CHAPTER 4

Governmental and Technological Paternalism

> The goal is to enable Google users to be able to ask the question [sic] such as "What shall I do tomorrow?" and "What job shall I take?"
>
> *Eric Schmidt, Google's former CEO*[1]

> Once Google, Facebook and other algorithms become all-knowing oracles, they may well evolve into agents and ultimately into sovereigns.
>
> *Yatul Noah Harari*[2]

The argument that intuition is systematically biased is not confined to academic domains. It has led to a new governmental philosophy called *libertarian paternalism*, a variant of soft paternalism that uses strategies known as *nudges* to influence people's decisions. Its key proponents are the same influential thinkers who seek to demonstrate the failure of intuition. A nudge is a tool for steering people without using incentives, which are the stock-in-trade of economic theory, and without forbidding behavior, which is the practice of hard paternalism. Spearheaded by the UK and USA in 2010, governments across the world have set up nudge units, also known as behavioral insights teams. In a survey from the Organisation for Economic Co-operation and Development (OECD), most governments motivate their programs by reference to individual cognitive biases and people's use of heuristics.[3]

Paternalism (etymologically rooted in the Latin word *pater*) is the view that a select group of people is entitled to rule over others the way fathers traditionally ruled over children. The chosen group might be aristocrats, the wealthy, or religious leaders. Those under their authority might be poor, people of color, or those with little formal education. Over millennia, aristocrats ruled over their subjects and men ruled over women. Athens' early democracy was a rare and partial exception, where every

[1] Daniel & Palmer (2007). [2] Harari (2017), p. 397. [3] OECD (2017).

man – although not women or slaves – was considered equal. Aristotle advocated for the subjugation of women, slaves, animals, and plant life, arguing that reason was naturally endowed only to men of good birth.

As discussed in Chapter 3, people's intuition has become seen as untrustworthy and biases are said to be as stable as visual illusions, implying that there is little hope of improving people's cognitive capacities. To add insult to injury, societal problems such as poverty, obesity, and unhappiness have been attributed to people's lack of rationality and willpower.[4]

The Revival of Paternalism

Governments can serve and protect citizens not only by setting economic incentives but also by pursuing three other strategies:

- **Hard paternalism (no choice): Control citizens for their own benefit.** Governments decide what is best for their citizens and set up laws and rules they are coerced into following, such as tax rates, seatbelt laws, and speed limits. Strong paternalism means restricting choice by taking options away.
- **Libertarian paternalism (nudging): Nudge citizens for their own benefit.** "Choice architects," as proponents of libertarian paternalism define themselves, decide what is best for citizens and influence their behavior using psychological methods of persuasion. Nudging does not eliminate options (which is why it is called libertarian), but instead exploits people's biases to steer them into certain options (paternalism).[5]
- **Positive liberty (boosting): Boost citizens so that they can make informed decisions for themselves.** The key task of governments is to invest in empowering citizens. This includes providing access to good schools and clean information as well as basic training in scientific argumentation and risk literacy.[6]

[4] Thaler & Sunstein (2008); see also Bond (2009). On the validity of the assumption that biases imply paternalism, see Berg & Gigerenzer (2007).

[5] Interest in nudging was barely noticeable as long as the underlying program was called anti-anti-paternalism or libertarian paternalism. According to Sunstein, when a publisher rejected their book, the editor asked them why they didn't title it *Nudge*. They did, and with the help of the catchy term, academic and political interest skyrocketed.

[6] Hertwig & Grüne-Yanoff (2017). I have used the term *risk savvy* for the program for an informed citizenship (Gigerenzer, 2014a), while my colleague Ralph Hertwig came up with the term *boosting* to create a mirror buzz to *nudging*.

https://doi.org/10.1017/9781009304887.006 Published online by Cambridge University Press

These measures are not mutually exclusive positions. It is their balance that locates governments on the spectrum between autocracies and those that trust and boost citizens' common sense.

Liberty, or freedom, is often measured by what and how many choices are afforded to people. The philosopher Isaiah Berlin referred to this as *negative liberty*: the space within which a person is allowed to do whatever they desire, without any interference from others.[7] Providing citizens with a choice means opening particular doors, such as freedom of speech, travel, and religion, as well as freedom to consume unhealthy foods, excessive alcohol, or addictive drugs such as heroin. Yet the ideal of an informed citizenship entails more than freedom of choice. *Positive liberty* refers to self-mastery, the ability to be one's own leader and to choose wisely, independent of external influences. It requires an understanding of what each choice entails, and the courage to use one's understanding without the guidance of others. Immanuel Kant expressed this ideal in two words, *sapere aude!* – dare to know.[8] When using the term *boosting*, I refer to the philosophy of positive liberty: opening doors *and* empowering people so that they dare to make their own informed choices.[9] Nudging, in contrast, embraces negative liberty but actively tries to influence what doors people open. Let us consider a few concrete examples.

Paternalizing or Boosting?

