

# Chapter 7

## Anticipatory Governance of Emerging Technologies

Daniel Sarewitz

In the last two centuries continual technological innovation has catalyzed wave after wave of social transformation, and is implicated in a few cataclysms as well. New waves now seem, perhaps, to be rapidly approaching, associated with accumulating and converging scientific and technological advances in such areas as nanotechnology, information technology, neuroscience, and human biotechnology. The power of these emerging technologies to remake society is thought by many to be on a scale comparable to the rise of steam power in the first industrial revolution, the emergence and convergence of electric power and the internal combustion engine in the late nineteenth century, and the proliferation of information technologies in the latter part of the twentieth century. Energy production systems, manufacturing systems, military weaponry, even the performance standards of the human brain and body are seen by some as subject to radical transformation in the coming decades. The accuracy of particular technological predictions is not really important. What is undeniable is that the scale and pace of the global research and innovation effort continues to grow, and that the consequences of this effort continue to permeate and transform society at every level.

What are the prospects for governing the societal implications and consequences of these emerging waves of technology? Current approaches are almost entirely reactive, ponderous, and bureaucratic, and are increasingly overmatched by the scale and pace of technological innovation and change. Standard regulatory regimes for dealing with chemicals in the environment, for example, have devolved into a miasma of litigation, politics, and scientific uncertainty that benefits neither the environment nor the economy. Governance of pharmaceutical products is perhaps justly criticized from all sides – useful drugs are not approved fast enough, harmful ones are not caught soon enough, useless ones seem to proliferate. Innovation is simply too fast, too pervasive, too decentralized to yield to approaches that demand comprehensive knowledge as a basis for taking action.

---

D. Sarewitz (✉)  
Arizona State University, Tempe, AZ, USA  
e-mail: daniel.sarewitz@asu.edu

Is there a way forward? I start with some general thoughts about what it actually means to talk about governing technological change, before moving toward some concrete examples of work now being done to develop theories, models, and tools that can improve the social capacity for guiding future technologies toward desired societal outcomes and away from undesired ones, a process termed anticipatory governance (Barben et al. 2008).

The dilemma for democratic societies created by our commitment to continual technological advance is obvious. If on the one hand we are committed to notions of pluralism, participation, and openness in charting the course of society, how on the other can we come to grips with the enormous transformational power of technology and technological systems, a power that often seems at once inscrutable, unconscious, overwhelming, and autonomous? Thirty years ago Langdon Winner (1978) developed the notion of “reverse adaptation” to describe the “adjustment of human ends to match the character of the available means” created by technological systems. At around the same time, David Collingridge (1980) articulated the fundamental dilemma of technological governance: in the early phases of technological evolution, many avenues of advance are available and promising, but too little is known about potential impacts to choose the best paths. Later on, when more is known about impacts, options are greatly restricted due to technological lock-in and concentration of power among vested interests. In light of such observations, and given that technology is among the most powerful forces for social transformation operating in the world today, it’s not unreasonable to wonder about the extent to which our commitment to democracy is an illusion or an opiate.

Of course one could say the same thing about, say, earthquakes, weather, or the motion of the solar system, that they make a mockery of democratic aspirations since they mediate our actions without our consent. But no one complains that the laws of gravity, or the motion of tectonic plates, are unfair and need to be governed more wisely. So we similarly could – and often do – place technological change outside of ourselves by conceiving it as an external phenomenon. This solves the democracy problem, because it allows us to treat technological change as a force to which we can only react. And indeed, for the most part, our approach is to pour tremendous resources into the creation of technological advance and then regulate and react to the outcomes as necessary to make them tolerable, just as we react to and accommodate weather or the motions of the Earth’s crust.

But this remains unsatisfying because technological innovation is, after all, a human endeavor, one that arises from human choices made in human cultures and institutions, and one whose importance for society depends on the continual willingness of humans to avidly make use of technological products and processes. Technology is our unruly child and we cannot evade some sense of accountability for its behavior.

