kate. 2010. *The Genome Incorporated: Constructing Biodigital Identity*. Farnham, UK: Ashgate.

Ogien, Albert. 1995. *L’esprit gestionnaire: Une analyse de l’air du temps*. Paris: EHESS.

Ogien, Albert. 2009. “L’hôpital saisi par la quantification: Une analyse de l’usage gestionnaire de la notion de qualité.” *Sciences de la société* 76: 31-51.

Ogien, Albert. 2010. “La valeur sociale du chiffre: La quantification de l’action publique entre performance et démocratie.” *Revue française de socio-économie* 5: 19-40.

Opam, Kwame. 2014. “Art Collective Aims to Humanize Drone Casualties with Massive Portraits of Victims.” *The Verge,* April 17. https://www.theverge.com/2014/4/7/5589920/art-collective-aims-to-humanize-drone-casualties-with-massive.

ORi. 2014a. “If Death by Autonomous Car Is Unavoidable, Who Should Die? Reader Poll Results,” Robohub.org, June 23. https://openroboethics.org/results-if-a-death-by-an-autonomous-car-is-unavoidable-who-should-die/.

ORi. 2014b. “My (Autonomous) Car, My Safety: Results from Our Reader Poll.” Robohub.org. June 30. https://robohub.org/my-autonomous-car-my-safety-results-from-our-reader-poll/.

Panofsky, Aaron, and Joan Donovan. 2019. “Genetic Ancestry Testing among White Nationalists: From Identity Repair to Citizen Science.” *Social Studies of Science* 49 (5): 653-81.

Paperman, Patricia, and Pascale Molinier, eds. 2013. *Contre l’indifférence des privilégiés: À quoi sert le care*. Paris: Payot.

Pasquale, Frank. 2015. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, MA: Harvard University Press.

Pavie, Xavier. 2014. "The Importance of Responsible Innovation and the Necessity of 'Innovation-Care.'" *Philosophy of Management* 13: 21-42.

Pavie, Xavier. 2020. *Critical Philosophy of Innovation and the Innovator.* Hoboken, NJ: John Wiley & Sons.

Pflanzer, Lydia Ramsey. 2016. "I Took a $30 Test that Told Me If I Had 'Superhero" Genes – and It Was By Far the Most Fun Test I've Taken." *Business Insider*, October 19.
https://www.businessinsider.com/a-30-genetics-test-for-superhero-genes-2016-8.

Philips, Andelka M. 2016. "Only a Click Away – DTC Genetics for Ancestry, Health, Love … and More: A View of the Business and Regulatory Landscape." *Applied & Translational Genomics* 8 (March): 16-22.

Piaget, Jean. (1932) 1965. *The Moral Judgment of the Child.* Translated by Marjorie Gabain. London: K. Paul Trench, Truber & Co.

Pietsch, Bryan. 2021. "2 Killed in Driverless Tesla Car Crash, Officials Say."
https://www.nytimes.com/2021/04/18/business/tesla-fatal-crash-texas.html.

Pinel, Clémence, Barbara Prainsack, and Christopher McKevitt. 2020. "Caring for Data: Value Creation in a Data-intensive Research Laboratory," *Social Studies of Science* 50 (2): 175-197.

Pinker, Steven. 2009. "My Genome, My Self." *New York Times Magazine,* January 7.
https://www.nytimes.com/2009/01/11/magazine/11Genome-t.html.

Poel, Ibo van de, and Zoë Robaey. 2017. "Safe-by-Design: From Safety to Responsibility." *Nanoethics* 11: 297-306.

Poel, Ibo van de. 2008. "How Should We Do Nanoethics? A Network Approach for Discerning Ethical Issues in Nanotechnology." *NanoEthics* 2 (1): 25-38.

Poel, Ibo van de. 2017. "Society as a Laboratory to Experiment with New Technologies." In Diana M. Bowman, Elen Stokes, and Arie Rip, eds., *Embedding New Technologies into Society: A Regulatory, Ethical and Societal Perspective*, 61-87. Singapore: Pan Stanford.

Popper, Karl. (1935) 1959. *The Logic of Scientific Discovery.* Translated by the author. London: Hutchinson.

Post, Stephen G. 2008. "Updating the Helper Therapy Principle: Recovery Rates for Alcoholism Doubled for Those Helping Other Alcoholics." September 3.
https://www.psychologytoday.com/us/blog/the-joy-giving/200809/updating-the-helper-therapy-principle.