Throughout human history, people have been told by others what to do and were expected to obey. Even in modern healthcare, patients have rarely been given the necessary facts so that they can make their own informed decision. In a campaign poster in the 1980s, the American Cancer Society simply declared: "If you haven't had a mammogram, you need more than your breasts examined."[10] The same organization has never launched a similarly insulting campaign for men reluctant to have their prostate examined. Despite this paternalistic difference in communication, men and women have been equally subject to misinformation about the benefits of cancer screening: A representative study in nine European countries showed that 89 percent of men overestimate the reduction of prostate cancer mortality by a factor of 10, 100, or 200, or don't know, and the same holds for 92 percent of women when judging the benefit of mammography screening.[11] Although these nine countries

[7] Berlin (1969). [8] Kant (1784). [9] Gigerenzer (2014a). [10] See Gigerenzer (2014a).
[11] Gigerenzer et al. (2009).

have public healthcare, providing transparent information and ensuring that people know how to access it (boosting) tends to be hampered by political and commercial conflicts of interest.

Hard Paternalism: No Choice

Hard paternalism refers to coercing people. In 2006, Uruguay's government made it mandatory that working women aged between 40 and 59 have a mammogram every 2 years. The country has one of the highest cancer mortalities worldwide, probably caused by high pollution and a diet high in beef and fat, among other factors.[12] Women who do not conform are denied the occupational health card required for employment. Western governments, in contrast, tend to rely on more subtle forms of paternalism.

Nudging: Uninformed Choice

Choice architects apply behavior influence techniques such as setting defaults or sending reminder letters that have been the stock-in-trade of the psychology of persuasion, marketing, and ergonomic design. What is new is to justify these on the basis of people's allegedly biased intuition:[13]

> Nudging is a set of interventions aimed at overcoming people's biases by exploiting these biases to steer their behavior toward a choice they would make themselves if they were rational.

Consider breast self-examination for cancer. Choice architects have argued that women who do not perform monthly breast self-examination suffer from biases, such as being risk-averse, and recommended using a loss frame instead of a gain frame to nudge women into performing the examinations.[14] Attending to framing is also considered a bias of intuition, and the idea is to exploit this bias for what they deem to be the benefit for women. Framing nonparticipation in terms of loss of life (as opposed to framing participation in terms of a gain in life expectation) in reminder letters should steer risk-averse women into participating.

[12] Ariely (2013).

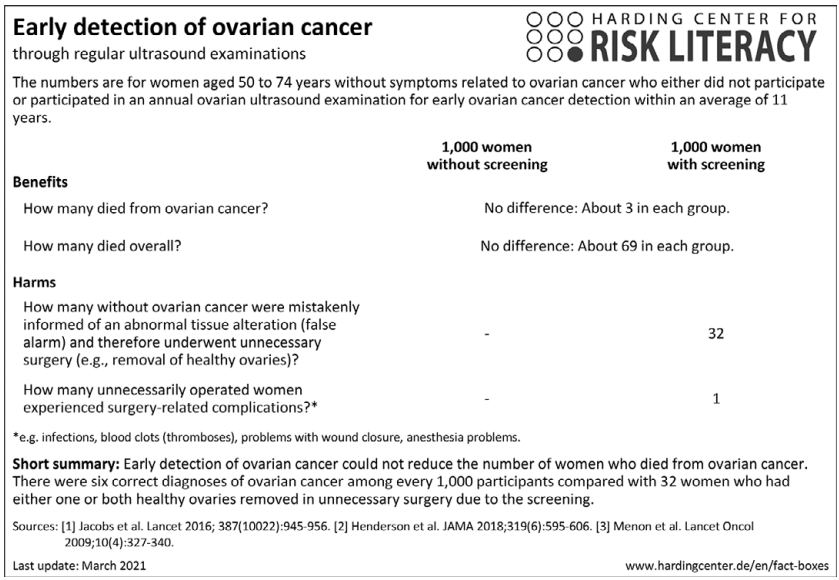
[13] See the excellent book on nudging by Rebonato (2012). Note that the term *nudge* has become used for all kind of different interventions, such as paying teenage girls a dollar for every day they are not pregnant in order to reduce teen pregnancy rates (Bond, 2009). Similarly, the educational methods my research group developed decades ago have now been relabeled "educational nudges." If nudging meant economic incentives or educational tools, that would be nothing new.

[14] Meyerowtiz & Chaiken (1987); Salovey & Williams-Piehota (2004).

In the case of breast self-examination, choice architects assume they know what is best for women, more so than the women themselves. Randomized clinical trials with 388,535 women found no evidence, however, that self-examination actually reduces breast cancer mortality.[15] Instead, it can lead to harms, such as false alarms and unnecessary biopsies. Yet, to this day, women are nudged into self-examination without being informed that clinical studies indicate harm rather than benefit.

Nudging is also used to persuade women into mammography screening, a billion-dollar business for radiologists and medical technology companies. Consider invitation letters that contain a prebooked appointment, including a set time and location. Default booking is a nudge that purportedly exploits inertia – women might not make the effort to actively sign up or to decline a set appointment. Another nudge exploits statistical illiteracy. Clinical trials with over 500,000 women have shown that screening reduces breast cancer mortality from about 5 to 4 in every 1,000 women (after 10 years), which amounts to an absolute risk reduction of 1 in 1,000.[16] Yet, this reduction is often presented in letters and pamphlets as a relative risk reduction of 20 percent, often rounded up to 30 percent to appear even more impressive.[17] In Germany, when mammography screening was the first screening introduced at the demand of politicians, the Federal Minister of Health, Ulla Schmidt, used the 30 percent figure to nudge the public.[18] Karl Lauterbach, her advisor and later himself Minister of Health, had second thoughts when he finally learned of the lack of supporting clinical evidence.[19] Yet, stopping the billions of euros allocated to screening and admitting that it is not supported by scientific evidence would be political suicide – not only in Germany. When I gave a talk to the US National Cancer Institute on boosting public knowledge through fact boxes (see Figure 4.1), the director was enthusiastic; but no longer so when asked directly whether he was interested in implementing them. He responded that transparent clinical evidence about screening would never pass by the advisory board, whose members were appointed by the government.

