evasive entrepreneurs

& the future of governance

how innovation improves economies and governments

—

Adam Thierer

Paperback ISBN: 978-1-948647-76-2
eISBN: 978-1-948647-77-9

Cover design: FaceOut Studio, Spencer Fuller.

Printed in the United States of America.

Library of Congress Cataloging-in-Publication Data available.

CATO INSTITUTE
1000 Massachusetts Avenue NW
Washington, DC 20001

www.cato.org

CONTENTS

INTRODUCTION 1

1. WHY INNOVATION MATTERS 19

2. EVASIVE ENTREPRENEURIALISM: THEORY AND EXAMPLES 53

3. DISOBEDIENCE: THEN AND NOW 109

4. WHY EVASIVE ENTREPRENEURIALISM IS ON THE RISE 129

5. INNOVATION AS CHECKS AND BALANCES 159

6. HUMANISM, ETHICS, AND RESPONSIBLE INNOVATION 183

7. SOFT LAW AND THE FUTURE OF TECHNOLOGICAL GOVERNANCE 205

8. DEFENDING INNOVATION: A BLUEPRINT 241

9. CONCLUSION: THE CASE FOR RATIONAL OPTIMISM 265

POSTSCRIPT: EVASIVENESS DURING A PANDEMIC 277

ACKNOWLEDGMENTS 285

NOTES 287

INDEX 349

ABOUT THE AUTHOR 369

"The higher-ups have measures.
Those lower down have countermeasures."
—old Chinese saying

"Exit has an essential role to play in restoring quality performance of government, just as in any organization."
—Albert O. Hirschman, *Exit, Voice, and Loyalty*

6

HUMANISM, ETHICS, AND RESPONSIBLE INNOVATION

Thus far, this book has mounted a defense of technological innovation and argued that even acts of evasive entrepreneurialism have great value to society. But there are downsides to them, too. The next two chapters respond to some common objections about the ethical or human dimensions of technological innovation, specifically the allegation that innovation somehow undermines our humanity.

In response, I will show that technological innovation is fundamentally about improving our humanity by bettering our lives as well as the lives of those around us and even those far away from us. Properly understood, technology and humanism are complements, not opposites. I will also explore the tension between permissionless and responsible innovation and argue that not only are these concepts compatible, but also they are already being balanced through a variety of governance practices

sometimes referred to as soft law, as defined in the Introduction and discussed in Chapter 2.

Adaptation in the Face of Adversity

As we have seen, the future is unfolding more rapidly than law's ability to keep up with it. Some techno-optimists vociferously cheer these developments and predict a better world will follow. By contrast, a great many techno-pessimists loathe them and live in fear of the supposed dystopian hellscape to come.

I obviously lean strongly in the techno-optimists' direction in most cases. My perspective is based not on blind faith in technology but rather on a factual evaluation of what has, again and again, lifted humanity to new heights throughout history. By creating new tools to solve basic problems and fulfill important needs, technological innovation has improved human well-being.[1]

The pessimists have some legitimate concerns about the potential dangers associated with new technologies, however, especially as traditional governance systems break down and the potential for more widespread legal evasion grows. Although some of their concerns are understandable, others, and their proposed solutions, often leave much to be desired.

For example, many critics today decry the frictionless nature of modern innovations and suggest that society should hit the pause button on technological developments, or at least find reasonable ways to slow things down a bit.[2] It is fine and well to add friction *voluntarily* to one's personal routine in an attempt to achieve a better balance with modern technologies.[3] It is an entirely different matter, however, to suggest that friction should be forced upon us through coercive mechanisms or even regulatory nudges that restrict our options and opportunities. I have already made it clear why it would be a serious mistake to throw a wrench

in the gears of progress in that fashion. Technological innovation is the fundamental driver of human betterment over time. Stopping or even slowing the rate of technological change is a call for stasis, and stasis will be our ruin as a species. It would have a profoundly negative effect on economic growth, living standards, our health and welfare, and our personal autonomy.[4]

Even when the potential for some harm exists and the case for adding friction in the form of some sort of regulatory intervention may be stronger, it does not mean the proposed remedies will work or be cost-effective. It is always easy for technological critics and concerned policymakers to insist that something must be done when a new technology is speeding ahead. It is quite another thing for critics to devise workable policies that won't result in enormous costs for society and discourage the development of innovations that could significantly improve our quality of living. Virtually every major regulatory agency is already grappling with this problem today as these technological capabilities expand and acts of evasive entrepreneurialism increase.

Still, the critics have a fair point: The pace of innovation feels overwhelming at times, and it truly does come into conflict with ideas and institutions that have great importance to many people. How, then, can all those concerns be addressed to ensure that humanist values are preserved in an age of rapid technological change?

The short and honest answer is that not everything can be perfectly addressed to make those critics happy. Some ideas and institutions will need to adapt. Yet values have *always* evolved throughout the course of civilization.[5] That does not mean that innovation should (or will) sweep away all that is old. The more sophisticated answer to the above question is that society typically finds a way to find balance, adapt, and muddle through.[6]

We humans are a remarkably resilient species, and we regularly find creative ways to deal with major changes through constant trial-and-error experimentation and the learning that results from it. In that process, we find a new baseline or equilibrium and incorporate new ideas, institutions, and values into our lives. We will continue to do so, but it will not always be according to the sort of script that many critics desire.

Humanism, Technology, and the Specter of Determinism

Because many of the calls for responsible innovation flow from humanist critiques of technological innovation, it is important to first address what some mean by "humanist." As Chapter 1 noted, there exists no shortage of critics who label themselves humanists and who decry the supposedly deleterious effects technological change has on individuals, institutions, or culture. Humanist critiques of innovation can be different from each other, but a core attribute of most of them is the notion that technology and technological change are somehow at odds with humanity and human flourishing. Accordingly, critics regularly use terms like dehumanizing or "re-engineering humanity" when discussing their fears about new technologies.[7] I have already labored to prove just how off base those critiques are, but it is worth diving a little deeper to understand why technological innovation and human flourishing are complements, not enemies.[8] First, though, we must address the accusation that defending innovation represents little more than an acceptance of a technological future that is devoid of any concern about other important values.

Across the field of Science and Technology Studies (STS), scholars have long decried what is known as "technological determinism." Generally speaking, technological determinism is defined as the belief that "technological developments take place

outside society, independently of social, economic, and political forces" and that "technological change causes or determines social change."[9] The opposite of technological determinism is referred to as "social determinism" or "social constructivism," which "presumes that social and cultural forces determine technical change."[10]

In STS discussions, to be labeled a technological determinist these days is akin to being affixed with a scarlet letter of shame. It implies that you are a naïve technology booster who sees no role for politics, society, or average people in shaping their own destinies. Technological determinism, as defined by its many critics, represents the height of anti-humanist thinking. Those critics have also come up with many creative labels to describe the same notion, including "technologism,"[11] "techno-fundamentalism,"[12] "technological solutionism,"[13] and even "techno-chauvinism."[14] Regardless of the monikers the critics choose to decry technological determinist thinking, they are unified in thinking that "people-based solutions" represent the morally superior approach to ensuring that future populations will live in a "people-centered economy."[15]

Those critics are creating a false dichotomy. When tech critics play the humanist card, they seem to imagine that they somehow have nobler intentions and a deeper concern for the plight of people than others do. Meanwhile they attack those who dare suggest that technological change has been a core driver of human betterment, even though it is an unambiguous fact. Consider the way technological determinism is typically described in STS literature. Sally Wyatt articulates the common conception of deterministic thinking as follows:

> One of the problems with technological determinism is that it leaves no space for human choice or intervention

> and, moreover, absolves us from responsibility for the technologies we make and use. If technologies are developed outside of social interests, then workers, citizens, and others have very few options about the use and effects of these technologies.[16]

Framed in that fashion, it is completely understandable why critics would lambaste anyone adhering to such a worldview. In reality, however, few people hold such an extreme view about technology being the only important force shaping the course of history or human affairs.

What is particularly ironic is that some of the most rigid technological determinists are technology critics themselves. "A primary characteristic of the antitechnologists," Samuel Florman once argued, "is the way in which they refer to 'technology' as a thing, or at least a force, as if it had an existence of its own" and which "has escaped from human control and is spoiling our lives."[17] For example, some of the most notable tech critics of the past half century were French philosopher Jacques Ellul, American historian Lewis Mumford, and American cultural critic Neil Postman. Their books painted a dismal portrait of a future in which humans were subjugated to the evils of "technique" (Ellul),[18] "technics" (Mumford),[19] or "technopoly" (Postman).[20] The narrative of their works read like dystopian science fiction books. Essentially, there was no escaping the iron grip that technology had on us. Postman claimed, for example, that technology was destined to destroy "the vital sources of our humanity" and lead to "a culture without a moral foundation" by undermining "certain mental processes and social relations that make human life worth living."[21]

When dour tech critics like these preach the gospel of technological gloom-and-doom, they usually get a free pass from their

fellow tech critics despite the clear deterministic overtones. Apparently it is acceptable to use deterministic reasoning when your intentions are "pro-human" and your preferences are in line with other innovation critics. If, however, one dares employ any sort of deterministic arguments when speaking *optimistically* about the future, that person is decried as uncaring and anti-human.

