

# TECHNOLOGICAL REVOLUTIONS: ETHICS AND POLICY IN THE DARK<sup>1</sup>

(2006) Nick Bostrom

[www.nickbostrom.com](http://www.nickbostrom.com)

[Published in *Nanoscale: Issues and Perspectives for the Nano Century*, eds. Nigel M. de S. Cameron and M. Ellen Mitchell (John Wiley, 2007): pp. 129-152.]

## Abstract

Technological revolutions are among the most important things that happen to humanity. Ethical assessment in the incipient stages of a potential technological revolution faces several difficulties, including the unpredictability of their long-term impacts, the problematic role of human agency in bringing them about, and the fact that technological revolutions rewrite not only the material conditions of our existence but also reshape culture and even – perhaps – human nature. This essay explores some of these difficulties and the challenges they pose for a rational assessment of the ethical and policy issues associated with anticipated technological revolutions.

## 1. Introduction

We might define a technological revolution as a dramatic change brought about relatively quickly by the introduction of some new technology. As this definition is rather vague, it may be useful to complement it with a few candidate paradigm cases.

Some eleven thousand years ago, in the neighborhood of Mesopotamia, some of our ancestors took up agriculture, beginning the end of the hunter-gatherer era. Improved food production led to population growth, causing average nutritional status and quality of life to decline below the hunter-gatherer level. Eventually, greater population densities led to vastly accelerated cultural and technological development. Standing armies became a possibility, allowing the ancient Sumerians to embark on unprecedented territorial expansion.

---

<sup>1</sup> I am grateful to Eric Drexler, Guy Kahane, Matthew Liao, and Rebecca Roache for helpful suggestions.

In 1448, Johan Gutenberg invented the movable type printing process in Europe, enabling copies of the Bible to be mass-produced. Gutenberg's invention became a major factor fueling the Renaissance, the Reformation, the scientific revolution, and helped give rise to mass literacy. A few hundred years later, *Mein Kampf* was mass-produced using an improved version of the same technology.

Brilliant theoretical work in atomic physics and quantum mechanics in the first three decades of the 20<sup>th</sup> century laid the foundation for the Manhattan project during World War II, which raced Hitler to the atomic bomb. Some believe that the subsequent buildup of enormous nuclear arsenals by both the U.S. and the USSR created a balance of terror that prevented a third world war from being fought with conventional weapons, thereby saving many tens of millions of lives. Others believe that it was only by luck that a nuclear Armageddon was avoided, which would have claimed the lives of many hundreds of millions, perhaps billions. These beliefs may both be true.

In 1957, Soviet scientists launched Sputnik 1. In the following year, the U.S. created the Defense Advanced Research Projects Agency to ensure that the U.S. would stay ahead of its rivals in military technology. DARPA began developing a communication system that could survive nuclear bombardment by the USSR. The result, ARPANET, later became the Internet, which made available the World Wide Web, email, and other services. The long-term consequences remain to be seen.

It would appear that technological revolutions are among the most consequential things that happen to humanity, perhaps exceeded in their impact only by more gradual, non-revolutionary technological developments. Technological change is in large part responsible for the evolution of such basic parameters of the human condition as the size of the world population, life expectancy, education levels, material standards of living, the nature of work, communication, health care, war, and the effects of human activities on the natural environment. Other aspects of society and our individual lives are also influenced by technology in many direct and indirect ways, including governance, entertainment, human relationships, and our views on morality, cosmology, and human nature. One does not have to embrace any strong form of technological determinism or be a historical materialist to acknowledge that technological capability – through its complex interactions with individuals, institutions, cultures, and the environment – is a key determinant of the ground rules within which the game of human civilization is played out at any given point in time.

In the course of a normal lifetime nowadays, we can all expect to be involved in one or more technological revolutions: if not as inventor, funder, investor, regulator, or opinion

leader, then at least as voting citizen, worker, and consumer. Given that technological revolutions have such profound consequences, one might think that they should be the focus of intense ethical deliberation and feature centrally in public policy analysis. If so much is at stake, it would seem to behoove us to dedicate a corresponding amount of effort to ensuring that we make the right decisions. How is humanity measuring up to this challenge?

## 2. ELSI research, and public concerns about science and technology

One can perceive a slow trend since World War II of intensifying endeavors to connect science and technology (hereafter, S&T) policy to a broader discussion about desired social outcomes. The rise of environmentalism in the 1960s fostered this trend, reflecting public demand that more S&T resources be devoted to the betterment of water and air quality. The congressional Office of Technology Assessment was created to improve understanding of the societal implications of technological choices. Disease lobbies have formed, seeking, among other things, increased funding for medical research into a variety of conditions. Concerns about global warming has pushed greatly increased resources into climate science and led to calls for more funding for research into alternative energy sources as well as more direct interventions to reduce greenhouse gas emissions.