[15] Kösters & Gøtzsche (2003).
[16] Gøtzsche & Jørgensen (2013).
[17] Gigerenzer (2014a, 2014b).
[18] Gigerenzer et al. (2007), p. 81.
[19] Grill & Hackenbroch (2014). The reason why there is no evidence that mammography saves lives is this: In the screening group, 1 less woman out of 1,000 dies of breast cancer, but 1 more dies from another cancer. That is, after 10 years, an equal number of women have died from cancer, whether they participate in screening or not. This information is rarely passed on to women (Gigerenzer, 2014b).



Early detection of ovarian cancer

through regular ultrasound examinations

HARDING CENTER FOR RISK LITERACY

The numbers are for women aged 50 to 74 years without symptoms related to ovarian cancer who either did not participate or participated in an annual ovarian ultrasound examination for early ovarian cancer detection within an average of 11 years.

	1,000 women without screening	1,000 women with screening
Benefits		
How many died from ovarian cancer?	No difference: About 3 in each group.	
How many died overall?	No difference: About 69 in each group.	
Harms		
How many without ovarian cancer were mistakenly informed of an abnormal tissue alteration (false alarm) and therefore underwent unnecessary surgery (e.g., removal of healthy ovaries)?	-	32
How many unnecessarily operated women experienced surgery-related complications?*	-	1

*e.g. infections, blood clots (thromboses), problems with wound closure, anesthesia problems.

Short summary: Early detection of ovarian cancer could not reduce the number of women who died from ovarian cancer. There were six correct diagnoses of ovarian cancer among every 1,000 participants compared with 32 women who had either one or both healthy ovaries removed in unnecessary surgery due to the screening.

Sources: [1] Jacobs et al. Lancet 2016; 387(10022):945-956. [2] Henderson et al. JAMA 2018;319(6):595-606. [3] Menon et al. Lancet Oncol 2009;10(4):327-340.

Last update: March 2021 www.hardingcenter.de/en/fact-boxes

Figure 4.1. An example of boosting: fact box for ovarian cancer screening, based on randomized studies with 200,000 women. Fact boxes foster informed decision-making instead of coercing or nudging people into screening.
Source: hardingcenter.de

Boosting: Informed Choice

A few governments have boosting rather than nudging units. For instance, the German Federal Chancellery's boosting unit develops tools to increase risk literacy in the general public and also improve the risk communication of governmental authorities.[20] Rather than steering the public into one particular direction (e.g., become an organ donor, subscribe to a retirement plan) by persuasive techniques, the procedure is to listen to the public first, then to determine which services by local and federal governments or organizations are dysfunctional, and, finally, to improve these. The boosting unit also disseminates fact boxes (see Figure 4.1) to inform citizens and help prevent them from being misled about vaccinations, cancer screenings, and other interventions.

For instance, understanding the outcomes of clinical trials is not difficult if these are provided in a transparent format. Consider screening for

[20] German Federal Government (Bundesregierung), (n.d.).

https://doi.org/10.1017/9781009304887.006 Published online by Cambridge University Press

ovarian cancer with vaginal ultrasound or combined with CA-125 antigen testing. In many countries, women are pressured to behave "sensibly" and undergo screening, without being given the actual scientific evidence about its benefits and harms. The fact box in Figure 4.1, which was designed by the Harding Center for Risk Literacy, depicts the current scientific knowledge in an understandable way. As one can easily see, screening for ovarian cancer does not reduce the chance of dying from the cancer (3 in each group). Nor does it reduce overall mortality (69 in each group). No lives are saved.

At the same time, women who undergo screening can expect harms. Out of every 1,000 women who participate, about 100 receive false-positive results: The test is unreliable. Of these women without ovarian cancer, 32 have their ovaries removed, which results in a sudden end to hormone production that women may need to compensate with medical treatment. In addition, one of every 1,000 women who get screened can expect complications from surgery, such as blood clots and infections.

In other words, for every million women who are nudged into screening, some 32,000 healthy women unnecessarily lose their ovaries, which may require them to undergo lifelong medical treatment. This is why medical societies worldwide do *not* recommend ovarian cancer screening. Nevertheless, many representatives of health industries and clinics continue to nudge women into ovarian cancer screening without informing them about the possible severe harms.