Generally speaking, we can dismiss extreme deterministic reasoning—regardless of whether it's tech optimists or pessimists making such claims—for a rather simple reason: technologies fail all the time. "If promising technologies can suffer fatal blows from unexpected circumstances," Florman correctly argued, then "[t]his means that we are still—however precariously—in control of our own destiny."[22]

Technologies fail for many reasons, but societal demands and citizen pushback are two underappreciated explanations for why so many technologies flounder or are rejected. For example, in 2013, Google launched Google Glass, a pair of augmented reality "smart glasses" that would let users access information about their surroundings via a pop-up interactive display. Within two years, however, Google had canceled the project for consumer use and instead moved to offer a version of Glass only for commercial enterprises to use for specific workplace tasks. Perhaps Google Glass failed because of its hefty $1,500 price tag, or maybe there was not much consumer need for such a product yet. An equally compelling explanation for the failure of Google Glass was the "creepiness" factor associated with it. Privacy advocates decried the device and critics used the derogatory term "Glassholes" when referring to Glass users.[23] This product was an example of what Nobel Prize–winning economist Alvin E. Roth once referred to as "repugnance as a constraint on markets."[24] The intensity of the public backlash forced Google and other

augmented reality companies to reconsider the wisdom of wearable smart glasses. "If the stigma surrounding Google Glass (or, perhaps more specifically, 'Glassholes') has taught us anything," argued *Wired* journalist Issie Lapowsky, "it's that no matter how revolutionary technology may be, ultimately its success or failure ride on public perception. Many promising technological developments have died because they were ahead of their times."[25] A similar sort of public repugnance about new facial recognition technologies appears to be growing and could limit the diffusion of that technology.[26]

This example shows why deterministic thinking is too simplistic—people push back against technology all the time, and tools are constantly being reformed to better suit our collective desires and demands. We need a more balanced perspective in these debates. For lack of a better term, we might think of the middle-ground position as "soft determinism." That is, one can believe that technology plays an important role in influencing history—and that innovation oftentimes moves faster than law's ability to keep pace with it—while also believing that society, governments, and each and every human being can and will play a major role in shaping technology's nature and evolution. Others have defined soft determinism as the idea that "technological change drives social change but at the same time responds discriminatingly to social pressures."[27]

Although soft determinism represents a more reasonable position in these debates—and one that also offers a more realistic explanation of how technological governance works in practice—a great many scholars and policy advocates continue to heighten their approval of humanist labels and rhetoric. But what exactly is a humanist critic?

Most self-declared humanist scholars would probably agree with philosopher L. M. Sacasas that "[h]umanism is a rather

vague and contested term with a convoluted history."[28] To some extent, humanist critiques of technology are simply meant to remind us that all people are important, as is the case when some claim the humanist position represents "a philosophical claim about the centrality of humankind to the universe."[29] Again, who could be against such an assertion, or the repeated claim made by other self-anointed humanists who insist technological change has many tradeoffs and downsides? In a 2015 essay, Andrew McAfee of the MIT Sloan School of Management noted that such observations are uncontroversial and widely agreed upon.[30] The problem, he correctly noted, is that such banalities should not be used to end any inquiry into the benefits of technological change. Unfortunately, that is exactly what often happens in the field of science and technology scholarship and policymaking today. McAfee describes this attitude as follows:

> The third sense of "humanist" is by far the most problematic. It's close to: "Because I am for the people I should be free from having to support my contentions with anything more than rhetoric." Or, more simply: "You can trust what I say, because I am on the side of people instead of the cold, hard machines." Well, no. We should evaluate what you say based on the quality and quantity of evidence you've marshaled, and on the rigour with which you have analysed and presented it. If this sounds like an argument in favour of the scientific method, that's because it's exactly what it is.[31]

As McAfee suggests, critics who insist that technological innovation is anti-human or dehumanizing and use such rhetorical ploys to reject a particular innovation bear some burden of proof of the alleged harms. They must be willing to acknowledge that there are tradeoffs associated not only with new technologies, but also with the remedies they propose to any alleged downsides.

As I discussed in my previous book, those who advocate slowing or stopping technological advances need to demonstrate that the harms they allege are highly probable, tangible, immediate, irreversible, catastrophic, or directly threatening to life and limb in some fashion.[32] In recent years, risk analysis tools have improved and cost-benefit analysis has become formalized within the regulatory policymaking process. These tools and methods can be used by those advocating preemptive, prohibitive controls on new tech.[33] Oftentimes, as will be noted later, the critics do not bother spelling out what sort of remedies they think are appropriate. They feel it is enough to decry the supposed downsides associated with technology, suggest that "something must be done," and then presumably expect someone else (usually government actors) to take up that cause. That is where their analysis all too often ends. Little effort is put into exploring the full range of tradeoffs associated with the various (but unspecified) innovation-limiting actions they argue are needed.

At worst, tech critics sometimes rest their case for limiting innovation on nostalgic arguments about some proverbial good old days—all the while deftly avoiding telling us precisely when those days were. The problem with all the punditry in what Richard Posner once aptly labeled "the declinist genre" is that it is flatly at odds with the actual historical record regarding the state of human affairs in the past.[34] Even a cursory review of history offers voluminous, unambiguous proof that the old days were, in reality, eras of abject misery. Widespread poverty, mass hunger, poor hygiene, short lifespans, and so on were the norm. What lifted humanity up and improved our lot as a species is that we learned how to apply knowledge to tasks in a better way through incessant trial-and-error experimentation.[35] In other words, humanity flourished by *innovating*, and the results of our innovative activities were called *technologies*.

Technology is not some mystical force that appeared out of thin air. Nor is it an autonomous entity with a will of its own. *All technology is the product of human design and action*.[36] The most straightforward definition of "technology" is simply the application of knowledge to a task, and as Benjamin Franklin once noted, man is a tool-making animal by his nature. "[T]he elementary pleasure of solving technical problems and successfully completing constructive projects," Samuel Florman once correctly observed, is "as old as the human race."[37]

Thus there are few things more humanist than crafting tools to solve important problems and to better our lives and the lives of our loved ones and others.[38] One can simultaneously believe in "the centrality of humankind to the universe" as well as the notion that technological innovation is central to humankind's ability to improve the little corner of the universe that we occupy.

How Technology Expands the Horizons of Our Humanity

Technology helps us better understand and address the needs of strangers at a distance. In his 1759 *Theory of Moral Sentiments*, the Scottish moral philosopher and economist Adam Smith observed the following:

> How selfish soever man may be supposed, there are evidently some principles in his nature, which interest him in the fortunes of others, and render their happiness necessary to him, though he derives nothing from it, except the pleasure of seeing it. Of this kind is pity or compassion, the emotion we feel for the misery of others, when we either see it, or are made to conceive it in a very lively manner. That we often derive sorrow from the sorrows of others, is a matter of fact too obvious to require any instances to prove it; for this sentiment, like all the other original passions of

> human nature, is by no means confined to the virtuous or the humane, though they perhaps may feel it with the most exquisite sensibility. The greatest ruffian, the most hardened violator of the laws of society, is not altogether without it.[39]

Smith believed that humans were both self-regarding and other-regarding and that we had an innate moral sensibility and sympathy for others, or what he called a "fellow-feeling." Thanks to this natural sensibility, we would first look to take care of ourselves and those closest to us, but we would then look to help others the best we could.[40]

During Smith's time, however, that "fellow-feeling" for the plight of others was limited by social, economic, and technical realities. Most people were confined to the family farm or working in a small shop or later in a factory in town. They were also unable to travel far beyond their immediate communities. Communication technologies did not yet give them the ability to learn much about the world beyond their own communities, except perhaps through newspaper accounts or secondhand information that trickled in weeks or months after developments occurred elsewhere.

Flash forward two centuries and consider how technology, in the words of American historian Thomas L. Haskell, "change[d] the moral universe in which we live."[41] In a two-part 1985 essay on the "Origins of the Humanitarian Sensibility," Haskell observed how "our feeling of responsibility for the stranger's plight, though nowhere near strong enough to move us to action, is probably stronger today than it would have been before the airplane."[42] The growth of ubiquitous, affordable transportation and other technological capabilities—most notably widespread, instantaneous communication and information

transmission—has expanded our moral universe. Haskell argued the following:

> Technological innovation can perform this startling feat, because it supplies us with new ways of acting at a distance and new ways of influencing future events and thereby imposes on us new occasions for the attribution of responsibility and guilt. In short, new techniques, or ways of intervening in the course of events, can change the conventional limits within which we feel responsible enough to act.[43]

By constantly expanding the horizons of our moral universe in this fashion, technology expands our humanitarian sensibility. It enables us to be more worldly, cosmopolitan, and compassionate. "The Humanist Manifesto," originally published by the American Humanist Association in 1933 and most recently updated in 2003, asserts that humanists "ground values in human welfare shaped by human circumstances, interests, and concerns and extended to the global ecosystem and beyond. We are committed to treating each person as having inherent worth and dignity, and to making informed choices in a context of freedom consonant with responsibility."[44]

This is a noble vision of life and living, but it should also be clear why innovation is central to that humanist narrative. Innovation is central to human betterment not simply because it betters *us*, but because it allows us to better the lot of our fellow humans. Innovation expands our responsibility for each other and allows us to better act upon our "fellow-feeling." "Progress consists of deploying knowledge to allow all of humankind to flourish in the same way that each of us seeks to flourish," notes Steven Pinker.[45] Technology helps us achieve this goal and enhances our humanity by helping us understand and address the

needs of our fellow humans across the globe, many of whom we will never meet. Again, what could be more humanist than that?

Making Permissionless Innovation and Responsible Innovation Compatible

Although humanist critiques of technology often go much too far, innovation's defenders should take seriously calls by critics to incorporate other values or rights into the process of technological development and governance. Concerns about the safety, security, and privacy-related implications of many emerging technologies are particularly notable in this regard because those issues pervade almost every emerging technology sector today.