Some 3% of the budget for the Human Genome Project was set aside for studying the ethical, legal, and societal issues (ELSI) connected to genetic information – not much in relative terms but still enough for Art Caplan to describe the move as the “full employment act for bioethics”. There is now a burgeoning of research in a number of technology-related fields of applied ethics such as computer ethics, neuroethics, and especially nanoethics. Research into the ethical, legal, and societal issues related to nanotechnology (NELSI) might over time outstrip that of the genetic ELSI program.

There is today also a large and diverse set of grassroots organizations, think tanks, and university centers that work on technology-related issues. In recent years, ethical issues related to human enhancement technology have particularly come to the fore. There is growing apprehension that anticipated technological developments – including nanotechnology but also artificial intelligence, neuro-technology, and information technology – are likely in the twenty-first century to have transforming impacts on human society, and perhaps on human nature itself. Some speak of an “NBIC”-convergence (referring to the integration of the neuro-bio-information- and cognitive sciences), and explicitly link this to the prospect of human enhancement.<sup>2</sup>

While the spectrum of opinion represented in these discussions is quite broad, there seem to be some points of consensus, at least within the Western “mainstream”:

---

<sup>2</sup> See e.g. (Roco and Bainbridge 2003; Bainbridge and Roco 2006).

- Technological development will have major impacts on human society.
- These developments will create both problems and opportunities.
- “Turning back” is neither feasible nor desirable.
- There is a need for careful public examination of both the upsides and downsides of new technologies, and for exploration of possible ways of limiting potential harms (including technological, regulatory, intergovernmental, educational, and community-based responses).

In addition to disagreements about the *content* of S&T policy, there are also disagreements about the *process* whereby such policy should be determined, with challenges being raised to the “official” model of the appropriate relationship between science and society, a model which harks back to the Enlightenment. According to the Enlightenment model, “the only scientific citizens are the scientists themselves. For science to engage in the production of properly scientific knowledge it must live in a ‘free state’ and in a domain apart from the rest of society. Historically, science’s grip on Truth is seen as having grown progressively stronger as society’s grip on science has grown progressively weaker and ever more closely circumscribed.”<sup>3</sup> The Enlightenment model pictures science as the goose that lays the golden egg, but only when it is protected from external interference.

This model has come under increased scrutiny since the 1960s. In Europe, broad efforts are underway to change the “social contract” between science and society in order to create a larger role for public participation and deliberation in setting the priorities and limitations of science and technology. The notion that science is unproblematically associated with progress is no longer widely accepted, a more critical approach having been stimulated by developments such as the nuclear arms race during the cold war, the Chernobyl disaster, and the increased salience of environmental concerns in later years. In Britain, the mismanagement of the BSE outbreak (“mad cow disease”) eroded public confidence in Government science policy. The experiences from the BSE crisis later helped foment public opposition to the introduction of genetically modified crops.

Initiatives to build more opportunities for the public to become engaged in science and technology issues can be seen as an effort to rebuild public confidence and to secure science’s “license to operate”. But beyond such public relations goals, there are also many who argue that the S&T enterprise needs much more guidance from society in order to ensure that scientific and technological research is really directed to achieve socially beneficial outcomes. The aim is not necessarily to restrict research, or to contest any particular scientific theory, but to yoke the science and technology behemoth to ends chosen by the people after due deliberation and debate. If S&T is such an important shaper of the modern world, it should be brought under democratic control, the thinking goes, and its

---

<sup>3</sup> (Elam and Bertilson 2002), p. 133.

workings should become more transparent to the people who have to live with the consequences.<sup>4</sup>

This view is reflected in a recent paper by Michael Crow and Daniel Sarewitz:

When resources are allocated for R&D [research and development] programs, the implications for complex societal transformation are not considered. The fundamental assumption underlying the allocation process is that all societal outcomes will be positive, and that technological cause will lead directly to a desired societal effect. The literature promoting the National Nanotechnology Initiative expresses this view.<sup>5</sup>

They continue:

The fact that societal outcomes are not a serious part of the framework seems to derive from two beliefs: (1) that the science and technology enterprise has to be granted autonomy to choose its own direction of advance and innovation; and (2) that because we cannot predict the future of science or technological innovation, we cannot prepare for it in advance. These are oft-articulated arguments, not straw men. Yet the first is contradicted by reality, and the second is irrelevant. The direction of science and technology is in fact dictated by an enormous number of constraints (only one of which is the nature of nature itself). And preparation for the future obviously does not require accurate prediction; rather, it requires a foundation of knowledge upon which to base action, a capacity to learn from experience, close attention to what is going on in the present, and healthy and resilient institutions that can effectively respond or adapt to change in a timely manner.<sup>6</sup>

Let us look in a little more detail at these two issues, the autonomy of the S&T enterprise and unpredictability, starting with the latter.

---

<sup>4</sup> Part of this intellectual trend is the conglomeration of science and technology into “technoscience”, the idea being (roughly) that science and technology are inextricably linked and that both are socially coded, historically situated, and sustained by actor networks consisting of both human and artifacts; see e.g. (Latour 1987). In this essay I for the most part do not sharply separate science and technology, but it seems to me that a more nuanced treatment would have to distinguish different components of the “science and technology enterprise” (to use the term favored by Crow and Sarewitz).