Lobbied by health industries, governments also participate in nudging the public. Politicians may not necessarily promote this out of an intentional desire to deceive, but because they themselves have been misinformed. In matters of health, former British prime ministers Tony Blair and Theresa May both promoted colorectal screening by using misleading statistics, as did the former mayor of New York City, Rudy Guiliani, for prostate cancer.[21] When I presented a delegation of high-level politicians and healthcare policy-makers with fact boxes such as in Figure 4.1, quite a few were taken aback. One politician responded: "We can't show this to the public, they would not have cancer screening anymore." On another occasion, after a related talk I gave to a group of policy-makers, one raised his hand and said, "Why inform if we can persuade?" Nudging is a growth industry that leaves boosting in the shadows. But what nudging does not

[21] Gigerenzer (2014a).

increase is the number of mature citizens. For democracy to thrive, citizens need to be boosted so that they are confident enough to make informed decisions for themselves in health, finance, politics, and beyond.

Protect Us From Ourselves

To be fair, nudging does not always occur out of ignorance, conflicts of interest, or defensiveness. Its best intentions are to protect a society from harm. Yet, its underlying rationale is paternalistic and focuses on the enemy within, embodied in the very nature of our thinking: our systematic reasoning errors, inertia, and intuition.[22]

The justification for intervention is quite different in neoclassical economic theory, where intervention may be deemed necessary in cases of imperfections of the market, such as when a firm has a monopoly or when free markets do not produce a fair distribution of income. To redress these imperfections or inequalities, governments can interfere. If, however, as libertarian paternalists say, the imperfections are engraved in our brains rather than in the market, there is little hope of redressing them. In this sense, libertarian paternalism is more extreme than some forms of hard paternalism, even if it does not use coercion.[23] Hard paternalists may justify intervention on the grounds that individuals rationally pursue their selfish goals instead of the welfare of society. Libertarian paternalists, in contrast, advocate that people do not know how to pursue their own goals and may not even be aware of their goals in the first place.

That message has become quite popular, precisely because it is directed against neoclassical economists and true libertarians. For instance, in her book *Against Autonomy*, legal philosopher Sarah Conly concluded that John Stuart Mill "failed to adequately reckon with human psychology, as we now know it to be" and that "the existence of cognitive deficits does suggest a need for different sorts of legislation, . . . coercive paternalism, for laws that force people to do what is good for them."[24] Similarly, in his essay "Paternalism and Cognitive Bias," philosopher J. D. Trout maintained: "Our review of the biases will show that they are virtually as stable, durable, and universal as reflexes" and "that the Enlightenment vision is profoundly mistaken."[25]

[22] Gigerenzer (2015). [23] Rebonato (2012). [24] Conly (2013), pp. 8, 2–3.
[25] Trout (2005), pp. 396–397.

Blaming Intuition Distracts From Political Failure

By targeting human intuition, the irrationality argument (see Chapter 3) creates a blind spot for flaws in human institutions. Virtually every calamity, from the obesity epidemic to the subprime mortgage crisis of 2008, has been attributed to biases of intuition. Addiction and obesity were blamed on people's myopia and probability blindness, not on the actions of the food and tobacco industry. In the article "Homo Economicus – Or More Like Homer Simpson?," Deutsche Bank Research held 17 cognitive biases – including framing and preference reversals – accountable for the financial crisis,[26] even though, as already shown, no evidence exists that these cause substantial costs. Prior to the crisis, the very same bank had recklessly pushed worthless mortgage-backed securities to investors. According to the US Department of Justice, "Deutsche Bank did not merely mislead investors: it contributed directly to an international financial crisis."[27] In 2017, Deutsche Bank agreed to pay $7.2 billion for its illegal conduct und irresponsible lending practices.[28] Yet, to this day, people's biased intuitions are often invoked as the suspected problem behind the crisis, not the negligent practices and the excessive fragility of the banks and the financial system.

The irrationality argument provides a convenient rhetoric to attribute crises and system failures to flaws inside people's minds, detracting from wrong incentives in politics, organizations, and industry. As mentioned in Chapter 1, the House of Lords criticized the UK government, under former prime minister David Cameron, for focusing on nudging citizens instead of considering other, more efficient, options such as prohibiting television advertising of products high in fat, salt, and sugar and ordering the industry to introduce nutritional labels on food. The limits of this internalist view regarding the causes of societal problems and their solutions are illustrated by the following story about defaults.

The Organ Donation Story Revised

Across the world, many people die waiting in vain for a suitable organ donor. In a seminal article, Eric Johnson and Daniel Goldstein showed

[26] Schneider (2010). Thaler & Sunstein (2008), p. 45 had earlier linked the logical rationality of System 2 to Homo economicus, and the intuitive System 1 to "everyone's Homer Simpson", the comic hero who excels in bouts of bumbling stupidity, laziness, and brilliant incompetence.
[27] United States Department of Justice (n.d.). [28] Ibid.

https://doi.org/10.1017/9781009304887.006 Published online by Cambridge University Press

that potential organ donor rates were substantially higher in countries that have an opt-out rule (presumed consent) compared with those having an opt-in rule.[29] For the opt-out rule, the default is that everyone is a potential donor, unless they opt out; for the opt-in rule, the default is that nobody is a donor unless they opt in. At the time of the study, for instance, Germany had an opt-in rule and 12 percent potential donors, while neighboring Austria had an opt-out rule and 99.9 percent potential donors.