Those concerns have led to a growing intellectual movement known as "responsible research and innovation" (RRI). Although this movement is more widespread in Europe, it is growing in the United States, but sometimes under the auspices of "technology ethics" or other labels.[46] In the United States, the term "upstream governance" is often used to refer to largely the same thing.[47] A great deal of work by STS scholars today revolves around these themes of "responsible innovation," "ethical innovation," and "upstream governance."[48] Definitions are still evolving, but a 2011 article by René von Schomberg, a leader in the RRI movement and the Director General for Research at the European Commission, defined RRI as follows:

> A transparent, interactive process by which societal actors and innovators become mutually responsive to each other with a view to the (ethical) acceptability, sustainability and societal desirability of the innovation process and its marketable products (in order to allow a proper embedding of scientific and technological advances in our society).[49]

Other scholars define RRI more simply, saying that it comes down to "taking care of the future through collective stewardship of science and innovation in the present."[50] Practically speaking, this definition means anticipating the potential adverse consequences associated with technological change and seeking to somehow mitigate them through some form of upstream governance.

In a sense, RRI is just an extension of corporate social responsibility (CSR), a widely discussed but quite amorphous concept in the United States and abroad. It is not really clear what CSR means in many contexts, or that it even works that well in practice.[51] Regardless, CSR has become a major part of modern business practices and decisionmaking. RRI builds on CSR, but RRI is more squarely focused on addressing the potential risks associated with specific technologies or technological processes. In 1970, the Nobel Prize–winning economist Milton Friedman observed that discussions about CSR "are notable for their analytical looseness and lack of rigor."[52] His statement is still somewhat true for CSR today, and it is especially true for RRI. Both concepts remain open to differing interpretations and incorporate many distinct values that vary by context. At root, however, what RRI and CSR have in common is the belief that, whatever those responsible values are, they should be baked in early during product decisionmaking and design.

At first blush, it may seem as if permissionless and responsible innovation are fundamentally at odds. To the contrary, RRI can very much be part of a policy regime that adopts permissionless innovation as its general tech policy default. These concepts can coexist so long as policymakers and RRI advocates are willing to think more broadly about what the term "governance" means as applied to technological processes. "Governance" can mean more than just formal regulation by legislatures, administrative

agencies, or other public bodies. Governance can also describe a much broader universe of norms and rules that are established and enforced by a wide variety of people (or groups of people) in a wide variety of ways.

When we consider questions of *technological* governance—and specifically the notion of anticipatory governance, which is a prominent feature of RRI discussions—it helps to specify whether we are speaking of governance in a broad or narrow sense. Whether consciously or not, RRI scholars and advocates often fail to make clear precisely what type of governance they desire. This distinction is important because if anticipatory governance involves the formal application of the precautionary principle by force of law, it will be a deal-breaker for many innovation advocates. Banning innovative acts on the basis of fears about hypothetical worst-case scenarios means that many best-case scenarios can never come about. It forecloses trial-and-error experimentation that could bring about life-enriching services and applications. For example, many tech critics suggest that robotic technologies should be preemptively regulated based on a host of worst-case *Terminator*-esque scenarios about killer machines or AI run amok. Those far-fetched fantasies make for great sci-fi stories, but meanwhile back in the real world, robotic exoskeletons are helping people with spinal injuries walk again, and inventors are working to create autonomous vehicle technologies that could save countless lives in the future. That is why worst-case thinking should not guide policy and why innovation should be innocent until proven guilty.

Precaution can be pursued through less restrictive approaches, however. Some government agencies allow innovations to be released into the wild but in accordance with established safety standards and with a recall regime for defective or unsafe products. For example, this is the way the National Highway Traffic

Safety Administration addresses motor vehicle safety and the Consumer Product Safety Commission deals with unsafe devices. This represents a softer form of precaution relative to harder constructions of the precautionary principle, such as outright bans.[53] Even less restrictive but still precautionary in orientation would be a mandatory labeling law or a government-led risk reduction educational campaign. The Food and Drug Administration uses that approach in many instances. As noted earlier and discussed at greater length later, soft-law governance approaches are also used in many sectors today. Soft law includes various tools and methods—multistakeholder processes, agency guidance documents, collaborative best practices, industry standards and self-regulation, and so on—that establish expectations about technological development or use but that lack the same level of enforcement that accompanies hard-law enactments.

In other words, there exists a broad spectrum of governance options for new technologies, and it is important to specify what sort of approach we are talking about when debating these issues. We will return to soft-law mechanisms in the next chapter and explain how they can help us find a sensible balance in tech governance discussions.

Responsible Innovation without the Precautionary Principle

Exactly how much formal upstream governance of a *regulatory* nature does the responsible innovation movement recommend? It is often not very clear. Many responsible innovation advocates seem sympathetic to policies based on the precautionary principle and highly skeptical of the wisdom of permissionless innovation as a policy default. Yet most of them never spell out the exact relationship between RRI and the precautionary principle as a matter of public policy.

Some advocates of responsible innovation argue that the focus "is more on mitigating wider societal long-term risks and so [RRI] favors incremental rather than radical innovation."[54] That focus suggests a closer connection between RRI and a formal application of the precautionary principle for emerging technologies. A 2015 Brookings Institution white paper by Walter D. Valdivia and David H. Guston provides a more concrete answer to this question. Valdivia and Guston insist that responsible innovation "is not a doctrine of regulation and much less an instantiation of the precautionary principle; the actions it recommends do not seek to slow down innovation because they do not constrain the set of options for researchers and businesses, they expand it."[55]

Unfortunately, this demarcation between the general notion of responsible innovation and the formal application of the precautionary principle is not nearly so well defined in most RRI literature. On the rare occasions when RRI proponents *do* define the line between them, the proponents often introduce other terms that are equally amorphous. Even Valdivia and Guston fall prey to this tendency when they go on to suggest that responsible innovation "considers innovation inherent to democratic life and recognizes the role of innovation in the social order and prosperity," but that RRI advocates desire "a governance of innovation where that choice is more consonant with democratic principles."[56] The problem is that this notion simply shifts the definitional challenge away from defining "responsible innovation" and toward a debate about how we define "democratic life" and "democratic principles" in any given context.

The RRI literature is rife with ambiguous terms such as those and many others, such as "the public interest," and yet these advocates consistently lack precision regarding what they mean and how their claims can be translated into concrete governance principles or policies.[57] For example, does making innovation

more "consonant with democratic principles" mean that each new technology is somehow subjected to a formal vote before it gets released? If so, it would be hard to imagine many important innovations ever seeing the light of day.

Even if they cannot be nailed down on the precise applicability of the precautionary principle within technology policy debates, most RRI advocates would reject permissionless innovation as a suitable default position. Some of them even imply that permissionless innovation is synonymous with anarchy or a complete disregard for human rights. But that stance is ludicrous. For my book about permissionless innovation, I surveyed almost countless essays and articles that cite the phrase. Not once did I see any advocate of permissionless innovation going to such extremes as these critics suggest. Perhaps even more surprising is that most advocates of permissionless innovation rarely propose abolishing laws or agencies that currently oversee existing technologies or sectors.

What permissionless innovation advocates generally are advancing is the notion that new ideas deserve a fair shake or that *entrepreneurs and innovations should generally be considered innocent until proven guilty*. Permissionless innovation means giving innovators a bit more breathing room and avoiding a knee-jerk rush to regulate the new and the different. It means innovators should have a green light to experiment with those new and different ideas unless we can agree that a compelling reason exists to disallow trial and error as the basis of innovation policy.

The precautionary principle, by contrast, recommends keeping the light red until innovators can prove that new products and services are perfectly safe, however that is defined. But there are many points along the spectrum between these two policy postures and even many values shared by advocates of both

perspectives. The difference will often come down to the processes we use to address the areas of difference.

A Willingness to Compromise

Properly understood, responsible innovation can be compatible with the permissionless governance vision and policy regimes that have the freedom to innovate as their operational default. To achieve that synthesis, however, those on both sides must agree to some compromises. Advocates from different perspectives need to be open to learning from each other and willing to take the other's concerns seriously. Flexibility is essential.

To begin, RRI advocates need to appreciate how they can accomplish a great deal of good even in the absence of formal regulatory action. Their first instinct should not be to decry permissionless innovation advocates as a bunch of uncaring anarchists. If that is their starting point in conversations about emerging tech, RRI advocates will be missing opportunities to work with diverse parties and instill wise principles into various technological development processes. Such a move would be particularly misguided during a time when the pacing problem has become an undeniable reality and has made traditional hard-law efforts more difficult.

If they hope to get some (or perhaps any) of the values and procedures that they care about incorporated into technological development processes, RRI advocates will need to be open to the idea that perhaps the only way to do so will be through less formal procedures, precisely because law will likely lag so far behind marketplace developments. They should also appreciate the limitations of traditional regulatory approaches and the deleterious effects those regimes have sometimes had on innovation and competition. Again, that will require compromise.

Many such scholars now speak of the need for new forms of technological governance "that move beyond traditional command-and-control policymaking and enforcement to improve the effectiveness and legitimacy of regulation."[58] "A good governance approach," notes Schomberg, "might be one which allows flexibility in responding to new developments."[59] He writes:

> The power of governments is arguably limited by their dependence on the insights and cooperation of societal actors when it comes to the governance of new technologies: the development of a code of conduct, then, is one of their few options for intervening in a timely and responsible manner.[60]

That sort of thinking will be required among RRI advocates who hope to find common ground with permissionless innovation advocates.