<sup>5</sup> (Crow and Sarewitz 2001), p. 97.

<sup>6</sup> (Crow and Sarewitz 2001), p. 98.

### 3. Unpredictability

Technological revolutions have far-reaching consequences that are difficult to predict. This poses a challenge for technology policy. The challenge, however, is not unique to technology policy. All major policy changes have far-reaching consequences that are difficult to predict. There does not exist an exact science that can tell us precisely what will happen in the long run when a government decides to abolish slavery, go to war, or give women the right to vote.

For more modest policy changes, such as a reduction of a sales tax or the introduction of stricter regulation on lead paint, expectations of the near-term consequences are more tightly constrained by economic and scientific models and by parallel experience in other countries. But social systems are complex, and even small interventions can have large unanticipated long-term consequences. Perhaps reduced lead levels will lead to increased intelligence in some children, and some of these might then grow up to become more successful scientists than they would otherwise have been. Some of these scientists might invent the future equivalent of the atomic bomb or antibiotics. Perhaps a reduced sales tax will increase profits in one sector of the economy, some of which might be used as campaign contributions that get a politician elected who will pass legislation that may, in turn, have wide-ranging and unpredictable ramifications.

Even the most trivial personal decisions can have monumental consequences that shape the fate of nations. Maybe one afternoon a thousand years ago in some Swiss village, a young woman decided to go for a stroll to the lake. There she met a lad, and later they married and had children. Thus she became the great-grandmother of Adolf Hitler. If she had gone to the forest instead of the lake, the Holocaust and perhaps World War II would not have happened.<sup>7</sup>

On the other hand, the unpredictability of the future should not be exaggerated. Crow and Sarewitz appear to concede that “we cannot predict the future of science or technological innovation”, but in doing so they concede too much. As Eric Drexler notes:

The future of technology is in some ways easy to predict. Computers will become faster, materials will become stronger, and medicine will cure more diseases.

Nanotechnology, which works on the nanometer scale of molecules and atoms, will be a large part of this future, enabling great improvements in all these technologies.<sup>8</sup>

<sup>7</sup> This example is borrowed from James Lenman (Lenman 2000). See also (Bostrom 2006).

<sup>8</sup> (Drexler 2003). Some technology impacts are equally predictable, e.g. some new medicines will be used, will save lives, some of those people whose lives have been saved will draw state pensions, vote, etc.

Predictability is a matter of degree, and the degree varies radically depending on what precisely it is that we are trying to predict.

Even such a seemingly platitudinous claim as “physics is better at prediction than social science” is on closer inspection quite problematic. Physics can predict some things and not others. We can use physics to predict with considerable accuracy where the planet Jupiter will be in the year 2020. But physics does not enable us to predict which particular atom will be closest to the center of gravity in the solar system in 2020. Social science can also predict some things: the prediction that there will exist some inequalities in social status and income in the U.S. in 2020 seems about as reliable as the prediction of where the planet Jupiter will be at that time.

If we want to make sense of the claim that physics is better at predicting than social science is, we have to work harder to explicate what it might mean. One possible way of explicating the claim is that when one says that physics is better at predicting than social science one might mean that experts in physics have a greater advantage over non-experts in predicting interesting things in the domain of physics than experts in social science have over non-experts in predicting interesting things in the domain of social science. This is still very imprecise since it relies on an undefined concept of “interesting things”. Yet the explication does at least draw attention to one aspect of the idea of predictability that is relevant in the context of public policy, namely the extent to which research and expertise can improve our ability to predict. The usefulness of ELSI-funded activities might depend not on the absolute obtainable degree of predictability of technological innovation and social outcomes but on how much improvement in predictive ability these activities will produce.

Let us hence set aside the following unhelpful question:

*Is the future of science or technological innovation predictable?*

A better question would be,

*How predictable are various aspects of the future of science or technological innovation?*

But often, we will get more mileage out of asking,

*How much more predictable can (a certain aspect of) the future of science or technological innovations become if we devote a certain amount of resources to study it?*

Or better still:

*Which particular inquiries would do most to improve our ability to predict those aspects of the future of S&T that we most need to know about in advance?*

Pursuit of this question could lead us to explore many interesting avenues of research which might result in improved means of obtaining foresight about S&T developments and their policy consequences.<sup>9</sup>

Crow and Sarewitz, however, wishing to side-step the question about predictability, claim that it is “irrelevant”:

preparation for the future obviously does not require accurate prediction; rather, it requires a foundation of knowledge upon which to base action, a capacity to learn from experience, close attention to what is going on in the present, and healthy and resilient institutions that can effectively respond or adapt to change in a timely manner.