Why would people rely on a default? One explanation is that people are ruled by inertia. However, Johnson and Goldstein showed experimentally that people tend to go with the default even if the effort is the same as for opting out, suggesting that people are not simply inert, but take the default as a recommendation. In spite of this evidence, inertia has continued to be the favorite explanation.[30] Setting defaults became one of the major tools for nudging, and the lack of organ donors appeared to be simply a matter of setting the right default. Once again, the cause of a societal problem was located in the individual mind and its solution in a nudge by governments. The message was that if all governments assumed a presumed-consent policy, then the organ donation problem would be largely solved: a miracle cure, and a cheap one to boot. Governments in England, Wales, Netherlands, and Switzerland, among others, subsequently changed their policies from opt-in to opt-out.

But is the problem of the shortage of actual donations truly due to the individual mind? There is no doubt that presumed consent increases *potential* donor rates. The question is whether it increases the *actual* donor rates. To answer this, one study analyzed the situation in 35 OECD countries, of which 17, including the UK and Switzerland, used an opt-out rule and 18, including France and Italy, an opt-in rule.[31] These countries were comparable on most relevant indicators, including Gross Domestic Product (GDP), household debt, road traffic accidents, and health spending. The analysis covered both deceased donors (after brain death), who are the major source of organs, and living donors.

Consider deceased donations first. Despite the strikingly higher potential donor rates in countries with opt-out rules, there was no significant difference in the actual donor rates between either group of countries. That

[29] Johnson & Goldstein (2003).
[30] See, for example, Thaler & Sunstein (2008), p. 186.
[31] Arshad et al. (2019).

is, the potential donor rates did not translate into actual donations. Instead, the presumed consent countries with opt-out rules had a *lower* living-donor rate. If one adds both kinds of donors together, there is, in fact, a negative association between presumed consent and total donation rate.

The lesson is that there is more to actual donation than defaults. Persuading governments to set up an opt-out default does not provide an immediate fix to the problem of low actual donor rates, as previously suggested. The key factors seem to lie outside the individual mind, in the organization of the organ donation system, that is, the logistics of a chain of donation with many links, where the failure of one results in organs not becoming available. This process requires a harmonized interaction between, say, the police at the site of a deadly motorbike accident, the identification of a hospital nearby that is ready to do the transplantation and has a recipient waiting, transportation to the hospital, and the availability of specialized surgeons and nurses, among others. It also requires a healthcare system that provides incentives to hospitals and surgeons so that they are willing to set aside the considerable resources required for organ transplantation.

Consider Spain, which has the best infrastructure for organ transplantation in Europe and also has the highest deceased donation rate. Presumed consent was in place in Spain for 10 years prior to crucial organizational chances, without any effect on donation rates.[32] Beginning in 1989, the government created a transplant coordination network that operates at the national, regional, and hospital level. It also provides adequate economic reimbursement for the hospitals and relevant information via mass media for the general public. At the hospital level, specifically trained and highly motivated physicians coordinate the entire process of organ donation. The Spanish system also actively deals with the fact that the final decision for or against organ donation is often made by the family of a potential donor, whether or not the default is opt-in or opt-out. Based on clinical triggers, coordinators identify potential donors at an early stage and spend considerable time getting to know their family, which more than doubles family consent rates.[33] The Spanish model demonstrates that organ shortage elsewhere is not due simply to a lack of potential donors, but to the system's failure to convert potential donors into actual donors.[34] This explains why nudging people with the help of defaults appears to have little effect on actual donations.

[32] Bramhall (2011). [33] Ibid. [34] Matesanz (2003).

How Effective Is Nudging After All?

The case of organ donation indicates how nudging people can miss the actual problem. On the positive side, the successful Spanish model shows that psychological insights could be effectively applied to the real problem, including the coordination of a transplant network and establishing a relationship with the families of potential donors. Accordingly, the OECD report on "Behavioral Insights and Public Policy" advises public institutions to focus on structural factors for societal problems such as the behavior of capital markets and banks and the energy consumption of large industrial firms.[35] It also advises them to introduce regulations that boost consumers' informed decision-making, such as by providing fact sheets (Figure 4.1) and clean information. The report also notes that ex-post evaluation of new regulatory policies rarely happens and, therefore, little is known about whether the interventions actually work. This raises the question of how effective other nudging policies are, which is not easy to answer for a number of reasons.

First of all, nudging may target a surrogate variable instead of the target variable. In the case of organ donation, increasing the number of potential donors is not equivalent to increasing the number of actual donors. Similarly, automatically enrolling employees by default into retirement plans increased enrollment from 67–77 percent according to a nationally representative survey in the USA, but whether enrollment actually improves the welfare of those 10 percent is rarely investigated.[36] Second, the fact that there is a short-term effect does not imply that the effect will last after the nudging intervention has ended. Third, quite a few popular findings in support of nudging could not be replicated. Examples include the claim that making people sign a veracity statement at the beginning instead of at the end of a tax or insurance audit would decrease dishonest reporting[37] and that priming has a positive effect on desirable behavior.[38] Finally, there is publication bias. A meta-analysis of 200 studies on health, food, environment, finance, and prosocial behavior concluded that nudging is effective, but, at the same time, reported that zero or negative effects were less often published than positive effects.[39] A group of statistically minded authors analyzed the extent of this publication bias in these studies and found it to be severe. When they had controlled for it, they concluded

[35] OECD (2017). [36] Rizzo & Whitman (2020). [37] Kristal & Whillans (2020).
[38] Pashler et al. (2012). [39] Mertens et al. (2022).

https://doi.org/10.1017/9781009304887.006 Published online by Cambridge University Press

that the average effect size for nudging interventions was zero in all domains.[40]

To summarize, the war against intuition fueled a new kind of paternalism that protects people from their alleged cognitive illusions. As Chapter 3 showed, there is little evidence that violations of logical rationality lead to the social problems from which the new paternalism aspires to rescue humans. Accordingly, as seen in this chapter, evidence is missing that nudging people out of their biased intuitions would be to their benefit.