Permissionless innovation advocates will need to be open to new ideas and perspectives, too. If the first instinct among them is to dismiss the entire RRI movement as little more than repackaged Luddism, hell-bent on derailing all the great inventions of the future, then permissionless innovation advocates are foolishly forgoing the chance to work with a diverse group of well-intentioned scholars and stakeholders who could ensure that new products and services gain more widespread acceptance and public trust. More practically, those who support permissionless innovation would be wise to accept that, although technological innovation is oftentimes outpacing the ability of government to keep up, well-established regulatory regimes or agencies are not necessarily going away any time soon. To repeat, few technocratic laws or regulatory bodies have been liberalized or eliminated in recent memory. It is unlikely that trend will reverse any time soon.

Thus, taking responsible innovation priorities seriously—and finding flexible ways to instill them in the development process—offers permissionless innovation advocates a chance to forge a rough peace with policymakers and issue advocates who often just want to have a small say in how technological processes are unfolding. But, if regulators seek to have a *big* say in such matters—namely, in the form of heavy-handed, preemptive restrictions—then major policy fights will no doubt ensue.

But responsible innovation advocates must certainly understand that the era of technocratic, overly bureaucratic, top-down, command-and-control regulation is being challenged by new realities.[61] Philip Weiser notes that "[t]he traditional model of regulation is coming under strain in the face of increasing globalization and technological change," and, therefore, governments must think and act differently than they did in the past.[62] "The new information environment," argues Taylor Owen, "may require states to adopt some characteristics of start-ups."[63] Similarly, Juma hoped to see "entrepreneurialism exercised in the public arena."[64]

The next chapter explores how entrepreneurial governance approaches are emerging today in the form of soft-law mechanisms. These mechanisms offer the greatest hope for compromise and sensible governance of emerging technology.

7

SOFT LAW AND THE FUTURE OF TECHNOLOGICAL GOVERNANCE

The previous chapter concluded with a call for compromise between responsible and permissionless approaches to innovation. This chapter shows how, as a practical matter, such compromises are already being negotiated in many policy deliberations about emerging technology governance through what has come to be known as soft law.

Soft law represents a messy amalgam of many different governance approaches that will often leave all sides somewhat dissatisfied because they will not get everything they want. That dissatisfaction, however, might be the best thing going for soft law. Much as Winston Churchill once famously said that democracy represented "the worst form of Government except for all those other forms that have been tried from time to time," it may be the case that soft law represents the worst form of technological governance, except for all those tried before.

This chapter explores what is meant by soft law. It explains how soft law is already becoming the dominant approach for modern technological governance (at least in the United States), and it shows how soft law is supplementing existing legal and regulatory hard-law remedies. Also explored will be the role that various other expert organizations play in facilitating technological governance today, as well as the importance of stepped-up risk education efforts in addressing various concerns. Finally, I discuss the shortcomings of soft law and more challenging issues regarding technological risks that might require more serious regulatory oversight.

Soft Law: The Basics

Soft-law mechanisms have already been mentioned many times throughout this book. Soft law includes a wide variety of informal, collaborative, and constantly evolving governance mechanisms that differ from hard law in that they lack the same degree of enforceability.[1] These soft-law systems and processes are multiplying at every level of government today: federal, state, local, and even global.

The easiest way to define soft law is to explain what it isn't: Soft law is not hard law. Soft law builds upon and operates in the shadow of hard law. But soft law lacks the same degree of formality that hard law possesses. Scholars who study such governance mechanisms note that soft law is used "as a shorthand term to cover a variety of nonbinding norms and techniques for implementing them."[2] Although some consider this informality and nonbinding nature to be a weakness of soft law, it also serves as a strength. Compared with hard law, soft law can be more rapidly and flexibly adapted to suit new circumstances and address complex technological governance challenges.[3]

Chapter 4 identified the many reasons that evasive entrepreneurialism is on the rise today. To reiterate, those problems include the accumulation of laws and regulations that are increasingly out of touch with common sense, a chronic inability of government institutions to reform broken governance systems or adapt regulatory policies to new realities, and the unrelenting reality of the pacing problem, which makes it increasingly hard for public policy to keep pace with the rate of technological change.

Those reasons are some of the same ones causing soft law to ascend. "Reinventing government" is a phrase that has been used widely in the past, but in light of these new realities, the need to get serious about the reinvention of governance processes is more urgent than ever. This need particularly applies to fast-evolving sectors in which "there is a growing consensus that traditional government regulation is not sufficient for the oversight of emerging technologies," as Wallach and Marchant argue.[4]

For those reasons, many tech policy scholars and governance experts have begun identifying new models that can help address pressing policy concerns without resorting to build-and-freeze hard-law approaches that are no longer working effectively, or which are inappropriate for newer, fast-moving technologies and sectors for which we hope to see accelerated innovation opportunities.

"Co-regulation" is a related term used to describe the give-and-take between regulators and regulated parties that is often a part of soft-law processes.[5] Phil Weiser describes the co-regulation approach as one in which an agency integrates "its efforts with private bodies with expertise in the field" and in which "integration involves the explicit embrace, oversight, and in which enforcement of actions by private bodies" to solve difficult problems outside traditional regulatory processes.[6]

An even broader term for these new approaches is "flexible regulation." A diverse body of scholarship has developed over the past quarter century that outlines flexible approaches to governance in various contexts. In her book *Innovation and the State*, Cristie Ford notes that these models go by many different names—reflexive law, management-based regulation, experimental governance, principles-based regulation, meta-regulation, and others—but that they all share a commitment to move away from the overly rigid and inefficient regulatory methods of the past.[7]

The consultancy firm Deloitte has produced several important reports on the need to reinvent governance frameworks for emerging technologies using flexible or co-regulatory approaches. It argues that these methods are essential today because "[i]nnovative technologies and new business models can catch regulators by surprise."[8] Deloitte recommends a set of five approaches to guide the future of emerging technology policy:[9]

- *Adaptive regulation*: Shift from "regulate and forget" to a responsive, iterative approach.
- *Regulatory sandboxes*: Prototype and test new approaches by creating sandboxes and accelerators, which are mechanisms that allow regulators to experiment with alternative and more flexible governance schemes without having to abandon laws or regulations altogether.
- *Outcome-based regulation*: Focus on results and performance rather than form.
- *Risk-weighted regulation*: Move from one-size-fits-all regulation to a data-driven, segmented approach.
- *Collaborative regulation*: Align regulation nationally and internationally by engaging a broader set of players across the ecosystem.

Many regulatory agencies are already using those approaches to cope with the pace of technological change and address thorny governance issues. But government agencies are not the only ones.

In a 2019 law article with Ryan Hagemann and Jennifer Huddleston, I cataloged a long list of soft-law methods and various case studies describing how such mechanisms are being tapped by government bodies today to deal with fast-moving technologies.[10] A partial inventory of soft-law methods includes multistakeholder processes, industry best practices or codes of conduct, technical standards, private certifications, agency workshops and guidance documents, informal negotiations, and education and awareness efforts. Again, this list of soft-law mechanisms is amorphous and ever-changing. Moreover, many of those soft-law methods and processes are used in conjunction with hard-law methods.

Multistakeholder processes are a particularly important type of co-regulatory soft law, and they have been the cornerstone of America's digital economy policy efforts for two decades.[11] In July 1997, the Clinton administration released *The Framework for Global Electronic Commerce*, a statement of the administration's principles and policy objectives toward the internet.[12] The document said that "governments should encourage industry self-regulation and private sector leadership where possible" and "avoid undue restrictions on electronic commerce."[13] The co-regulatory multistakeholder model promoted by the *Framework* was instrumental in helping transition internet governance and policymaking efforts from the National Science Foundation to the National Telecommunications and Information Administration (NTIA) and the Internet Corporation of Assigned Names and Numbers.[14] That collaborative governance vision has been the cornerstone of internet policy ever since.

Through collaborative efforts, regulators have been working with innovators and various civil society organizations to formalize "privacy-by-design," "safety-by-design," and "security-by-design" efforts.[15] Through ongoing conferences, meetings, negotiations, and guidance documents, these parties have hammered out best practices that bake important values and safeguards directly into the product design process. This work is a way of introducing what some call anticipatory ethics into the early stages of technological developments.[16]

For example, over the past two decades, soft-law mechanisms have been used extensively to address concerns about online safety and youth activities on the internet. Between 2000 and 2010 alone, six major online safety task forces or blue-ribbon commissions were formed to study online safety issues and consider what should be done to address them.[17] Three of those task forces were convened by the U.S. government, and the British government commissioned another. Two additional task forces were formed through universities and private associations during this period. Each of those six task forces was made up of, or received input from, a diverse set of experts from academia and think tanks, corporations and professional trade associations, advocacy organizations, and various government agencies. In other words, they were multistakeholder processes. The task forces recommended a variety of best practices, educational approaches, and technological empowerment solutions to address various safety concerns.

More recently, multistakeholder processes have formulated privacy, safety, and cybersecurity-related best practices. Many of the meetings were convened by the U.S. Department of Commerce (the NTIA in particular); the White House Office of Science and Technology Policy; and a wide variety of federal regulatory agencies, including the FTC, FDA, FAA, and FCC.