This answer is too quick. Each of the elements they mention as required for the preparation for the future relies in some way on accurate prediction. A capacity to learn from experience is not useful for preparing for the future unless we can correctly assume (predict) that the lessons we derive from the past will be applicable to future situations. Close attention to what is going on in the present is likewise futile unless we can assume that what is going on in the present will reveal stable trends or otherwise shed light on what is likely to happen next. It also requires prediction to figure out what kind of institutions will prove healthy, resilient, and effective in responding or adapting to future changes. Predicting the future quality and behavior of institutions that we create today is not an exact science.

It is possible, however, to reconstruct Crow and Sarewitz’s argument in a way that makes more sense. Effective preparation for the future does require accurate prediction of at least certain aspects of the future. But some aspects are harder to predict than others. If we despair of predicting the future in detail, we may sensibly resort to courses of action that will do reasonably well independently of the details of how things turn out. One such course of action is to build institutional capacities that are able to respond effectively to future needs as they arise. Determining which institutional capacities will prove effective in the future does require prediction, but this is often a more feasible prediction task than predicting the details of the situations that they will have to respond to. The more the future is veiled in ignorance, the more it makes sense to focus on building general-purpose capabilities.

Recast in this way, the argument is more defensible as far as it goes. But it does not go very far. Its limitations become clear when we consider it within the context of S&T policy.

---

<sup>9</sup> See for example (Tetlock 2005).



The tasks of S&T policy include setting priorities for the allocation of funding to research projects. It is hard to predict which lines of research will bear fruit and which will not. There are several possible ways of responding to this predicament. One is to concentrate funding on those research avenues which we can be fairly certain will bear at least some fruit. Another is to diversify the research portfolio and fund a little bit of everything. A third approach is to bet on a few research avenues that seem especially promising and accept the risk of total failure. Depending on the funders' attitude to risk, and other factors, some mixture of these approaches might be optimal.<sup>10</sup>

But the question of predictability does not go away. Of course, it is not possible to fund a little bit of (literally) *everything*, and spreading out funding as evenly as possible among all seekers seems unlikely to be the smartest way of going about things. So how much funding should be concentrated on a few promising fields, such as nanotechnology? How tightly focused should that funding be on particular approaches, methods, and research centers? There is no simple answer. The optimal strategy will depend on just how confident the funders are in their ability to pick winners, i.e. predict future advances.

The situation becomes even more complicated if we consider that some research projects might not simply fail to come to fruition but might bear poisoned fruit – produce results that we would be better off without. Certain possible weapons technologies fall into this category, but some critics of technoscience would argue that it includes a great deal else beside. Presumably, the majority's feeling that humanity ought to pursue scientific and technological research rests on the assumption that the value of the consequences of such advances is likely to be, on balance, positive. But if this assumption is true, and if it is also granted that *some* technological advances will prove detrimental, then again the question becomes whether we can be confident enough in our ability to predict in advance which particular trees will produce poisoned fruit in order to be justified in cutting them down now, or whether we should instead let them all grow, in the name of our epistemic modesty. Universal cultivation seems to require that there be just the right amount of predictability: enough so that we can expect that on balance the orchard will be beneficial to humanity and that our cultivation of it will in fact promote its growth, but not so much that we would be better off by chopping down selected trees because their growth may in the long run cause harm.

The complexity of our prediction problem increases even further when we consider that the payoff of an individual research project is not independent of what happens with other research projects. Different advances may work synergistically (as in the case of NBIC technologies), or one might preempt another and make it obsolete. When such dependencies exist, the development of an optimal research portfolio becomes more difficult. If

---

<sup>10</sup> The need to make these kinds of tradeoff, of course, is not confined to funding of the natural sciences and technology, but applies to funding of the social sciences and ELSI programs too.

predictability is low, we might decide to ignore such dependencies; if it is higher, on the other hand, we would be remiss not to take them into account in deciding our research priorities. The question of dependencies between potential future advances might also have to be reflected in what institutional structures we should create for the process of S&T agenda-setting and implementation; for example, whether to establish a separate committee for a particular subfield. Again, the question of the degree of predictability of various aspects of the future – far from being one that can be trivially answered or sidestepped – is in fact highly difficult and central to many S&T policy issues.

From this brief discussion of predictability we can already draw several conclusions. *First*, while some scientific advances and technological innovations are hard to predict with accuracy far in advance, the problem is not unique to the science and technology context. Big policy decisions, small policy decisions, and trivial personal decisions all have important consequences that cannot be predicted in detail. *Second*, there are many aspects of scientific and technological developments that *can* be predicted. *Third*, all meaningful preparations for the future rely, explicitly or implicitly, on prediction. *Fourth*, the issue of the *relative* predictability of different aspects of the future, and how much the predictability can be *improved* by various kinds of investment, is important in thinking about how R&D programs should be structured.

A further lesson is that improvements in our ability to predict various potential S&T advances and their consequences could make a very valuable contribution to our capacity to make wise S&T policy decisions.

#### 4. Strategic considerations in S&T policy

Let us now turn to consider the source of an additional level of complexity in S&T governance: strategy and politics. These impose constraints on what can be done, and therefore on what it would make sense to attempt to do. As Ralph Waldo Emerson once wrote:

Web to weave, and corn to grind;  
Things are in the saddle,  
And ride mankind.