Starting in the late 1960s, in the shadow of the Cold War and the emergence of the environmental movement, aspirations for the control of technological change began to grow. In particular, the technology assessment movement was rooted in the notion that future trajectories of technological evolution could be predicted and governed. As explained in 1976 by Harvey Brooks (p. 20), one of the founders of the study

of post-War science policy: “Ideally the concept of TA [Technology Assessment] is that it should forecast, at least on a probabilistic basis, the full spectrum of possible consequences of technological advance, leaving to the political process the actual choice among the alternative policies in the light of the best available knowledge of their likely consequences.”

But a more pessimistic vein of analysis, represented by people like Winner, Collingridge, and, before them, Lewis Mumford and Jacques Ellul, viewed such hyper-rational ambitions as implausible, due to the pervasive embeddedness of technological innovation in human institutions and political arrangements. Winner (1978) talked about “technics out-of-control” and “technological somnambulance” to convey a sense of powerlessness and resignation in the face of continual technological transformation of society.

What lies between an implausible commitment to control and a fatalistic embracing of passivity? Certainly the expectation that democratic societies (or any other societies, for that matter) can dictate technological futures is neither coherent nor desirable. We know that efforts to control most forms of social activity turn out to create more problems than they prevent. And we know that most efforts to predict technological pathways as an input into decision-making have been failures, and often absurd failures at that. But neither is there a need to abandon all hope. Another way to look at the problem is to start with the recognition that, like procreation, technological innovation is an innately human activity, and as such it acts as a mirror on, and amplifier of, the ambiguities and contradictions of the human condition. If this recognition tempers our expectations, then the alternative to control is not abdication, but reflection – on what we are actually doing – and governance – based on our reflections, and carried out in the context of our democratic aspirations. In most other domains of important human action and choice, the role of democratic decision making is not to exercise control but to reveal and adjudicate value disputes that underlie choices. Yet it is precisely in this domain that governance of technological change, for the most part, has gotten a free pass.

Why should this be? The key reality is that the products of science and technology do not appear magically; rather, they emerge from choices made by people working in institutions designed by people. In the United States after World War II, a series of strategic decisions were made about which areas of science should be advanced, and those decisions led, over a period of several decades, to revolutions in such areas as computer science, solid-state physics, materials science, molecular biology, genomics, and electrical engineering, and to linked technological revolutions in weapons, communication, information, transportation, and bio-technologies. These developments were not designed in advance and implemented in an ordered or predictably way. But neither did they happen accidentally, serendipitously, randomly, surprisingly. It was all a product of decisions made in government, in industry, in universities, by people with a strong, if evolving, sense of what they were trying to accomplish over the long term. The process was powerfully driven by the role of the U.S. Department of Defense as both leading investor in, and principal consumer of, advanced technology (Alic John 2007). Yet the approach was dominated not by top-down planning, but by catalyzing close

relationships among a relatively small number of leading universities, corporations, and government agencies. These linkages led to tightly coupled networks and feedbacks across a growing innovation enterprise that was at once institutionally highly complex, yet highly focused, as a matter of mission, on rapid technological advance. In other words, the explosive growth of the U.S. R&D enterprise in the Cold War era was an exercise in the governance of science and technology within the broader context of a complex, adaptive innovation system. Similarly, any effort to govern the societal implications of rapidly emerging technologies must contend with, and indeed exploit, the decentralized, networked essence of the innovation process.

The systems view renders standard cause-effect thinking irrelevant. For example, the question of whether the long-term results of some specific discovery or line of research actually were predictable was quite besides the point. Decisions were being made with a view toward future outcomes, not by tossing dice, and such decisions strongly determined what types of knowledge and innovation were created, and who was likely to benefit from that knowledge and innovation. Decision makers were acting in response to values, interests, aspiration, power, etc., just as decision makers always do. *The key questions, then, are these: who is making the decisions? And how do these decisions emerge from and interact with the complex socio-technical context within which they are being made?*

Why has technological change, unlike other areas of human activity, largely been exempted from the rigors of democratic debate? Certainly part of the reason, as I've suggested, is the sense that technological change is simply too complicated, too unpredictable, and too inevitable, to yield to collective engagement with its meanings in democratic forums. Another reason is that technology is closely aligned with science, and so with the powerful cultural belief that the process works best when it is left alone. A related supporting belief is that benefits are inherent in science and technology, whereas the problems they create are the fault of society, or politics. Perhaps most importantly, however, is the alignment of technological innovation with the ideologies of the marketplace, which tell us that the appropriate measures of technological value are monetary, and the appropriate mode of intervention is hands-off.