Those multistakeholder efforts and agency best-practice reports have contained assorted "responsible innovation" principles for technologies as wide ranging as the following:

- Big data, machine learning, and artificial intelligence[18]
- The Internet of Things (i.e., internet-enabled devices and applications)[19]
- Online advertising practices[20]
- Autonomous vehicles policy[21]
- Motor vehicle cybersecurity[22]
- Cybersecurity of advanced medical devices[23]
- Facial recognition technologies[24]
- Health and medical smartphone applications[25]
- Medical advertising on social media platforms[26]
- Mobile phone privacy disclosures[27] and mobile applications for children[28]
- 3D-printed medical devices[29]
- Small unmanned aircraft systems (i.e., drones)[30]

This list just scratches the surface of soft-law and multistakeholder processes. Moreover, the recommendations flowing out of those soft-law efforts can be quite detailed and are too numerous and context-specific to itemize here. But to illustrate, one common best practice recommended in many of those efforts involves devising appropriate data collection and storage procedures. As part of various soft-law efforts, innovators are typically encouraged to use commonly accepted encryption techniques and ensure that data are properly handled, used only for clearly specified and sensible purposes, and then deleted after a certain period of time. In some cases, technical specifications and procedures are worked out during multistakeholder negotiations. In other cases, those tasks are left to industry bodies or third-party accreditors to address and enforce.

The bottom line is that, just as software is eating the world, soft law is now eating the world of technological governance.[31] Strangely, however, many tech critics or responsible research advocates rarely mention such efforts in their writings. Perhaps they are simply unaware of the many ways in which the principles they advocate already infuse multistakeholder processes and soft-law efforts. But that seems unlikely because those efforts are widely discussed and reported. The more likely reason they ignore soft-law efforts is that they probably do not believe that those efforts are comprehensive or stringent enough. To the extent that they discuss soft-law efforts at all, tech critics or responsible research and innovation (RRI) scholars will often say that such initiatives lack teeth and that anything short of a full-blown regulatory regime (or significant expansion of an existing one) is insufficient to address their concerns. It seems clear that, for many tech critics, only a comprehensive federal law and corresponding regulatory regime for each emerging technology sector will be enough.

That perspective is unfortunate, and if it is the line in the sand that RRI scholars wish to draw, then they will be left with little wiggle room in conversations about governance options for emerging technologies. The prospects for comprehensive regulation of most emerging technologies are dim; at the very least, comprehensive regulation will take many years to get in place. Again, the unrelenting pace of technological change means the clock is always ticking. In many cases, the law would be outdated by the time it got on the books. Practically speaking, therefore, it is a mistake to make the perfect the enemy of the good, and it would be wise to have backup governance plans that are more adaptable in this era of rapid technological change. In this sense, soft-law efforts might be

viewed as the minimum necessary governance needed to address various social needs and values. If RRI principles already infuse soft-law governance processes, and if they can be improved in a flexible way to adapt to new challenges, that means a rough-and-ready set of principles and policies are in place while more formal rules are pursued, if they continue to be needed at all.

Some RRI scholars, and potentially many innovators, will decry the messy, uncertain nature of soft-law governance processes. For many such scholars and tech critics, soft law is simply not enough. But there is a great deal to be said for the way soft-law mechanisms have already started adapting in a dynamic, iterative fashion to deal with rapid changes in various fields. Additionally, although soft-law systems may embody various uncertainties, they still often provide *more* governance than hard law. As Chapter 4 documented, hard-law systems regularly struggle to adapt to changing technological realities, even though that reluctance to change creates serious problems and makes those governance systems less effective (and potentially more likely to be evaded by innovators). If soft law offers a better chance than hard law of getting some principles and values baked into technological design processes, then RRI supporters should acknowledge that potential benefit and build on it.

Finally, it is vital to understand that soft law does not develop in a vacuum. There is not an either-or choice between soft law and hard law so much as there is a constantly sliding scale of governance options that ideally are used in cooperation with each other. In most soft-law schemes, government officials initiate the process or are at least a major part of it. Moreover, soft law often builds upon, or operates alongside, many other governance

mechanisms, including many that are reactive and remedial in character. These mechanisms include the following:

- Federal and state consumer protection agencies (such as the FTC), which police unfair and deceptive practices and other harms
- Courts and common law, including legal solutions like product liability, negligence, design defects law, failure to warn, breach of warranty, contract law, and other assorted torts and class action claims
- Insurance markets, which serve as risk calibrators and correctional mechanisms
- Third-party accreditation and standard-setting bodies, discussed later
- Education and awareness efforts, both by government bodies and third parties, also as discussed later
- Social norms and reputational effects, especially the growing importance of reputational feedback mechanisms[32]
- Media, academic institutions, nonprofit advocacy groups, and the general public, all of which can put pressure on technology developers
- New entry and competition combined with the power of consumer choice

Only by taking into account the full range of players and activities at work can we develop a more robust understanding of how technology is actually governed in our modern world. I suspect that many RRI scholars *do* appreciate these other factors, even though they sometimes fail to account for all of them in their writing and advocacy. But, again, many of those advocates generally do not favor the remedial, ex post nature of some of these governance tools and will continue to insist that more ex

ante anticipatory planning must be at the heart of technological design and development processes.

In reality, a mix of these two approaches is already at work today in soft-law processes and will likely continue to dominate governance well into the future. As long as anticipatory efforts do not become formal regulatory proposals, this mix of responsible innovation governance tools and methods should be embraced by a diverse array of scholars and innovators alike.

Risk Education

Educational approaches are a particularly important part of the soft-law toolkit, yet they are often underappreciated. With traditional regulatory approaches being strained by new realities, public awareness campaigns and risk communication efforts can be an effective way of providing citizens with better information about some of the risks associated with new technologies they increasingly need and demand. Improved risk education can also help address the problem of technological illiteracy that can fuel the sort of technopanics discussed throughout this book.

Social scientists frequently debate the degree to which scientific or technological illiteracy among the general public ends up driving poor decisionmaking—both by individuals and policymakers. The *knowledge deficit model* holds that public skepticism about science or hostility to certain technologies is related to the level of public ignorance or misunderstanding of the technology at hand. In theory, better education about such matters should correct that ignorance and reduce opposition to science and technology.

Other social scientists and economists contend that a certain degree of rational ignorance about new scientific or technological developments exists. Because individuals are limited in time

and ability to process highly technical matters, they rely on cognitive shortcuts or trusted sources to inform their attitudes and decisions about science and technology. That model is sometimes referred to as the low-information rationality or bounded rationality model. In either case, improved risk communication and technological literacy efforts can help better inform individuals about complex science and technology issues.

Consider the GMO example discussed earlier. When public understanding of risk tradeoffs is based on myths or misperceptions, it can result in backlashes and technopanic-based policymaking about genetic modification.[33] Better education and risk communications can help reverse that problem.

Regulators can play an important role in this regard as public risk educators. Again, consider digital technologies and online safety. As noted earlier, soft-law mechanisms have been tapped repeatedly to deal with various concerns related to online safety, harassment, and hate speech. The many task forces and blue-ribbon commissions that were organized to address these issues generally agreed that educational approaches would be both more effective and less restrictive than regulatory solutions.[34] The accelerating pace of technological change was a primary factor cited by most of the task forces when reaching that conclusion. More specifically, the task forces outlined how a combination of media literacy, awareness-building efforts, public service announcements, targeted intervention techniques, and better mentoring and parenting strategies could help prepare youngsters to be better digital citizens and to better adapt to changes in technology than could top-down regulations. Many government agencies, such as the FTC and FCC, already work together and with technology developers to facilitate education and awareness efforts about online safety and security threats and best practices. This model is a good one for many other fast-moving,

hard-to-classify technologies such as the Internet of Things and artificial intelligence.

Another example of how education can be helpful in communicating risks involves the Food and Drug Administration. Risk communication and health literacy are already important parts of the FDA's mission, even though they do not receive much attention and the agency does not allocate nearly as much resources to them compared to traditional regulatory responsibilities.[35] Risk regulation has always been the primary focus of FDA efforts to ensure safe and effective drugs and medical devices. It will likely remain that way. But as noted in earlier chapters, the old build-and-freeze model of regulation is increasingly under strain from new realities.

Stepped-up risk education efforts can help fill the emerging gaps. In its 2009 *Strategic Plan for Risk Communication*[36] as well as its 2011 report *Communicating Risks and Benefits: An Evidence-Based User's Guide*,[37] the FDA provided a roadmap for what a more comprehensive risk education campaign would entail. The FDA also engages in various product labeling efforts as well as other public education campaigns and strategies.[38] Yet those efforts have always been secondary for the agency, which has instead focused on trying to preemptively guarantee the safety and efficacy of drugs and devices. And much of the education the FDA does is basically explaining to companies and the public how to comply with its voluminous body of regulation.

A more robust focus on risk education would aim to better inform citizens about the relative risk tradeoffs they face with new technologies and technological capabilities.[39] Such risk education should focus on both the general public and the innovators who are providing new devices and treatments. These approaches will be essential in a world of highly personalized

medicine, where citizens are more empowered to make their own wellness decisions. As Chapter 2 made clear, new technological capabilities are already giving people more options about how to address their health or augment their abilities using digital health technologies, 3D printers, genetic technologies, and biohacking techniques. Stepped-up risk education and health literacy are desperately needed as these capabilities accelerate and outstrip the ability of traditional laws and regulations to keep up with breaking developments and a more technologically empowered public.[40]

Some skeptics will argue that government will often get things wrong when it engages in risk education or health literacy. For example, many health experts criticize the U.S. Department of Agriculture's Food Pyramid because of specific dietary recommendations that those experts feel have undermined public health.[41] To be sure, government health officials—and officials engaging in risk education in other contexts—will not always get it right. Government-led risk education efforts must pivot and adapt to changing technical and social realities or else they will be ignored. Of course, no one is forced to follow the government's advice. Moreover, governments are not the only ones doing such education. As noted next, many other organizations are helping to advance the understanding of various risks and provide guidelines for acceptable development and use of new technological capabilities.

The Importance of Professional Associations and Ethical Codes

Professional organizations, trade associations, and various consortia can and do develop guidelines and codes of ethics to address the responsible development or appropriate use of various emerging technologies. Such organizations can serve as independent

standard-setting bodies and can help hold innovators accountable by designing guidelines and best practices established through soft-law processes.