In particular, one needs to question whether mankind is really riding science and technology, or whether it is the other way around.

One obvious sense in which mankind is not in the saddle is that the S&T policy decisions on this planet are not made by some one unified body of rational and beneficent representatives of humanity who are trying to get us to some particular destination. Instead, there are countless agents, pursuing different and often opposing objectives, influencing

various aspects of our species' S&T activities – national and regional governments, corporations, private philanthropic foundations, special interest lobbies, journal editors, research councils, media organizations, university presidents, prize committees, consumers, voters, scientists, public intellectuals, and so forth. More specifically, we know two things: (1) there is no unified decision-making entity that has the power to direct or halt all research worldwide in any area, and (2) that many of the decision-making entities that influence S&T policies at various locations are themselves subject to influence from a variety of agents with diverging goals and agendas. Both of these facts have profound consequences for our thinking about S&T policy.

The first fact, the absence of global control of the world's S&T, makes it difficult or impossible to stop research and innovation in a particular direction even if it would be a good thing to do so. For example, even if some detailed study or public consultation concluded that a nanotech revolution would be detrimental to humanity, there is no clear path to preventing such a revolution from happening anyway. The government of one country might rescind public funding for research in certain areas thought likely to enable advances in nanotechnology; but research would continue (albeit perhaps at a slower pace) using funding from private sources. The government might then ban all such research, whether publicly or privately funded, but other nations would almost certainly continue to push forward, and if the technology is feasible, it will eventually see the day anyway. Global bans on technological developments are very difficult to negotiate and even harder to police. The difficulties are amplified in cases where significant incentives exist for some groups for moving forward, where development can be conducted with modest resources, where concealment is possible, and where there is no salient demarcation between the hypothetically proscribed activity and legitimate research. Nanotechnology satisfies many of these conditions, so the prospect of global relinquishment appears to be close to nil, at least in lieu of dramatic advances in both surveillance technology and global governance.<sup>11</sup>

The infeasibility of halting certain kinds of research is a point often repeated in S&T policy discussions: "If our country does not go forward with this, someone else will and we will fall behind," or "If nanotechnology is outlawed, only outlaws will have nanotechnology."<sup>12</sup> The appeal to national competitiveness seems to be one of the rhetorically most effective arguments both for increased spending on research, and against regulation that would slow development.

---

<sup>11</sup> The latter proviso is not insignificant if we are thinking about longer timescales. One might also imagine that support for tough international action could increase dramatically following a big disaster such as an act of nuclear terrorism. And of course, another way in which nanotechnology research could come to a halt is as a result of a civilization-destroying global catastrophe. This proviso is also not insignificant; see (Bostrom 2002).

<sup>12</sup> E.g. (Vandermolen 2006).

It is worth comparing this argument from economic competitiveness with another appealing argument for more research funding: that research is a global public good and should be supported out of love of humankind. The two arguments stand in some tension. The global public goods argument suggests that it might be in a nation's self-interest to free-ride on other nations' S&T investment (particularly foundational research, the benefits of which are especially difficult for the producer to monopolize). If this is the case then "national competitiveness" might actually suffer from the diversion of resources away from other sectors of society to S&T research. Yet both arguments could be true. There might be high returns for a nation to its investments in R&D, and additional returns that cannot be captured domestically and instead become a positive externality benefiting other nations. In this case, both the love of humankind and the appeal to national advantage would work in tandem as reasons for increasing R&D investment.<sup>13</sup>

As both of these arguments illustrate, there are important consequences for S&T policy from the fact that S&T policy is not perfectly globally coordinated. But the complexity of the strategic situation increases vastly when we take into account that even within a particular country, S&T decisions are not made by a single unified perfectly rational and perfectly beneficent agency. Policy recommendations directed to an imaginary ideal global or national decision-maker may form useful focal points for interim discussion, but ultimately they need to be transformed into recommendations addressed to some identifiable real agent. At that stage, recommendations must take into account the limitations of that agent's powers, understanding, attention, and interests. This transformation of "what should be done" in an abstract sense into sensible recommendations to an agent that can actually do things is far from straightforward.

For example, one argument that has been given for moving forward with nanotechnology research as rapid as possibly is as follows:<sup>14</sup>

- (1) The risks of advanced nanotech are great.
- (2) Reducing these risks will require a period of serious preparation.

---

<sup>13</sup> Most studies of the economic returns to R&D have not focused on the international dimension. Domestically, it appears that the social returns of R&D – although they are difficult to measure – are very high and that optimal R&D investment substantially exceeds the actually level. See e.g. (Jones and Williams 1998; Salter and Martin 2001).

<sup>14</sup> See (Drexler 1992a), p. 242. Drexler (private communication) confirms that this reconstruction corresponds to the point he was making. Obviously, a number of implicit premises would have to be added if one wished to present the argument in the form of a deductively valid chain of reasoning. By "advanced nanotechnology" I here refer to a possible future form of radical nanotechnology, sometimes called molecular nanotechnology, or "machine-phase" nanotechnology; see also (Drexler 1992b).