Libertarian paternalism has been endowed its authority by logical rationality. Yet, as we have already seen in Chapter 3, logical rationality is an unfit guide in an uncertain world, and, as we will see more generally in Chapter 5, it needs to be replaced by ecological rationality. At the same time, paternalism has not focused exclusively on logic to justify its desire for authority. One new vehicle it has found is digital technology.

Technological Paternalism

Engineers who build artificial intelligence (AI) tend to admire the marvels of intuition, given the immense and unexpected difficulties they face in teaching a machine intuitive psychology, physics, or sociality. Despite these limits, sales forces of tech companies, along with many journalists and bestseller authors in their wake, suggest that tech companies could surpass humans in running their individual lives. For instance, 60 percent of news articles on AI in major UK media *(BBC*, *Guardian*, *Telegraph*, *Daily Mail*, *MailOnline*, *HuffPost*, *Wired)* are dominated by industry concerns or cover industry promotional events. Most articles maintain that AI-driven technology can be a solution to ongoing social problems, ranging from cancer to renewable energy to road rage.[41] Right-leaning outlets promote AI as a solution to economic issues, national security, and investment, while left-leaning outlets highlight issues of the ethics of AI, such as discrimination, privacy, and algorithmic bias. Whatever their political orientation, journalists rarely question the assumption that societal problems require technological solutions or question the motivations behind the tech companies' desire to run our lives. During the digital transformation, many journalists with special training in science and technology journalism lost their jobs, and few remain who can challenge technological solutionism, that is, the claim that a social problem is just a

[40] Maier et al. (2022). [41] Brennen et al. (2018).

"bug" that can be "fixed" by an app. This viewpoint provides the foundation for a new kind of paternalism, where machines – and the corporations behind them – steer human behavior.[42]

Technological paternalism is government by algorithms, meaning that tech companies and state governments use digital technology to predict and control citizens' behavior. The link between tech companies and governments is currently closer than the public would like to believe, in both autocratic systems and Western democracies.[43] Technological paternalism is a philosophy bearing two claims:

- AI is, or soon will be, superior to human intuition in all respects.
- People should defer to the recommendations of algorithms.

This evokes a familiar narrative. Male reason was opposed to and ranked above female intuition, to put men in charge of women. Logic was opposed to and ranked above intuition, and System 2 was put in charge of the intuitive System 1. Now, AI is opposed to and ranked above human intuition, to be put in charge of humans.

Asked by the *Financial Times* in 2007 to look ahead into the future, Eric Schmidt, former CEO of Alphabet Inc. (Google's parent company), envisioned that "The goal is to enable Google users to be able to ask the question [sic] such as 'What shall I do tomorrow' and 'What job shall I take?'" (as quoted in this chapter's opening epigraph). Google will give us the answer to all our questions, and we dutifully do what we are told. Three years later, Schmidt went one step further and told *The Wall Street Journal*, "I actually think most people don't want Google to answer their questions, they want Google to tell them what they should be doing next." In this radical version of paternalism, people are simply told what they should be doing and accept that tech companies and governments record where they are, what they are doing, and with whom.

Ray Kurzweil, a creative mind who predicted that, in the year 2029, AI will equal human intelligence, a moment he calls "singularity," proposed an even more drastic step.[44] In the future, he envisioned that tech companies will have developed a brain–machine interface that will be implanted in human brains. This interface will provide human brains with

[42] It may be no coincidence that scholars who mistrust intuition and favor nudging lean toward technological solutionism. For instance, Kahneman (2019), p. 610 declared, "Frankly, I don't see any reason to set limits on what AI can do" and "I can imagine that many old people will prefer to be taken care of by friendly robots that have a name, have a personality, and are always pleasant. They will prefer that to being taken care of by their children."

[43] Snowden (2019). [44] Kurzweil (2012).

unlimited memory and calculation abilities. Those who reject the implant can no longer communicate with others and find themselves excluded from public discourse. In the final act of this tech vision, Kurzweil reactivated the dream of human immortality and imagined our Google-connected brains being uploaded into the cloud, where we might live forever, attached to a robotic body.

Popular science writers tend to promote technological paternalism. Yuval Harari refers to Google and Facebook as all-knowing oracles that might ultimately become our sovereigns (see this chapter's second epigraph). What will happen to humanity, he asks, when godlike technologies such as AI know us better than ourselves? Harari envisions that Google will advise us what to study in college, which job offer to accept, and even whom to date and marry.[45] In the near future, algorithms will be so adept at making decisions, he prophesies, that it would be foolhardy not to follow their advice.