Various trade associations have already worked with government agencies to formulate some of the best practices and codes of conduct documented earlier. Other organizations have focused on developing high-level codes of professional conduct for innovators in their sectors. Some of the most notable examples involve the Association of Computing Machinery (ACM), the Institute of Electrical and Electronics Engineers (IEEE), the International Organization for Standardization (ISO), and UL (which was previously known as Underwriters Laboratories), among others.

For example, the ACM developed a Code of Ethics and Professional Conduct in the early 1970s, refined it in the early 1990s, and then updated it again just recently in 2018.[42] Each iteration of the ACM Code reflected ongoing technological developments, from the mainframe era to the PC and internet revolution and on through today's machine learning and AI era. The latest version of the ACM Code "affirms an obligation of computing professionals, both individually and collectively, to use their skills for the benefit of society, its members, and the environment surrounding them" and insists that computing professionals "should consider whether the results of their efforts will respect diversity, will be used in socially responsible ways, will meet social needs, and will be broadly accessible."[43] The document also stresses the following:

> An essential aim of computing professionals is to minimize negative consequences of computing, including threats to health, safety, personal security, and privacy. When the interests of multiple groups conflict, the needs of those less advantaged should be given increased attention and priority.[44]

Other organizations formulate more targeted or applied best practices and codes of conduct. The field of artificial intelligence and machine learning is a particularly good example. Several initiatives are already underway:

- The IEEE's Ethically Aligned Design project is an effort to craft "A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems."[45] With more than 420,000 members in more than 160 countries, IEEE boasts of being "the world's largest technical professional organization dedicated to advancing technology for the benefit of humanity."[46] IEEE's new effort seeks to incorporate into AI design five key principles that involve the protection of human rights, better well-being metrics, designer accountability, systems transparency, and efforts to minimize misuse of these technologies. The second iteration of the group's report was 263 pages long and contained a litany of recommended best practices to satisfy each of those objectives. The effort included almost a dozen working groups with detailed reports and a variety of certification proposals as well.
- The Partnership on AI began as an industry-led effort formed by Apple, Amazon, Google, Facebook, IBM, and Microsoft, but it has grown to include more than 80 members from industry, civil society organizations, academic institutions, and other groups. The Partnership is billed as a multistakeholder organization that brings those diverse groups together "to study and formulate best practices on AI, to advance the public's understanding of AI, and to provide a platform for open collaboration between all those involved in, and

affected by, the development and deployment of AI technologies."[47]

- OpenAI is a nonprofit research organization created in 2015 with seed money from notable tech innovators and investors like Elon Musk of Tesla (and formerly PayPal), Sam Altman of Y Combinator, venture capitalist Peter Thiel (also formerly of PayPal), Reid Hoffman of LinkedIn, and others. OpenAI publishes research reports discussing how to make sure that AI development "is used for the benefit of all, and to avoid enabling uses of AI or (artificial general intelligence) that harm humanity" and to ensure it does not become "a competitive race without time for adequate safety precautions."[48] OpenAI is also a member of the Partnership on AI.
- In late 2016, the British Standards Institute published a "Guide to the Ethical Design and Application of Robots and Robotic Systems."[49] Developed by a committee of scientists, academics, ethicists, and philosophers, the guide "recognizes that potential ethical hazards arise from the growing number of robots and autonomous systems being used in everyday life" and aims to "eliminate or reduce the risks associated with these ethical hazards to an acceptable level." Specifically, the guide's protective measures create best practices for the safe design and use of robotic applications in a wide range of fields, from industrial services to personal care to medical services.[50]
- The ISO is a global standards–making body that was formed in 1946 and continues to play an important role in establishing international norms for emerging technologies. The ISO "is an independent, non-governmental international organization with a

membership of 163 national standards bodies"[51] that seeks to build global consensus through multistakeholder efforts. The ISO uses dozens of technical committees that include global experts in diverse fields, such as industry, consumer associations, academia, nongovernmental organizations, and governments.[52] It has already played an important role in formulating global best practices for robotics and AI-based applications. In 2014, for example, the ISO crafted requirements and guidelines "for the inherently safe design, protective measures, and information for use of personal care robots."[53] That standard is just one of dozens of robotics-related ones that ISO has published.[54]

Other industry groups and professional societies in fields as diverse as drones and biotechnology are developing guidelines and best practices for their sectors. Such efforts often complement governmental efforts to explore issues surrounding emerging technologies.

It goes without saying that codes of conduct, voluntary standards, and professional ethical codes are not cure-alls for the problems associated with technological development, and they cannot magically ensure that all innovation will be responsible. Additional efforts will sometimes be needed, as will be discussed later. But efforts such as those described here can go a long way toward improving accountability and responsibility among various emerging technology companies and individual innovators. Standards, codes, ethical guidelines, and multistakeholder collaborations create powerful social norms and expectations that are often equally important as, or even *more* important than, what laws and regulations might seek to accomplish.[55] There are powerful reputational factors at work in every sector that—when

combined with efforts such as these—create a baseline of accepted practice. These efforts are also likely to get more initial buy-in among private innovators, at least compared to heavy-handed regulatory proposals. Finally, these efforts deserve more attention if for no other reason than the continuing reality of the pacing problem. Soft-law mechanisms will always be easier to adopt and adapt as new circumstances demand.

Critics might also insist that, even if they do *some* good, privately negotiated best practices or codes of conduct will be hard to coordinate and enforce both domestically and abroad. This concern is valid, but it may be intractable in light of the different values that various countries and cultures possess.

To help address this problem, Gary Marchant and Wendell Wallach propose the formation of what they call governance coordinating committees (GCCs). GCCs would help coordinate technological governance efforts among governments, industry, civil society organizations, and other interested stakeholders in fast-moving emerging technology sectors.[56] Because "no single entity is capable of fully governing any of these multifaceted and rapidly developing fields and the innovative tools and techniques they produce," they suggest that GCCs could act as a sort of "issue manager" or "orchestra conductor" that would "attempt to harmonize and integrate the various governance approaches that have been implemented or proposed."[57] They have also called for the formation of an International Congress for the Governance of AI as "a first step in multistakeholder engagement over the challenges arising from these new technological fields."[58]

In essence, Marchant and Wallach are proposing the creation of what is commonly known in Europe as a *quango*, or quasi-autonomous nongovernmental organization. Quangos have been effective in Europe and some other areas in helping devise solutions to governance coordination challenges in technically

complicated fields. Like quangos, GCCs could help provide another mechanism whereby technological governance issues are addressed through ongoing collaboration among various parties, both domestically and globally. They could help craft or enforce voluntary best practices, or at least offer a forum for ongoing discussions around thorny issues. Difficult details and hard questions remain, including the following: how do GCCs get formed, who would be on them, and how would they be supported financially? Moreover, even to the extent that international consensus could be found on ethically complicated issues (like genetic modification of human embryos), how would a GCC be able to enforce restrictions across so many countries and cultures?[59] As a forum for conversation and collaboration, however, GCCs could still hold great promise and should be given greater consideration.

The Challenge of Defining Harm

Soft law cannot serve as a complete substitute for hard law. Some technological developments can give rise to significant harms or intractable problems that will sometimes require a heightened level of regulatory scrutiny and action. That fact does not mean hard-law solutions will always be completely effective in solving those problems, and at times hard law may create more problems in the process. Nonetheless, laws and regulations will sometimes need to be considered to help discourage the most serious harms associated with certain technological developments.

How are policymakers supposed to determine which technologies and theoretical harms deserve more regulatory scrutiny, and how should they determine the likelihood or measure the potential severity of the harms? These are notoriously hard questions to answer because there are many different ways to judge

what constitutes acceptable risk[60] or catastrophic risk.[61] Such questions also demand a fuller exploration of what theories of rights and responsibilities animate the discussion. Here I offer only a brief sketch of a much-needed theory of technological harm that can help us answer these questions. But even a robust theory would not be able to preemptively answer every question critics pose about the alleged dangers of various emerging technologies. A major point of my previous book was that many, if not most, of these questions can only be answered through real-world, trial-and-error experiences and responses.[62]

To understand why that is the case, again consider the many contentious debates about online safety and privacy. Many innovation policy squabbles—both today and in the past—have involved heated battles over what are best thought of as cognitive or psychological harms. Those harms are not to physical life and limb but rather are potential harms to one's feelings or cognitive processes or the creation of a general sense of unease.

Supposed informational harms were also at the heart of past policy skirmishes over indecent or obscene content. Raging debates surrounded these issues long before the internet came along, and policymakers imposed many censorial prohibitions on content creators and distributors in the name of upholding "community standards" and "public decency."[63] To be sure, a great many people thought that there was some harm to themselves, their children, or the public more generally because of such content. Yet large numbers of people disagreed with the proposed regulations and felt it was their right to consume whatever sort of content they desired. How should harm be calibrated when "objectionable content" is in the eye of the beholder? In this case, the First Amendment generally won out over time, and most content controls gradually went away. Even though some rules are still on the books today, they are largely ignored as

agencies and courts engage in the sort of rule departure discussed earlier. In essence, we have witnessed the end of traditional censorship efforts in the United States over the past two decades.[64]

Eye-of-the-beholder spats and calls for information controls have not gone away, however. Instead, they have moved into new areas. For example, subjective theories of informational harm have become a major fault line in debates over digital security and privacy. Tech critics often insist that new data collection and dissemination capabilities pose a threat to their security and privacy "rights." Others do not seem to understand what all the fuss is about and worry more about how their regulation might impede their "right" to collect and receive more information or speech, or perhaps their "right" to better service, greater convenience, or lower prices for important services. Whose rights should prevail, and are these really rights at all? In 2017, the Federal Trade Commission even launched an inquiry into the question of what constituted informational harms in the context of online data collection or data security incidents.[65] Among the filings that the agency received in this and related proceedings, little consensus existed regarding what constituted information rights or wrongs.[66]

These debates might never cease, because we might never be able to reach strong consensus regarding the nature and extent of such rights or determine how to protect those values.[67] Moreover, how those rights are conceived of varies widely by country, making global enforcement more challenging. Many European laws conceive of privacy as a "dignity right" that trumps most other economic and social values, including freedom of speech. The United States has taken a different approach. Privacy rights have been generally associated with other, more well-established rights that are more tangible in character, including the property

rights people hold in their bodies, their homes, their personal property, and their financial accounts. To the extent that there is an overarching information imperative in the United States, it has been shaped by the First Amendment to the Constitution, which generally disallows regulation of the collection or use of data with a few important exceptions (personal health and financial information, for example).