- (3) Serious preparation will only begin once the prospect of advanced nanotech is taken seriously by broad sectors of society.
- (4) Broad sectors of society will only take the prospect of advanced nanotech seriously once there is a large research effort underway.
- (5) The earlier a serious research effort is initiated, the longer it will take to deliver advanced nanotech (because it starts from a lower level of pre-existing enabling technologies).
- (6) Therefore, the earlier a serious research effort is initiated, the longer the period during which serious preparation will take place, and the greater the reduction of the risks that will eventually have to be faced.
- (7) Therefore, a serious research effort should be initiated as soon as possible.

I present this argument not in order to evaluate it, but to illustrate the point about strategic complexity. What naively looks like a reason for going slowly or stopping (the risks of advanced nanotech being great) ends up, on this line of thinking, as a reason for the opposite conclusion. When one attempts to integrate such second-guessing of the responses of various actors in one's R&S policy recommendations, the result might differ radically from what one would get from a more simple-minded approach that ignores the strategic dimensions of the situation.

It is interesting to consider to whom it is that this kind of argument addresses itself. The "broad sectors of society", which will supposedly begin serious preparation only after a large research effort is already underway, are presumably not the intended recipients of the message. If they were capable of and willing to understand, agree with, and act on an argument like this, then they would not need to wait for a large research effort to get underway in order to take the need for preparation seriously. The argument appears to be esoteric. There are some people who are "in the know" about the prospects of advanced nanotechnology, and these people would have a reason (so the argument goes) to direct their efforts towards accelerating the implementation of a serious nanotech research effort even if they thought that the risks of advanced nanotech outweighed the benefits.

One way the cognoscenti could do this would be by publicizing another argument for the acceleration of nanotech research, such as the argument that if *we* don't move forward quickly, then *somebody else*, perhaps a hostile state, will get there first – and that would be the worst possible outcome of all. Note that even this second argument addresses itself not to everybody but to a select group – in this case our compatriots, or at least the citizens of "good" states. It would not be desirable that the citizens of "bad" states urge *their*

compatriots and government officials to launch a crash program for the development of advanced nanotech so that *they* get there first.<sup>15</sup>

There are actually people (although perhaps not many) who think at this level of sophistication and attempt to take strategic considerations such as those above into account in deciding what they ought to do. Some of these people are well-meaning and honest and would not consent to putting forward an argument or an opinion that they did not sincerely hold to be true. Esoteric arguments do not require deception, or even active concealment, because to a significant extent audiences self-select which arguments they hear and absorb. The “nano-cognoscenti” might be the only ones who are receptive to the argument about a large research program being necessary to get broad sectors of society to take the risks seriously and start preparing. The citizens of “good” nations might be more likely to follow their compatriots’ advice on the need to move forward with the research to avoid falling behind in a future arms race than the citizens of the “bad” nations who we fear might otherwise take the lead. (But how sure can we be that this is always the case?)

Predictability, or the lack thereof, again emerges as an important issue. Clearly, anticipating the responses of many different agents, how these responses will interact, and more generally how the ecology of ideas and opinions will be affected by the promulgation of one argument or another, is a daunting task – in many cases even more difficult than forecasting future developments in S&T.

## 5. Limiting the scope of our deliberations?

It is tempting to ignore all of these difficulties and focus on the simpler task of figuring out what we have most reason to do, subtle strategic considerations aside. At the individual level, a person might simply try to decide: is nanotech likely to do more good than harm? If yes, then be in favor of its development; if no, oppose it. Or alternatively: is there some path involving nanotech development combined with certain regulations and or public policies that would bring great benefits and only moderate risks? If yes, then promote that path.

We may note immediately that the choice here is not a dichotomous one, either to ignore all strategic considerations or else to take them all on board without limitation. Even in relatively uncomplicated deliberations, *some* strategic considerations might be admissible. For instance, we might take into account the fact that other nations will proceed with nanotech development even if our nation doesn’t, while setting aside all considerations having to do with the cognitive limitations of the hoi polloi or the way that special interests are likely to influence the implementation of any officially adopted policy directive.

---

<sup>15</sup> One may of course insist that the good states should develop only *defensive* nanotech capabilities. But offensive and defensive applications would require largely the same underlying technological advances.