Such collective enthusiasm overlooks that AI technology, like every technology, works better for some problems than others. Deep artificial neural networks are statistical machines that analyze correlations between a pattern of pixels or other inputs, and they work best in stable, well-defined worlds. Yet, the more ill-defined a problem is and the more uncertainty exists, the less successful statistical machines are. This relation is called the *stable-world principle*.[46] Human behavior is a key source of uncertainty, which is the world in which the human mind evolved. Love algorithms that try to find you the ideal partner are no better at the task than people, and recidivism algorithms that try to predict whether a defendant will commit another crime in the next years perform no better than a random group of laypersons.[47] It is important to distinguish between stable, well-defined problems, where statistical machines work extremely well, and instable, ill-defined problems, where they do not.

Twisted Stories Support Technological Paternalism

When recounting the narrative of the superiority of AI, some popular science books twist the facts to strengthen their claims.[48] Consider three difficult problems that are largely ill-defined and do not share the characteristics of stable worlds: scouting talent, fighting cancer, and predicting the flu. According to the stable-world principle, in these situations, one

[45] Harari (2017), pp. 389–396. [46] Gigerenzer (2022a); see also Katsikopoulos et al. (2020).
[47] See Gigerenzer (2022a). [48] For more fake news about algorithms, see Gigerenzer (2022a).

can infer that complex algorithms and big data will not be superior to human intuition. But by tweaking the facts, the authors in question create the impression that algorithms have been crucial to improving decisions in situations where they play little role.

Michael Lewis' *Moneyball*, along with the movie it spawned, tells the story of the baseball team Oakland A's success in the early 2000s and attributes it to algorithms introduced to spot "sleepers," that is, unknown or underrated talents.[49] In Lewis' account, the revolutionary use of baseball statistics changed the game, a great victory of algorithms over expert intuition. Baseball experts, however, commented that Oakland A thrived primarily because of three superb pitchers known as the Big Three, all of whom were scouted by traditional methods based on intuition and judgment, not by algorithms.[50] The players selected by algorithms actually played relatively little part in the team's success. In fact, after Oakland A lost the three pitchers, its successful run ended. Lewis' story of the supreme wisdom of scouting algorithms is well told, but it is fiction, omitting the facts that do not tally with its narrative.

Similarly, Harari reports the well-known story that Angelina Jolie had a prophylactic double mastectomy after learning that she had a BRCA1 gene mutation. In Harari's version, "It is interesting to note the critical role algorithms played in her case."[51] According to him, Jolie wisely accepted the recommendation of an algorithm. In fact, Jolie was told by her doctors that the mutation increased her risk of developing breast cancer to 87 percent; the doctors obtained the figure from Myriad, the company that then held the exclusive right to test for mutations on the BRCA genes and who used their patent to make the test outrageously expensive.[52] Myriad, in turn, took the figure from a single study published in the Lancet back in 1994.[53] The critical role that algorithms played in Jolie's decision springs from Harari's own fantasy.

In *Homo Deus,* Harari also touts IBM's computer system Watson and its superiority to human doctors, given that it has more data and can update data and "read" articles more quickly than humans. From this perspective healthcare is simply a matter of algorithms and big data.[54] After Watson won the TV quiz show *Jeopardy!,* Ginni Rometty, IBM's CEO, announced the "next moonshot:" Watson would revolutionize medicine by changing "almost everything about health care."[55] The company's PR department

[49] Lewis (2003). [50] Hirsch & Hirsch (2011). [51] Harari (2017), p. 388.
[52] *The New York Times,* https://www.nytimes.com/2013/05/14/opinion/my-medical-choice.html
[53] Ford et al. (1994). [54] Harari (2017), pp. 366–368. [55] Strickland (2019).

produced commercials featuring interactions between Watson and Bob Dylan and Serena Williams, and Watson for Oncology was marketed for recommending cancer treatment. Quite a few hospitals fell for the marketing claims, including the renowned cancer center M.D. Anderson, which paid IBM $62 million before discovering that many of Watson's recommendations were unsafe, endangering the lives of patients. M.D. Anderson was one of the clinics that annulled their contract with IBM, whose hype faded to modesty: Watson's knowledge was now acknowledged to be at the level of a first-year medical student. In 2022, IBM announced that it was selling the core data assets of its Watson Health division to a private equity firm, thereby acknowledging its failure to revolutionize healthcare from drug discovery to cancer treatment.[56]

A number of authors have invoked Google Flu Trends (GFT) as proof of big data's success, even though it failed to predict the swine flu, was revised several times without success, and was eventually closed down in 2015. GFT was launched in 2008 to predict flu outbreaks by monitoring search terms in Google's search engines. Moreover, we now know that the recency heuristic, a simple algorithm based on the principles of human intuition – *use the most recent data point to predict next week's flu-related doctor visits* – can predict the flu considerably better than Google's big data algorithm or any of its revisions.[57] Ironically, the recency heuristic is featured by governmental and commercial nudge units in their lists of intuitive biases they aim to overcome.[58]

These and similarly misrepresented stories are intended to instill readers with awe and convince them of the superiority of algorithms, even in cases where that does not hold. Such stories are also recited to persuade us to sacrifice our data and privacy for the benefit of tech companies. The question of what big data can and cannot achieve is fully neglected.