Privacy advocates in the United States insist that policymakers should mimic the European approach and adopt a comprehensive privacy law that might even incorporate a formal "privacy bill of rights." Throughout the Obama administration, privacy advocates pushed such efforts but failed to get any traction. At the time of this writing, however, America appears on its way to potentially advancing a new federal privacy framework, if for no other reason than to preempt a confusing patchwork of state privacy laws. Even if these measures pass, however, they will be severely challenged by all the same realities documented throughout this book. Government passage of major bills claiming to protect privacy does not necessarily mean such laws accomplish the goal envisioned. Technologies, individual desires, and societal values continue to evolve faster than laws in many instances. A great many Americans enjoy the benefits associated with data collection, including lower-cost services, various conveniences, expanded competition and choice, and other benefits, and they would likely choose to continue to share their data in exchange for those things. Sometimes social values and individual choices are as hard to control as technological change.

Soft law is preferable to hard law when consensus is elusive, as it is in many of these cases. An adaptive multistakeholder framework performs better than codified laws when harms are amorphous, speculative, or subjective. This reason is partially why various soft-law processes are being tapped more regularly. When coupled with

ex post judicial remedies, soft law continues to represent the better approach for most online safety, privacy, and security concerns, as well as a variety of other concerns about emerging technologies. The threat of hard law can sometimes help discourage the worst types of misbehavior by some actors, or at least encourage them to come to the table as part of a multistakeholder process and agree to commonly accepted best practices going forward.

When Soft Law Isn't Enough: Existential Risks

What about more serious alleged harms where widespread agreement exists that more should be done to preemptively address risks to life, limb, health, and so on? The most problematic category of such harms is often referred to as "existential" or "catastrophic" risks. As noted in Chapter 1, "existential" is a term some tech critics throw around far too casually when decrying a variety of innovations or particular companies they do not care for. We can dismiss assertions of existential threats when they are alleged for lesser matters, such as whether Facebook is destroying civilization as we know it. That sort of threat inflation cheapens the meaning of the term "existential"; there are no plausible mechanisms by which Facebook could pose such a threat. There may be legitimate existential threats out there that we *should* be spending more time addressing, but that threat probably isn't one of them.[68]

Nick Bostrom, Director of the Future of Humanity Institute at the University of Oxford, has written extensively about the dangers of "superintelligence" and what he calls the "vulnerable world hypothesis." What makes Bostrom's work distinctive among modern technology critics is his willingness to finish his sentences. That is, critics usually heap scorn on various technologies, but most do not follow through with concrete recommendations for what to do about their litany of woes. By contrast,

Bostrom provides a roadmap with various options about how to address the new technological risks he believes exist. This roadmap makes Bostrom's work deserving of greater attention because it signals what sort of regulatory approaches other critics and policymakers might eventually support.

"Our approach to existential risks cannot be one of trial-and-error," Bostrom argues, because with such risks, "[t]here is no opportunity to learn from errors."[69] In other words, some theoretical risks are so potentially catastrophic that permissionless innovation is no longer the optimal default for tech policy. Does that automatically mean that the precautionary principle should be the default? Not necessarily. As Bostrom himself notes, "stopping technological development would require something close to a cessation of inventive activity everywhere in the world. That is hardly realistic; and if it could be done, it would be extremely costly—to the point of constituting an existential catastrophe in its own right."[70]

On the other hand, Bostrom argues, "*limited* curtailments of inventive activities" might be a sensible policy.[71] That approach was adopted by governments to address the use of chemical weapons after World War I, and then nuclear proliferation after World War II. After the horrific uses of chemical weapons during World War I, the Geneva Protocol for the Prohibition of the Use in War of Asphyxiating, Poisonous or Other Gases, and of Bacteriological Methods of Warfare was formulated in 1925 to limit the uses of such weapons in future conflicts.[72] Later, after World War II, international treaties and other agreements were formulated that sought to limit the ability to possess or enrich uranium, or to traffic nuclear weapons. The Treaty on the Non-Proliferation of Nuclear Weapons (NPT) was created in 1968 to advance the peaceful uses of nuclear technology while seeking to limit the dangerous ones. The International Atomic Energy Agency (IAEA), formed a

decade earlier, helps advance this mission "to accelerate and enlarge the contribution of atomic energy to peace, health and prosperity throughout the world."[73]

This book is not the place for a comprehensive evaluation of the success of the Geneva Protocol for chemical weapons or the NPT and the IAEA for nuclear proliferation. Nonetheless, we can draw two high-level conclusions from these efforts. First, the worst fears about chemical and nuclear weapons have not come to pass. Although rogue actors still exist and develop such weapons, the most concerning applications of these technologies have been constrained for the most part. But how much of that success can be attributed to treaties and nonproliferation agreements versus the simple fact that it is costly to obtain and produce such weapons? If the cost and complexity of weaponization *were* the primary factors limiting the worst-case applications of chemical and nuclear technologies, what happens in a world in which newer dangerous technologies are cheaper, more widely available, and easier for greater numbers of people to access or develop?

That fear leads Bostrom to propose the "Principle of Differential Technological Development," which would

> Retard the development of dangerous and harmful technologies, especially ones that raise the level of existential risk; and accelerate the development of beneficial technologies, especially those that reduce the existential risks posed by nature or by other technologies.[74]

Although Bostrom admits that "correctly implementing differential technological development is a difficult strategic task," he believes "it is worth making the attempt" if for no other reason than "to buy a little time."[75] Alas, his proposed specific measures and countermeasures to mitigate vulnerabilities all

have rather serious tradeoffs and limitations. Bostrom recommends that we consider efforts that would do the following:

- Prevent dangerous information from spreading.
- Restrict access to requisite materials, instruments, and infrastructure.
- Deter potential evildoers by increasing the chance of their getting caught.
- Be more cautious and do more risk assessment work.
- Establish some kind of surveillance and enforcement mechanism that would make it possible to interdict attempts to carry out a destructive act.

In thinking about enforcement options, one must consider both the practicality and the wisdom of each approach. To move away from the theoretical and toward the practical, one can apply Bostrom's framework to modern existential threats that are commonly discussed, such as concerns about 3D-printed weapons or the development of robotics and AI technologies. Much like chemical and nuclear technologies before them, 3D printers, robots, and AI technologies already have many important peaceful and socially beneficial uses, and many more are sure to be developed. But, if used improperly, these technologies could produce horrific consequences. As Chapter 1 already noted, all one needs to do is read or watch just about any sci-fi book or show about robots or AI to see every worst-case scenario explored *ad nauseam*.

So, what should be done to prevent the rise of 3D-printed "ghost guns," "killer robots," and *Terminator*-esque scenarios? Although far-fetched, such occurrences are, at least, risks that might warrant some degree of precautionary regulation.

Returning to Bostrom's proposals for dealing with existential risks of this sort, the first of them—restricting the spread of dangerous

information about new technologies or technological capabilities—is probably the least feasible in a world of ubiquitous, low-cost information transmission. In addition, it is not wise to propose a global censorship regime in this regard because of the potential collateral damage it would have for beneficial types of information flows.

Bostrom's last suggestion—preventive policing through stronger interventions by various levels of government—raises new risks of its own. "If the continued survival of humanity depended on successfully imposing worldwide surveillance," responds Kelsey Piper of *Vox*, "I would expect the effort to lead to disastrous unintended consequences."[76] A mass surveillance apparatus would not necessarily guarantee workable containment solutions to the sort of disasters that Bostrom fears, but it certainly would open the door to a different type of disaster in the form of highly repressive state controls on communications, individual movement, and other activities. Even assuming we could look beyond the specter of mass surveillance in the name of reducing technological risk, the questions of cost and resource constraints remain a problem. Bostrom does not consider those downsides, however.[77] Finally, mass surveillance schemes could discourage research into a great many risk-*reducing* technological applications and thus undermine Bostrom's other goal of "accelerat[ing] the development of beneficial technologies, especially those that reduce the existential risks posed by nature or by other technologies."

That leaves his three other options, all of which have greater merit. To repeat, those options are as follow:

- Restrict access to requisite materials, instruments, and infrastructure.
- Deter potential evildoers by increasing the chance of their getting caught.
- Be more cautious and do more risk assessment work.