At a collective level, too, we might decide to exclude certain kinds of reasons from various contexts of public deliberation. This is the idea, for example, behind John Rawls' concept of *public reason*.<sup>16</sup> In dealing with constitutional essentials and matters of basic justice, citizens abiding by the idea of public reason are, according to Rawls, entitled to justify the position they want adopted only on the basis of reasons that could reasonably be accepted by other citizens who do not necessarily share the same metaphysical, religious or cultural views.<sup>17</sup> This constraint has been criticized on grounds that it would require insincerity (not putting forward the real reasons for one's views) and lead to a shallowness in public discourse as participants are required to confine themselves to the lowest common denominator of shared background assumptions. Rawls believes that we should nevertheless abide by such a constraint

... given our duty of civility to other citizens. After all, they share with us the same sense of its imperfection, though on different grounds, as they hold different comprehensive doctrines and believe different grounds are left out of account. But it is the only way, and by accepting that politics in a democratic society can never be guided by what we see as the whole truth, that we can realize the ideal expressed by the principle of legitimacy: to live politically with others in the light of reasons all might reasonably be expected to endorse.<sup>18</sup>

As a much more mundane example of discourse restriction, consider the convention against the use of *ad hominem* arguments in science and many other arenas of disciplined discussion. The nominal justification for this rule is that the validity of a scientific claim is independent of the personal attributes of the person or the group who puts it forward. Construed as a narrow point about logic, this comment about *ad hominem* arguments is obviously correct. But it overlooks the epistemic significance of heuristics that rely on information about *how* something was said and *by whom* in order to evaluate the credibility of a statement. In reality, no scientist adopts or rejects scientific assertions solely on the basis of an independent examination of the primary evidence. Cumulative scientific progress is possible only because scientists take on trust statements made by other scientists – statements encountered in textbooks, journal articles, and informal conversations around the coffee machine. In deciding whether to trust such statements, an assessment has to be made of the reliability of the source. Clues about source reliability come in many forms – including

---

<sup>16</sup> For a review of philosophical views on the idea of publicity, including those of Rawls, see (Gosseries 2005).

<sup>17</sup> This is a simplification of Rawls' view, but the details are not essential for present purposes; see also (Gosseries 2005).

<sup>18</sup> (Rawls 1999), pp. 242-243.

information about factors such as funding sources, peer esteem, academic affiliation, career incentives, and personal attributes such as honesty, expertise, cognitive ability, and possible ideological biases. Taking that kind of information into account when evaluating the plausibility of a scientific hypothesis need involve no error of logic.

Why is it, then, that restrictions on the use of the *ad hominem* command such wide support? Why should arguments that highlight potentially relevant information be singled out for suspicion? I would suggest that this is because experience has demonstrated the potential for abuse. For reasons that may have to do with human psychology, discourses that tolerate the unrestricted use of *ad hominem* arguments manifest an enhanced tendency to degenerate into personal feuds in which the spirit of collaborative, reasoned inquiry is quickly extinguished. *Ad hominem* arguments bring out our inner Neanderthal.

The instrumental proscription of *ad hominem* and the more deeply normative Rawlsian idea of public reason both illustrate the general concept of constructing discourses partly by ruling out of court some types of consideration. There exist many such discourse boundary constraints, both in science and in other arenas, varying from context to context. It is perhaps not implausible to suppose that some strategic considerations, such as the ones introduced above, may also be appropriately excluded in some contexts of S&T policy deliberation.

How should we determine where the appropriate boundaries for a particular discourse should be placed? The complexities of this question protrude far beyond the scope of this essay, but it might be worth listing a few potentially relevant factors.<sup>19</sup> These might serve to hint at how which future investigations could explore this area further.

- (1) Self-deception and bias. We know that human cognition is susceptible to self-deception and biases of various kinds. Certain types of consideration might offer more foothold than others to irrationality and prejudice.
- (2) Manipulability. Similarly, in public deliberation, certain types of consideration might be easier to manipulate for partisan purposes, suggesting that they should be given less weight or barred altogether.
- (3) Unpredictability. As argued already, the complexity of a certain level of strategic consideration is such that we might quickly lose ourselves in a fog of unknowability if we venture there.
- (4) Direct vs. mediated consequences. This harks back both to unpredictability and foregoing discussion of strategic considerations. But one might also maintain, as a matter of basic ethical theory, that we are not responsible for (and should hence not take into account?) those consequences of our actions that are mediated by

---

<sup>19</sup> These would be complementary to the “micro-level” maxims identified by Paul Grice in his work on conversational implicatures (Grice 1975).



many external causal factors or other by people's choices, even if they should be partly predictable.

- (5) Division of labor. In a public deliberation, it is not always necessary that all parties attempt to incorporate and digest all types of consideration. A sensible majority view might emerge even if some or all parties are biased and blinkered in what considerations they take into account.
- (6) Accessibility to stakeholders. Some types of consideration might effectively exclude some stakeholders from participating meaningfully in a public deliberation. Does this sometimes make the appeal to such considerations inadmissible?
- (7) Moral side-constraints – honesty and truthfulness included, but perhaps there are also other moral constraints and desiderata that apply.
- (8) The potential for cumulative progress. It might be that discussion of certain types of consideration produce results that are easier to preserve and carry over to future dialogs and new situations, enabling cumulative progress.
- (9) The potential for connection to other domains. Work that clarifies some types of consideration might more readily be integrated with results from other disciplines or cognitive domains. This might be an argument for paying more attention to such considerations.