Whatever hypothesis one prefers, the overriding fact is that, in contrast to almost every other important area of human endeavor, the pursuit of technological transformation is largely exempted from formal democratic processes of eliciting value preferences and adjudicating value disputes about desired future states, even though technological innovation strongly expresses those very things.

This exemption perhaps explains why Technology Assessment began as a technocratic exercise, with rational analysis formally separated from political decision processes. TA was something added on to the innovation process, done in different places, like the defunct Congressional Office of Technology Assessment. TA also bought into the notion of science and technology as essentially autonomous enterprises that could be governed by introducing new technical information into political discourse as a basis for regulation. Thus, TA harbored the expectation that decision makers would potentially be willing to make controversial decisions on the basis of

highly contestable, non-verifiable probabilistic statements about the future of a technology. It was destined to disappoint. As Harvey Brooks (1976, p. 21) wrote: “The record on the implementation of TA has not been particularly happy. The outcome, whether negative or positive, tends to be more determined by political momentum and bureaucratic balance of power than by a rational process.”

Surprise. But this then led to the wrong conclusion: that, since technological assessment could not be based on technocratic predictions, it could not be done at all. This wrong conclusion was inherited by the next generation of technology governance through the Ethical, Legal, and Societal Implications (ELSI) program of the Human Genome Project. In the early 1990s, ELSI was grafted onto the Genome Project to support research by social scientists and humanists on some of the complex dilemmas raised by the coming proliferation of genomic information (Cook-Deegan 1996). ELSI was about understanding emerging social dilemmas, but it included no mechanisms for feeding back into decision making about science, or feeding forward into decisions about genome politics. It codified the separation of the science from the study of the social outcomes of science, and marks the end of the first era of Technology Assessment.

In contrast, over the past several decades, growing insight into the dynamics of innovation systems has stimulated new approaches to technological governance aimed at resolving the Collingridge dilemma. These new approaches are rooted in the idea that, by making the human choice contexts implicated in innovation processes visible and open to multiple perspectives, conscious governance can emerge at earlier stages, when more options are available and when uncertainty about future impacts is higher. This work was pioneered by Arie Rip and colleagues in the Netherlands, who termed it “constructive technology assessment” (Schot and Rip 1997) and has more recently gained beachheads in Britain, for example with work done at Lancaster University and the think-tank Demos on “upstream engagement” (Wilsdon and Willis 2004), and in the U.S., for example with work I’ve been involved with at Arizona State University (ASU), which we term “real-time technology assessment” (Guston and Sarewitz 2002), and which I will describe in more detail below. The goal of these efforts, most broadly, is to inject pluralistic reflection into the innovation process as a means of improving the public value of new technologies. The overall goal is to create a capacity for anticipatory governance, by building reflexivity into institutions where key technoscientific choices are being made.

If we understand that we are all participants in a great experiment in social transformation being carried out without our consent or even our understanding, the self-imposed limits of TA now become almost painfully obvious. If we understand technological transformation as emerging not from the autonomous, automatic advance of science and technology but from a complex set of decisions made within a variety of institution contexts, then a different way to think about and implement TA can emerge. This new approach to TA will reflect the following realities:

1. The pace and direction of advancing knowledge and applications is determined by human choice.

2. The specific directions that technoscience is steered, and the pace of its advance, reflect who is making the decisions – their interests, values, motives, perspectives.
3. The decisions that are made are determined within a complex social setting that encompasses a range of socioeconomic, cultural, and political components.
4. This complex social setting interacts with the results of technoscientific advance to yield social outcomes. The setting, the science, and the outcomes mutually evolve over time.