Prainsack, Barbara, and Yael Hashiloni-Dolev. 2009. *Religion and Nationhood: Collective Identities and the New Genetics*. In Paul Atkinson, Peter Glasner, and Margaret Lock, eds., *The Handbook of Genetics & Society: Mapping the New Genomic Era*, 404-21. London: Routledge.

Prainsack, Barbara, Silke Schicktanz, and Gabriele Werner-Felmaye, eds. 2014. *Genetics as Social Practice: Transdisciplinary Views on Science and Culture*, 47-164. Burlington, VT: Ashgate.

Prainsack, Barbara. 2014, "Understanding Participation: The 'Citizen Science' of Genetics." In Prainsack et al. 2014, 47-164.

Prainsack, Barbara. 2017. "Research for Personalised Medicine: Time for Solidarity." *Medicine and Law* 36 (1): 87-98.

Prainsack, Barbara. 2017b. *Personalized Medicine: Empowered Patients in the 21st Century?* New York: New York University Press.

Pugliese, Joseph. 2013. *State Violence and the Execution of Law: Biopolitical Caesurae of Torture, Black Sites, Drones*. New York: Routledge.

Puig de la Bellacasa, Maria. 2017. *Matters of Care: Speculative Ethics in More Than Human Worlds*. Minneapolis: University of Minnesota Press.

Rabinow, Paul, and Gaymon Bennett. 2012. *Designing Human Practices: An Experiment with Synthetic Biology*. Chicago: University of Chicago Press.

Rajan, Kaushik Sunder. 2006. *Biocapital: The Constitution of Postgenomic Life*. Durham, NC: Duke University Press.

Rawls, John. 1971. *A Theory of Justice*. Cambridge, MA: Belknap Press of Harvard University Press.

Reardon, Jenny. 2017. *The Post Genomic Condition: Ethics, Justice, and Knowledge after the Genome*. Chicago: University of Chicago Press.

Reeves, Byron, and Clifford Nass. 1996. *The Media Equation: How People Treat Computers, Televisions and New Media Like Real People and Places*. Cambridge, UK: Cambridge University Press.

Regalado, Antonio. 2019. "More Than 26 Million People Have Taken an At-home Ancestry Test." *MIT Technology Review*, February. https://www.technologyreview.com/s/612880/more-than-26-million-people-have-taken-an-at-home-ancestry-test/.

Rejeski, David. 2008. "Nanotechnology and the Trust Gap." *Nanotechnology Columns*. https://www.nanotech-now.com/columns/?article=109.

Richardson, Kathleen. 2015. "The Asymmetrical 'Relationship': Parallels Between Prostitution and the Development of Sex Robots." *SIGCAS Computers & Society* 45 (3): 290-93.

Richardson, Kathleen. 2018. *Challenging Sociality*. London: Palgrave Macmillan.

Riessman, Frank. 1965. "The 'Helper' Therapy Principle." *Social Work* 10 (2): 27-32.

Rinard, Ruth G. 1996. "Technology, Deskilling and Nurses: The Impact of the Technologically Changing Environment." *Advanced Nursing Science Journal* 18 (4): 60-69.

Roberts, J. Scott, Michele C. Gornick, Deanna Alexis Carere, Wendy R. Uhlmann, Mack T. Ruffin, and Robert C. Green. 2017. "Direct-to-Consumer Genetic Testing: User Motivations, Decision Making, and Perceived Utility of Results." *Public Health Genomics* 20 (1): 36-45.

Roco, Mihail C., and William Sims Bainbridge. 2003. *Converging Technologies for Improving Human Performance: Nanotechnology, Biotechnology, Information Technology, and Cognitive Science*. Dordrecht/Boston: Kluwer Academic Publishers. http://www.wtec.org/ConvergingTechnologies/Report/NBIC_report.

Romero, Simon. 2018. "Indian Slavery Once Thrived in New Mexico. Latinos Are Finding Family Ties to It." *New York Times*, January 28. https://www.nytimes.com/2018/01/28/us/indian-slaves-genizaros.html.

Rose, Nikolas. 2007. *The Politics of Life Itself: Biomedicine, Power, and Subjectivity in the Twenty-first Century*. Princeton, NJ: Princeton University Press.

Rouvroy, Antoinette. 2009. "Généticisation et responsabilité: Les habits neufs de la gouvernance néolibérale." In Christian Hervé, Michèle Stranton-Jean, Patrick A. Molinari, and Marie-Angèle Grimaud, eds., *Généticisation et responsabilités*, 109-18. Paris: Dalloz.

Royakkers, Lambèr M. M., and Rinie van Est. 2016. *Just Ordinary Robots: Automation from Love to War*. Boca Raton, FL: CRC Press.