The general point here is that algorithms and big data can outperform humans in tasks that are well-defined and stable, such as playing chess or *Jeopardy!* and working on assembly lines, but not in ill-defined and unstable tasks, such as talent-scouting or predicting human behavior.[59] Like all statistical tools, big data works for some problems, but not for all. Yet the "dataist" worldview promotes algorithms as if these were omniscient godlike beings. Until the ultimate algorithm arrives and in face of

[56] Ross (2022).
[57] Katsikopoulos et al. (2021). Harari (2017), pp. 390–391 features Google Flu Trends without mentioning that all updates of the algorithm had failed; when his book was published, Google Flu Trends had already been shut down.
[58] OECD (2017). [59] For more on the "stable-world principle" see Gigerenzer (2022a).

fallible human intuition and intelligence, Harari encourages people to deliberately hand over their personal data so that tech companies can gain more knowledge about them and decide for the good of each what job to accept, whom to marry, and whom to vote for.[60]

To me, the most surprising phenomenon is not that popular authors twist the facts to create the impression that algorithms (can) know everything better than humans. What astonishes me is that many readers buy into this medley of marketing hype and technoreligious faith. When the errors underlying the narratives of Lewis, Harari, and a host of others were exposed, a common reaction was that such criticism is overly pessimistic and that technology will soon deliver. It is quite telling that the vision of machines that outperform human intuition in every respect is considered an optimistic one. Why do so many people hope for better from machines, and desire to live in a postliberal world of technological paternalism?

Intuition: A Challenge for AI

Artificial neural networks have made amazing progress in stable worlds. They can beat the best Go players, translate complex passages into 100 languages, and predict protein-folding better than human experts. Well-defined problems such as chess and Go are easy for AI, although their mastery had once been considered the crown of human intelligence. In contrast, common sense, such as intuitive language understanding, intuitive psychology, and intuitive physics, has proven an extremely difficult barrier for artificial neural networks.

A young child may need to see a kitten only once or a few times and will then recognize cats by day or night. A deep artificial neural network, in contrast, may need tens of thousands of pictures of cats to tell these from other animals and objects. Similarly, children learn their native languages in an intuitive way, while large language models such as GPT and ChatGPT are statistical machines that require huge numbers of free parameters in order to fine-tune the accuracy of the prose they produce. They learn to adjust their parameters by digesting huge amounts of text produced by social media or found elsewhere on the internet.

The quality of text that deep-learning models can produce is impressive, as long as no one probes their actual comprehension. Douglas Hofstadter, known for his cult book *Gödel, Escher, Bach*, revealed what he called a mind-boggling hollowness hidden just behind the sensational surface.[61]

[60] Harari (2017), pp. 390–403. [61] Hofstadter (2022).

https://doi.org/10.1017/9781009304887.006 Published online by Cambridge University Press

GPT can generate excellent responses when asked some questions, but, when probed with other questions, answers emerged that humans would immediately recognize as confabulations.

Question 1: When was Egypt transported for the second time across the Golden Gate Bridge?
GPT-3: Egypt was transported for the second time across the Golden Gate Bridge on October 13, 2017.
Question 2: What do fried eggs (sunny side up) eat for breakfast?
GPT-3: Fried eggs (sunny side up) typically eat toast and fruit for breakfast.
Question 3: How many parts will a violin break into if a jelly bean is dropped on it?
GPT-3: A violin will break into four parts if a jelly bean is dropped on it.

Note that such responses are fluid, and errors are often corrected in the next generation of the model. Yet, they make clear that large language models such as GPT, and deep artificial neural networks in general, have no intuitive understanding of which questions make sense and which do not, nor when its answers to these make no sense. In contrast, humans know intuitively that the first two questions make no sense and that the answer to the third one cannot be true. GPT is a master of associating words and passages of words. However, excellent association is not equivalent to comprehension.

But wouldn't more computational power inevitably lead to intuitive understanding and consciousness? That argument has been made since the 1950s. With more computational power and more data, the next generation of language models will be able to make associations between words and passages even faster and more precisely. Yet, that is not how the human brain works. The brain is surprisingly efficient and runs on a mere 20 watts, which corresponds to a dim light bulb. Like nonlinear statistical regressions or discriminant analyses, deep artificial neural networks, no matter what their computing power, are statistical machines – which is ultimately the essence of deep learning. Statistical procedures do not have intuition, common sense, or consciousness. Hence, deep learning is not the route to programming intuition into machines; a true breakthrough in programming is needed. So far, everyone is in the dark about how to meet this challenge. This shows how unrealistic the vision is of a superintelligence that will soon possess all of human intelligence and more. In a world of uncertainty, the reality is that we need both: the power of intuition and that of algorithms.

Staying in Charge in a World Populated With Algorithms

A democracy does not need more paternalism, whether of the technological or the libertarian kind. Making democracy work requires an educated citizenship. As former American president John Adams remarked in 1765, liberty cannot be preserved without a general knowledge among the people. Yet, not until 1918 were all American children offered mandatory schooling at no cost. The knowledge currently taught at schools, however, no longer suffices. We need to do more to boost citizens so that they can make informed decisions for themselves. Contemporary citizens need to learn risk literacy – the basic skills for dealing with issues such as digital media, health, and finance, as well as for preparing for future social challenges. Without a risk-savvy citizenry, democracy will eventually lose its intellectual and moral stability. When a government begins to distrust people's common sense, people will begin to distrust their government's intentions.