These realities raise the following questions:

1. What is the range of choices available to people making decisions about science?
2. What are the interests, values, motives, and perspectives of people making decisions about science?
3. How do these interests, values, motives, and perspectives relate to the complex social setting within which decisions are made?
4. How do the results of scientific advance interact with socioeconomic, cultural, and political factors to yield social outcomes?

These questions can be researched and understood to various extents and would constitute *both the intellectual and the operational agenda* for the new approach to technological governance – an anticipatory approach, not in the futile sense of first predict, then take action, but in the sense of building institutional capacities to reflect on contexts and choices. The goal is to build a capacity for *reflexiveness* – social learning that expands the realm of conscious and available choice – into science and technology institutions and decision processes themselves. The process of understanding the dynamics of decision making about science and technology simultaneously provides knowledge and insight that can improve decision making processes and enable the participation of a broader and more diverse community of decision makers. Decision making is improved because previously implicit decisions become explicit, because expanded choices relevant to the decisions become apparent, and because greater diversity of actors relevant to the decisions can recognize themselves as potential stakeholders, thus creating the potential for improved deliberation.

While this capacity for reflexivity and anticipatory governance can and should be enhanced at many points in the innovation system, work that I've been involved in at Arizona State University focuses on the very upstream end of an emerging class of technology – nanotechnology – in the laboratory setting itself. The Center for Nanotechnology in Society at ASU (CNS; [cns.asu.edu](http://cns.asu.edu)), funded by the National Science Foundation, is in essence a test-bed for the idea that reflexivity can be built into the research process via a suite of social science methods termed by that constitute “real-time technology assessment,” or RTTA (Guston and Sarewitz 2002).

Above all else, RTTA, is about institutional innovation. It's about taking the closed environment of the research institution and opening it up so that the complex social dynamics of early-stage innovation processes become apparent to those

who are most centrally involved in those processes. RTTA includes four activities that, when taken together, create what are intended to be the necessary components of an inherently reflexive research process. The four components are:

- RTTA 1: Innovation system analysis. This activity builds knowledge of the technical landscape and the opportunities it is enabling.
- RTTA 2: Analysis of scientists' and the public's values and attitudes. This creates a generous awareness of the diverse values and aspirations of current and potential stakeholders who inhabit that landscape.
- RTTA 3: The creation of abundant opportunities for deliberation and stakeholder participation, informed and structured by what we learn in RTTA 1 and 2. This allows expansive exploration of alternative potential futures and landscapes.
- RTTA 4: Assessment of the social learning that actually occurs as a result of the prior three activities. This activity builds empirically grounded insight into how the system itself is evolving, and feeds that understanding back into the system.

RTTA seeks to make conscious and explicit the complex social, political, and economic setting within which nanotechnology research and innovation occurs, as it is occurring. RTTA does not try to predict the future of nanotechnology, but it does aim at stimulating discussions about what types of futures are possible, and what types are desirable. RTTA is certainly not in the business of telling researchers what to do, but it is in the business of allowing them to understand what they are doing in a manner that is much more contextually rich than in usual laboratory settings.

As I have emphasized, institutional innovation is at the heart of the effort. The goal is to move toward a research setting that is highly permeable to ideas and concerns that are normally excluded from lab settings. It took much of CNS's first 2 years simply to put the collaborative networks in place, build the necessary trust among partners, and begin to fully implement the wide array of opportunities for reflexive engagement, including collaborative teaching, joint research activities, and informal discussions between nanoscientists and social scientists; science cafes and other events in the community; scenario workshops and other future-visioning activities; and shared support of graduate students.

In March 2008, as the most important of CNS's participatory, RTTA 3 activities, we held the first National Citizen's Technology Forum (NCTF), bringing lay citizens together at six sites across the national to discuss, in highly mediated settings, the social implications of rapidly emerging and converging technologies. (For a full description, see Hamlett et al. 2008.) Some of the major issues of interest that emerged included:

- Need for effective regulation of emerging technologies;
- Demand for effective programs of public information;
- Concern about equitable and needs-based access to new technologies;
- Ambivalence about privacy, safety, and human performance enhancement.