Royal Society & Royal Academy of Engineering. 2004. "Nanoscience and Nanotechnologies, Opportunities and Uncertainties." https://royalsociety.org/topics-policy/publications/2004/nanoscience-nanotechnologies/.

Russell, Stuart J. 2019. *Human Compatible: Artificial Intelligence and the Problem of Control*. New York: Viking.

Rutter, Michael, and the English and Romanian Adoptees Study Team. 1998. "Developmental Catch-up, and Deficit, Following Adoption after Severe Global Early Privation." *Journal of Child Psychology and Psychiatry and Allied Disciplines* 39 (4): 465-76.

Saifi, Sophia. 2014. "Not a 'Bug Splat': Artists Give Drone Victims a Face in Pakistan." CNN, April 9. https://edition.cnn.com/2014/04/09/world/asia/pakistan-drones-not-a-bug-splat/.

Sayeed. Sarah. "Bugsplat." *Sojourners*, February 25. https://sojo.net/articles/bugsplat

Schilit, Samantha L. P., and Arielle Schilit Nitenson. 2017. "My Identical Twin Sequenced Our Genome." *Journal of Genetic Counseling* 26 (2): 276-78.

Schiller, Amy, and John McMahon. 2019. "Alexa, Alert Me When the Revolution Comes: Gender, Affect, and Labor in the Age of Home-Based Artificial Intelligence." *New Political Science* 41 (2): 173-91.

Schomberg, René von. 2013. "A Vision of Responsible Research and Innovation." In Richard Owen, John Bessant, and Maggie Heintz, eds., *Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society*, 51-74. Chichester, West Sussex, UK: John Wiley & Sons.

Schramowski, Patrick, Cigdem Turan, Sophie Jentzsch, Constantin Rothkopf, and Kristian Kersting. 2020. "The Moral Choice Machine." *Frontiers in Artificial Intelligence* 3 (36).

Schwartz, Daniel. 2013. "Drone-Speak Lexicon: From 'Bugsplat' to 'Targeted Killing.'" CBC News, February 8. https://www.cbc.ca/news/world/drone-speak-lexicon-from-bugsplat-to-targeted-killing-1.1342966.

Searle, John. 1980. "Minds, Brains, and Programs." *Behavioral and Brain Sciences* 3 (3): 417-24. https://www.cambridge.org/core/journals/behavioral-and-brain-sciences/article/abs/minds-brains-and-programs/DC644B47A4299C637C89772FACC2706A.

Sebbah, François-David. 2010. *Qu'est-ce que la "technoscience"? Une thèse épistémologique ou la fille du diable? Enquête sur les usages de la notion de "technoscience" dans l'espace de la pensée française contemporaine*. Paris: Les Belles Lettres, 2010.

Sen, Amartya. 2009. *The Idea of Justice*, Cambridge, MA: Harvard University Press.

Shelley-Egan, Clare, Diana Bowman, and Douglas K. R. Robinson. 2018. "Devices of Responsibility: Over a Decade of Responsible Research and Innovation Initiatives for Nanotechnologies." *Science and Engineering Ethics* 24: 1719-46.

Singer, Peter W. 2009. *Wired for War: The Robotics Revolution and Conflict in the Twenty-First* Century. New York: Penguin.

Skrabanek, Petr. 1994. *The Death of Humane Medicine and the Rise of Coercive Healthism*. London: Crowley Esmonde.

Smith, Adam. (1759) 2002. *The Theory of Moral Sentiments*. Cambridge, UK: Cambridge University Press.

Solove, Daniel. 2011. *Nothing to Hide: The False Tradeoff between Privacy and Security*. New Haven, CT: Yale University Press.

Sparrow, Robert, and Mark Howard. 2017. "When Human Beings Are Like Drunk Robots: Driverless Vehicles, Ethics, and the Future of Transport." *Transportation Research Part C: Emerging Technologies* 80 (July): 206-15.

Stacy, Christian Plerhoples, and Brady Meixell. 2018. "Self-Driving Cars Could Harm Low-Income People If We Don't Prepare Their Rise." *Urban Institute*, September 14.
https://www.urban.org/urban-wire/self-driving-cars-could-harm-low-income-people-if-we-dont-prepare-their-rise.

Stahelski, Anthony, Amber Anderson, Nicholas Browitt, and Mary Radeke. 2021. "Facial Expressions and Emotion Labels are Separate Initiators of Trait Inferences from the Face." *Frontiers in Psychology* 12. December 8.