These recommendations more closely track the approaches and instruments that were developed to deal with high-risk uses of chemical and nuclear weapons. In fact, since 2012, there has been a major effort underway called the Campaign to Stop Killer Robots, which seeks a multinational treaty to stop the most nefarious robotic applications.[78] At the time of this writing, almost 30 countries, 86 nongovernmental organizations, and more than 25,000 AI experts had pledged support of this effort "to ban fully autonomous weapons and thereby retain meaningful human control over the use of force." Meanwhile, almost 250 organizations and more than 3,000 individual experts have signed the Future of Life Institute's "Lethal Autonomous Weapons Pledge," which "call[s] upon governments and government leaders to create a future with strong international norms, regulations and laws against lethal autonomous weapons."[79] Signatories vow that they "will neither participate in nor support the development, manufacture, trade, or use of lethal autonomous weapons."[80]

It remains unclear how enforcement will work, but one could imagine that "killer robot" applications might be limited through international accords and actions, perhaps using the Geneva Protocol for chemical weapons or the NPT and the IAEA for nuclear proliferation as models. A similar framework might be considered for 3D-printed weapons or even certain synthetic biology or genetic engineering applications that involve extreme forms of human modification. The UN's Biological Weapons Convention framework might provide a model in these cases. The International Criminal Court, whose mission is to "to hold those responsible accountable for their crimes and to help prevent these crimes from happening again," could also play a role in addressing lethal uses of emerging technologies.[81] Over time, the body of laws, accords, and general principles that make up the law of armed conflicts will evolve to accommodate these new technological capabilities.

To be sure, such approaches are not foolproof. How should we deal with rogue states or other holdouts who refuse to play a constructive role in such agreements and treaties? We already face that problem with nuclear nonproliferation efforts and states like North Korea. More problematic is the question we have already alluded to regarding the regulation of dual-use technologies. Namely, how can we address the harmful applications of various general-purpose technologies (computing, robotics, 3D printers, genetic editing,[82] etc.) without undermining the many beneficial and life-enriching applications of those same technologies?[83]

In this regard, global regulation of genetic editing will soon become an important test case. In March 2019, several of the world's leading genetic scientists came together and called a worldwide five-year moratorium on DNA editing for purposes of producing genetically modified children.[84] The scientists asked governments to "publicly declare that they will not permit any clinical use of human germline editing for an initial period [of five years]."[85] Interestingly, many other top geneticists refused to sign the call for a ban, even though it was just a request for a voluntary moratorium, not a formal treaty. Those not signing the call for a moratorium cited a variety of factors, including the fact that it seemed to be too late for such a ban to be meaningful, with the proverbial genie already well out of the bottle.[86] The scientists agreed that there were serious ethical issues surrounding genetically edited children, but consensus proved elusive about the regulatory specifics. There were unanswered questions about who would enforce the moratorium and how they would do so, especially against rogue actors operating in states that will not honor such a ban.

In a world where innovation arbitrage is only getting easier, dual-use technologies will be harder to control because they and

their creators will, as Richard Posner has noted, "simply gravitate to another country."[87] That observation is particularly true today because *physicality* matters less than it did in the past. In a world of ubiquitous and near-instantaneous information flows, how can we really control the ultimate threat: the spread of knowledge about dangerous ideas and applications? But that still leaves open the wisdom and practicality of regulating dual-use technologies more generally. If not properly targeted and limited in nature and scope, overzealous bans on broad classes of technologies could undermine scientific discovery and the many accompanying life-enriching and life-saving benefits specific technologies could bring about.

Risk Prioritization Is Essential

To reiterate, these questions are extraordinarily challenging, with no easy answers. Even though I have repeatedly stressed the benefits of allowing most innovation to develop relatively unencumbered, there will always need to be some limits on those technologies that have the potential to bring about more serious risks to humanity. Precautionary restraints are most justifiable when the alleged harms are highly probable, tangible, immediate, irreversible, catastrophic, or directly threatening to life and limb in some fashion.[88] The argument for permissionless innovation as a *general* default should not be viewed as a demand for unfettered freedom to innovate in *every* instance. That was obviously the case for nuclear and chemical weapons, and it is why Bostrom and others are correct to raise questions about future technological developments that could produce similar existential risks to civilization. We need some prior restraints on technological innovation in such instances.

But, again, perspective is essential. The three important things to remember about technological risk are the following:

- Not all technological risks are equal.
- Almost all technological risks have corresponding *rewards* that must also be considered (or, stated differently, there can be no reward without risk taking).[89]
- Knowledge and resource constraints challenge our ability to predict the course of technological developments.

Because of all these factors, it is vital to weigh the full range of tradeoffs associated with any proposed solution(s) to alleged technological risks. The most important thing that policymakers can do in this regard is to get smarter about risk prioritization and to stop making risks seem greater than they are. Debates about technological risk are haunted by false equivalence in technological risk assessment. Chapter 1 offered several examples of technology critics resorting to false equivalence and threat inflation when discussing tech policy issues. Those views can lead to a paradox in that society might spend so much time and energy panicking over lesser risks that it fails to properly address the ones that are truly significant.[90] In other words, it is the proverbial boy who cried wolf.[91]

Consider, for example, a 2016 address by the then–UN secretary-general Ban Ki-Moon on the Non-Proliferation of Weapons of Mass Destruction (WMDs).[92] In his remarks, the secretary-general advocated a stepped-up disarmament agenda "to prevent the human, environmental and existential destruction these weapons can cause."[93] Ban rightly pressed the need to remain vigilant in addressing the horrors of chemical, biological, and nuclear attacks. He did not stop there, however. The secretary-general went on to discuss his concerns about "new global threats

emerging from the misuse of science and technology, and the power of globalization."[94] His speech included a diverse class of emerging technologies that are not usually mentioned in the same breath as those traditional WMDs, including information and communication technologies (ICTs), artificial intelligence, 3D printing, and synthetic biology. Ban said such technologies "will bring profound changes to our everyday lives and benefits to millions of people" but worried that "their potential for misuse could also bring destruction. The nexus between these emerging technologies and WMD needs close examination and action."[95]

There is nothing wrong with Ban raising concerns about many of these emerging technologies.[96] Yet by so casually moving from a heated discussion of traditional WMDs into a brief discussion about the potential risks associated with ICTs, AI, 3D printing, and synthetic biology, he implies that these technologies and their potential risks are roughly equivalent. But it is simply not the case that all these risks are equal. The secretary-general is using what rhetoricians refer to as an appeal to fear. Douglas Walton, author of *Fundamentals of Critical Argumentation*, outlines the argumentation scheme for fear-appeal arguments as follows:[97]

- *Fearful Situational Premise*: Here is a situation that is fearful to you.
- *Conditional Premise*: If you carry out A, then the negative consequences portrayed in the fearful situation will happen to you.
- *Conclusion*: You should not carry out A.

This logic pattern is known as *argumentum in terrorem* or *argumentum ad metum*.[98] Tech critics and other concerned parties sometimes use fear appeals in an attempt to shake the public or policymakers out of a perceived slumber and get them to pay

more attention to new technological risks. The problem with fear appeals, however, is that they are often logical fallacies built on poor risk analysis or even outright myths.[99] Yet if such appeals are successful, they can lead to unnecessary anticipatory regulation of emerging technologies.

Ban's speech presents that problem. When important international officials like Ban group all these technologies together in a speech about weapons of mass destruction and sandwich them between impassioned opening and closing statements about the need "to take action" because "the stakes are simply too high to ignore," we are witnessing a fear appeal in action. The conclusion that follows from such appeals is obvious: global controls of some sort are needed. Again, this is a false equivalence. Ban's mistake is to equate all these technologies and risks and then suggest that sweeping action is needed for all of them when, in reality, such actions are probably only appropriate for a smaller class of technologies that legitimately pose an existential risk to humanity.

Policymakers and international figures of importance should be extremely cautious about the language they use to describe new classes of technologies, lest they cast too wide a net. Suggestions that every new technology poses a catastrophic or existential risk will desensitize people to actual risks that may be associated with a narrower class of innovations. That does not mean we should ignore risks associated with other technologies or technological capabilities. Instead of using fear appeals and advocating extreme (and likely unworkable) global regulatory schemes, however, it will often be wiser to build on existing laws, norms, and alternative governance frameworks. It is important to be practical. It most contexts, it remains highly unlikely that a global governance solution will work. We are not likely to witness the development of strict global laws and regulatory bodies, at least not any with serious teeth. The better role

that international bodies and actors can play is as coordinators of national policies and conveners of ongoing deliberation about multinational concerns.[100]

For example, a variety of transparency laws and other efforts already exist in many national and global governance regimes. These include know-your-customer guidelines and whistleblower processes that aim to identify problematic actors in various contexts. More resources could be plowed into such efforts. Again, education and awareness-building efforts can also be tapped in many cases. Soft law still has a role to play in this regard, too. Even if we cannot achieve global consensus on the potential harms associated with particular technologies or figure out how to successfully craft a formal global regulatory regime to address those concerns, less formal governance efforts can still help create important ethical norms. Best practices or codes of conduct for researchers and developers can also go a long way toward fostering a culture of responsibility and a greater commitment to safety, as even Bostrom has acknowledged.[101] These options should at least be given a greater chance to help start a conversation about wise technological development and responsible innovation.[102]

Finally, Marchant and Wallach's GCCs idea, discussed earlier, might have some merit in this regard—assuming we can figure out how to create them and make them work in various contexts. In the field of digital communications coordination and internet domain name management, the Internet Society (founded in 1992) and the Internet Corporation of Assigned Names and Numbers (ICANN) are examples of governance coordinating committees of sorts.[103] But ICANN deals with more technical matters that do not involve existential risks. Consensus and coordination will likely prove more challenging in the same areas where it is potentially most needed. If, however, policymakers

can get risk priorities right and zero in on the most serious harms, it at least gives society a chance to better address those issues in a rational fashion while allowing other important innovations to develop freely.