Crucially, the deliberative process itself led to shifts in attitudes over the month-long course of the NCTF – that is to say, people actually did deliberate. While participants were almost uniformly optimistic about emerging technologies both before and after the NCTF, concerns about the downsides increased markedly. Doubts about the benefits of applications for human enhancement, about equitable access to technologies, about risks, and about economic implications all increased as a consequence of the extended deliberative process. That is to say, the deliberations enhanced the intellectual sophistication of the participants by allowing them to hold internally conflicting views of nanotechnology. This outcome is encouraging as well, but it also predicts that technical communities may seek to advance their own interests by opposing efforts to expand RTTA-like institutional innovations in the R&D system.

Within the laboratory setting itself, RTTA is intended to enhance awareness of the choices that researchers face as they pursue their experiments. Several types of questions seem to be surfacing at CNS, for example:

- Given several research project options, which one is likely to yield the most social benefit in the near term?
- Given several molecules that can serve a particular function, which one is the most environmentally benign?
- Should a neural enhancement device be implanted in the brain or be worn externally?
- Are the potential benefits of a human memory-enhancement implant obviously going to be greater than the potential downsides?

Yet such questions are perhaps overly concrete, because they might seem to suggest that the idea is to directly link individual choices upstream in the laboratory to complex downstream consequences in society. Innovation system complexity means that cause-effect chains will always be difficult to trace and that, except in exceptionally rare cases, the consequences of individual decisions will not be discernible in broad societal outcomes. Rather RTTA is a tool to build systemic capacity – the capacity to reflect on context and choice at a multitude of times and places in the innovation process.

At CNS, one early place where we expect to see evidence of this enhanced capacity is in the evolution of values and attitudes of scientists, engineers, and social science researchers to reflect greater awareness of the political, social, and economic contexts of innovation. We would also expect to see institutional values and norms evolve as well, for example in terms of expanded notions of scientific responsibility and productivity, and of what successful graduate education should look like. These are hypothesis that are still being tested as CNS moves into its fifth full year of activity.

The overarching hypothesis behind CNS and RTTA is that an emerging reflexive capacity will favor more socially beneficial choices – that is, choices that steer toward articulated public values – and CNS researchers are continually testing this hypothesis at the micro-level of partner research laboratories. Indeed, CNS is motivated by the belief that the very process of turning laboratories at a research



university from insular to openly reflexive is inherently beneficial because it creates openness, transparency, and broader capacity for engaged deliberation than existed previously. This benefit is in part procedural, in that open and aware deliberation is more democratically satisfactory than closed and clueless deliberation, or than a lack of conscious deliberation. But CNS also tests the idea that the benefit is instrumental: that reflexivity moves innovation toward more socially desirable outcomes, and away from undesirable ones, as diverse decision makers reflect more deeply on the context of their decisions. And of course this can happen either through a change in innovation paths, or through a change in the conceptions of desirability, or, more likely, through the interaction of both.

This conscious yet non-deterministic evolutionary process is at the heart of anticipatory governance. Anticipatory governance is an appropriate aspiration for democratic engagement with technological transformation, one that succumbs neither to the illusion of control, nor the resignation of technological somnambulance.

Anticipatory governance comprises three areas of simultaneous activity: engagement, foresight, and integration. Engagement encompasses the suite of activities that stimulates public deliberation; foresight describes the process of developing plausible and evolving scenarios of possible futures that can be the subject of the public deliberation; and integration brings the engagement and foresight activities into the domain of scientific practice to enhance reflexiveness (Barben et al. 2008).

RTTA represents one suite of methods aimed at advancing the goal of anticipatory governance, at one site, at one university. As I have mentioned, there are a few other similar exercises taking place, mostly in western Europe. If we are to escape from our self-imposed subjugation at the hands of the Collingridge dilemma, then the challenge is to move from local experiments and pilot projects to a scaled up, society-wide capacity to innovate reflexively, rather than unconsciously.