Star, Susan Leigh. 1999. "The Ethnography of Infrastructure." *American Behavioral Scientist* 43 (3): 377-391.

Stempsey, William E. 2006. "The Geneticization of Diagnostics." *Medicine, Health Care, and Philosophy* 9 (2): 193-200.

Stiegler, Bernard. 2018. "Les algorithmes et la bêtise artificielle."
https://www.bnf.fr/fr/mediatheque/les-algorithmes-et-la-betise-artificielle.

Styron, William. 1979. *Sophie's Choice*. New York: Random House.

Suard, François. 1994. *Chanson de geste et tradition épique en France au Moyen-Age*. 2nd edition. Caen: Paradigme.

Sung, Ja-Yung, Lan Guo, Rebecca E. Grinter, and Henrik I. Christensen. 2007. "'My Roomba Is Rambo': Intimate Home Appliances." In J. Krumms, G. D. Abowd, A. Seneviratne, and T. Strang, eds., *UbiComp 2007: Ubiquitous Computing, Lecture Notes in Computer Science*, vol. 4717, 145-62. Berlin: Springer.

Swiersta, Tsjalling, and Arie Rip. 2007. "Nano-ethics as NEST-ethics: Patterns of Moral Argumentation about New and Emerging Science and Technology." *Nano-Ethics* 1: 3-20.

Tandy-Connor, Stephany, Jenna Guiltinan, Kate Krempely, et al. 2018. "False-positive Results Released by Direct-to-Consumer Genetic Tests Highlight the Importance of Clinical Confirmation Testing for Appropriate Patient Care." *Journal of Genetic Medicine* 20 (12): 1515-21.

Terranova, Tiziana. 2000. "Free Labor: Producing Culture for the Digital Economy." *Social Text* 18: 33-58.

Thatcher, Margaret. 1980. Press conference. June 25. https://www.margaretthatcher.org/document/104389.

Thomas, Philip S., Bruno Castro da Silva, Andrew G. Barto et al. 2019. "Preventing Undesirable Behavior of Intelligent Machines." *Science* 366 (6468): 999-1004.

Thompson, Dennis F. 1908. "Moral Responsibility of Public Officials: The Problem of Many Hands." *American Political Science Review* 74 (4): 905-16.

Thomson, Judith Jarvis. 1985. "The Trolley Problem." *Yale Law Journal* 94 (6): 1395-1415.

Threlkeld 2018. "Living DNA Announces Move from Illumina Microarray." October 23.
https://casestone.com/threlkeld/blog/114-living-dna-announces-move-from-illumina-microarray.

Timmermans, Stefan, and Mara Buchbinder. 2010. "Patients-in-Waiting: Living Between Sickness and Health in the Genomics Era." *Journal of Health and Social Behavior* 51 (4): 408-23.

Tolmeijer, Suzanne, Markus Kneer, Cristina Sarasua, Markus Christen, and Abraham Bernstein. 2020. "Implementations in Machine Ethics: A Survey." ACM *Computing Surveys* 53 (6): 1-38.

Tranøy, Knut Erik. 1972. "'Ought' Implies 'Can': A Bridge from Fact to Norm (Part 1)." *Ratio* 14: 116-30.

Tremoulet, Patrice, Thomas Seacrist, Chelsea M. Ward McIntosh, and Helen Loeb. 2019. "Transporting Children in Autonomous Vehicles: An Exploratory Study." *Human Factors: The Journal of the Human Factors and Ergonomics Society* 62 (2): 278-87.

Tronto, Joan. 1993. *Moral Boundaries: A Political Argument for an Ethic of Care*. New York: Routledge.

Tronto, Joan. 2012. *Le risque ou le Care?* Paris: Presses universitaires de France.

Tronto, Joan. 2013. *Caring Democracy: Markets, Equality, and Justice*. New York: New York University Press.

Turiel, Elliot. 2002. *The Culture of Morality: Social Development, Context, and Conflict*. Cambridge, UK: Cambridge University Press.

Turing, Alan M. 1948. "Intelligent Machinery."
https://turingarchive.kings.cam.ac.uk/unpublished-manuscripts-and-drafts-amtc/amt-c-11.

Turing, Alan M. 1950. "Computing Machinery and Intelligence." *Mind* 40 (236): 433-60.
https://academic.oup.com/mind/article/LIX/236/433/986238.

Turkle, Sherry. 2011. *Alone Together: Why We Expect More from Technology and Less from Each Other*. New York: Basic Books.

Turkle, Sherry. 2015. *Reclaiming Conversation: The Power of Talk in a Digital Age*. New York: Penguin.

Umoja, Safiya. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: New York University Press.

Vallor, Shannon. 2015. "Moral Deskilling and Upskilling in a New Machine Age: Reflections on the Ambiguous Future of Character." *Philosophy & Technology* 28: 107-24.

Veatch, Robert M. 2009. *Patient, Heal Thyself: How the New Medicine Puts the Patient in Charge*. Oxford, UK: Oxford University Press.

Venter, Craig. 2013. *Life at the Speed of Light: From the Double Helix to the Dawn of Digital Life*. New York: Viking.

Verbeek, Peter-Paul. 2014. "Technology Design as Experimental Ethics." In Simone V. D. Burg and Tsjalling Swierstra, eds., *Ethics on the Laboratory Floor*, 79-96. New York: Palgrave Macmillan.

Vlasits, Anna. 2017. "How 23andMe Won Back the Right to Foretell Your Diseases." *Wired*, April. https://www.wired.com/2017/04/23andme-won-back-right-foretell-diseases/.

Wadhwa, Vivek, and Alex Salkever. 2017. *The Driver in the Driverless Car*. Oakland, CA: Berrett-Koehler Publishers.

Wallace, Susan E., Elli G. Gourna, Viktoriya Nikolova, and Nuala A. Sheehan. 2015. "FamilyTree and Ancestry Inference: Is There a Need for a 'Generational' Consent?" *BMC Medical Ethics* 16 (1): 87. https://doi.org/10.1186/s12910-015-0080-2.

Way, Niobe, Alisha Ali, Carol Gilligan, David E. Kirkland, and Pedro Noguera. 2018. *The Crisis of Connection: Roots, Consequences, and Solutions*. New York: New York University Press.

Way, Niobe. 2011. *Deep Secrets: Boys' Friendships and the Loss of Connection*. Cambridge, MA: Harvard University Press.

Weiner, Kate, Paul Martin, Martin Richards, and Richard Tutton. 2017. "Have We Seen the Geneticisation of Society? Expectations and Evidence." *Sociology of Health and Illness* 39 (7): 989-1004.

West, Mark, Rebecca Kraut, and Chew Han Ei. 2019. "I'd Blush If I Could: Closing Gender Divides in Digital Skills Through Education." UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000367416.page=1.

Widdows, Heather. 2013. *The Connected Self: The Ethics and Governance of the Genetic Individual*. New York: Cambridge University Press.

Wiener, Norbert. (1950) 1954. *The Human Use of Human Beings: Cybernetics and Society*. Boston: Houghton Mifflin.

Wild, John. (1936) 1962. *George Berkeley, A Study of His Life and Philosophy*. New York: Russell & Russell.

Winner, Langdon. 1986. *The Whale and the Reactor: A Search for Limits in an Age of High Technology*. Chicago: University of Chicago Press.

Withers, Paul. 2018. "Robots Take Over: Machine to Run for MAYOR in Japan Pledging 'Fair Opportunities for All.'" *Express*. April 17. https://www.express.co.uk/news/world/947448/robots-japan-tokyo-mayor-artificial-intelligence-ai-news.

Wittgenstein, Ludwig. (1948) 1998. *Culture and Value*. Edited by Georg Henrik von Wright. Translated by Peter Winch. Oxford, UK: Blackwell.

Wolfe, Alexandra. 2014. "Anne Wojcicki's Quest for Better Health Care: The 23andMe CEO on the Promise of Genetics and the Future of Health Care." *Wall Street Journal*, June 27. https://www.wsj.com/articles/anne-wojcickis-quest-for-better-health-care-1403892088.

Woods, Heather Suzanne. 2018. “Asking more of Siri and Alexa: Feminine Persona in Service of Surveillance Capitalism.” *Critical Studies in Media Communication* 35 (4): 1-16.

Worms, Frédéric. 2009. *La philosophie en France au XXe siècle: Moments*. Paris: Gallimard.

Wynsberghe, Aimée van. 2015. *Healthcare Robots: Ethics, Design and Implementation*. Burlington, VT: Ashgate.

Xiang, Chloe. 2022. “The Synthetic Party in Denmark is Dedicated to Following a Platform Churned out by an AI, and Its Public Face Is a Chatbot Named Leader Lars.” October 13. https://www.vice.com/en/article/jgpb3p/this-danish-political-party-is-led-by-an-ai.

Young, Iris. 2006. “Katrina, Too Much Blame, Not Enough Responsibility.” *Dissent* 53 (1): 41-46.

Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: PublicAffairs.