Part of the challenge is simply to make it safe to talk about innovation in terms of a range of public values and choices, rather than in the simple input–output, more-is-always-better mode. For example, it's not hard to think of some fairly simple questions that could *always* be discussed in public venues when decisions are being made about what R&D will be done. Instead of just asking: How much should we spend on this program or that? We can also ask:

- What are the values that motivate a particular investment in innovation?
- Who holds those values?
- Who is most likely to benefit from the translation of the research results into social outcomes? Who is unlikely to benefit?
- What alternative approaches are available for pursuing such goals?
- Who might be more likely to benefit from choosing alternative approaches?
- Who might be less likely to benefit?

The habit of asking these sorts of questions has not yet been formed. But habits do change. Important norms of scientific practice, for example, have evolved greatly in the past several decades. Issues of human subjects research, of the use and

treatment of animals in research, of environmentally safe practice, of the gender and ethnic diversity of the scientific community, have all become mainstream concerns of policy makers and researchers alike, whereas in the recent past, serious consideration of such issues was often labeled as “anti-scientific.”

Moreover, these changes in norms have come about along with changes in institutional structure. For example, concern about the ethical governance of human subjects research in the U.S. has led to nationwide institutional reform. Every publicly funded research project involving human subjects is monitored by an institutional review board (IRB) that must approve the research before it can be conducted, and ensure that ethical principles such as prior informed consent are enforced. There are thousands of such boards operating in the United States, thus demonstrating that comprehensive governance of innovation activities is a reasonable goal. While IRBs are far from perfect in protecting the rights of research subjects, and while they also impose a cost in terms of the efficiency of conducting research, they are nonetheless an accepted element of a scientific infrastructure that respects and protects human dignity.

The IRB experience demonstrates that comprehensiveness is possible when the stakes are high – and the stakes associated with emerging and converging technological revolutions are enormous and radical. Just as the IRB process is an accepted part of *all* human subjects research, institutionalizing anticipatory governance activities as part of the publicly funded science and technology enterprise could be done by requiring an RTTA-like component for *all* major public programs and projects related to transformational technoscience. This capacity-building could be funded by a small tithe, perhaps 2%, on research and innovation expenditures. And while such a scenario may seem, right now, to be ridiculously ambitious, one could easily imagine a time, perhaps several decades in the future, when every major research institution would be continuously engaged in the process of reflecting upon the values and choices that are implicated in its work. At such a time in the future, what will truly seem ridiculous is the fact that major research institutions in the first decades of the twenty-first century were committed to a rejection of the need for continually reflecting on the social meanings of the emerging technologies that they help to create.

## References

- Alic, John. 2007. *Trillions for military technology: How the pentagon innovates and why it costs so much*. New York: Palgrave Macmillan.
- Barben, D., E. Fisher, C. Selin, and D. Guston. 2008. Anticipatory governance of nanotechnology: Foresight, engagement, and integration. In *The handbook of science and technology studies*, eds. E. Hackett, O. Amsterdamska, M. Lynch, and J. Wajcman, ch. 38, 979–2000. Cambridge, MA: MIT Press.
- Brooks, Harvey. 1976. Technology assessment in retrospect. *Newsletter on Science, Technology, and Human Values*, no. 17 (Oct).
- Collingridge, D. 1980. *The social control of technology*. London: Frances Pinter.
- Cook-Deegan, R. 1996. *The gene wars: Science, politics, and the human genome project*. New York: W.W. Norton.

- Guston, D., and D. Sarewitz. 2002. Real-time technology assessment. *Technology in Society* 24: 93–109.
- Hamlett, P., M. Cobb, and D. Guston. 2008. *National citizens technology forum: Nanotechnologies and human enhancement*. CNS-ASU Report #R08-0003. Tempe, AZ: Center for Nanotechnology in Society, Arizona State University.
- Schot, J., and A. Rip. 1997. The past and future of constructive technology assessment. *Technological Forecasting and Social Change* 54 (2–3): 251–268, Feb Mar 1997.
- Wilsdon, J., and R. Willis. 2004. *See-through science: Why public engagement needs to move upstream*. London: Demos.
- Winner, L. 1978. *Autonomous technology: Technics-out-of-control as a theme in political thought*. Cambridge, MA: MIT Press.