Recall the two beliefs or “oft-articulated arguments” to which Crow and Sarewitz attribute responsibility for the lack of consideration given to the implications for complex societal transformation in the allocation of resources to R&D programs. One was the argument about unpredictability. The other is the argument that the science and technology enterprise has to be granted autonomy to choose its own direction of advance and innovation. This latter argument is, in Crow and Sarewitz's view,

contradicted by reality... The direction of science and technology is in fact dictated by an enormous number of constraints (only one of which is the nature of nature itself).

It is unclear that how this is supposed to be an objection to the view that the S&T enterprise should be granted autonomy. Those who hold this view might well agree with Crow and Sarewitz that the spirit of free inquiry is currently fettered by an enormous number of constraints while at the same time insisting that it would be preferable if at least some of these shackles could be abolished, or at least relaxed, and that imposing additional

constraints would make things worse. One needs to distinguish the normative question about how things ought to be from the positive question of how things are.<sup>20</sup>

Democracies often find it advisable to insulate certain functions from direct public influence. The U.S. constitution, for example, deliberately creates a judiciary that is substantially insulated from direct political influence. Central banks in many countries are similarly autonomous. In all democracies, the day-to-day operation of most government departments is significantly shielded from the direct impact of public opinion. In the universities, senior research staff enjoy a great deal of intellectual freedom and frequently have the opportunity to pursue research projects of their own choosing, even though their salaries might come from the tax payer. In the private sector, corporations and individuals are free to pursue almost whatever direction of research they desire, provided they have the funding to do so. And independent intellectuals and writers are accountable to nobody but themselves.

What advocates of increased public involvement and direct democracy in S&T are arguing is not that matters of scientific controversy ought to be settled by popular referendum, or that the public be brought in to micro-manage the conduct of scientific research projects. Nor are they minded to restrict freedom of speech or any of the other intellectual freedoms currently enjoyed by individuals. What is being proposed is generally some more moderate position, for example that the setting of the overall parameters and priorities of publicly funded S&T research ought to be made more transparent and more directly subject to democratic input than is currently the case. Efforts might be advocated to improve public understanding of science and to create more opportunities for genuine, two-way dialogue between scientists and the public. Often, the focus is on the implementation side of the S&T enterprise, urging greater direct democratic control over which new technologies are permitted, under what forms of regulation, and with what ancillary policies to modulate their societal impacts. Also sometimes of concern are specific methods of scientific research, such as animal experimentation or the use of human embryos.

The question of governance of S&T issues can probably not be separated from questions of governance in general. My focus here, however, is not on the structures of governance – with *who* should make the decisions – but rather on the terms of the discourse: what kind of considerations should be taken into account, and in what ways.<sup>21</sup> Of course,

---

<sup>20</sup> A distinction that is often overlooked in discussions about the future. There is a lamentable tendency to let prediction and evaluation blend into a confused blob of wishful (or fearful) thinking.

<sup>21</sup> Another variable here would be what we may term *the format* of deliberation, e.g. whether it should occur behind closed doors or in the full glare of publicity, and what the tone or spirit of a deliberation should be, e.g. how much and what kind of emotion should be displayed. It is plausible that such format factors have a substantial effect on the nature of the deliberation. There is some direct empirical evidence for this. For example, one recent study of the deliberations of the Federal Reserve's Federal Open Market Committee found that transcript publication suppressed the expression of dissenting

who makes decisions does, in reality, influence what considerations get taken into account. But we can still distinguish the normative question of what kinds of reasons ought to be guiding the S&T policy.

We have considered some possible grounds for various limitations of what reasons should be considered in the determination of S&T policy. Let us end the section with the words of Marie Curie:

We must not forget that when radium was discovered no one knew that it would prove useful in hospitals. The work was one of pure science. And this is a proof that scientific work must not be considered from the point of view of the direct usefulness of it. It must be done for itself, for the beauty of science, and then there is always the chance that a scientific discovery may become like the radium a benefit for humanity.<sup>22</sup>

## 6. Expanding the scope of our deliberations?

The quote by Marie Curie expresses an extreme version of the view that we ought to narrow the considerations taken into account. According to Curie, scientific work must be done “for itself, for the beauty of science” and with no view to its direct usefulness.

One may contrast the innocence of Marie Curie’s words with a well-known remark made some two and a half decades later by another distinguished physicist, Robert Oppenheimer, who had spearheaded the development of the nuclear bomb:

In some sort of crude sense which no vulgarity, no humor, no overstatement can quite extinguish, the physicists have known sin; and this is a knowledge which they cannot lose.<sup>23</sup>

The explosion of the first nuclear weapon in the Trinity test, and the later use in Hiroshima and Nagasaki, are sometimes seen as emblems of the failure of scientists to concern themselves with the societal implications of their work. In fact, many of the scientists involved in the Manhattan project (and others, such as Linus Pauling, who declined to participate) were quite deeply concerned about societal implications. Among those agreeing to lend their skills to the project, a major motivation was the concern that Nazi Germany might otherwise get to the fission bomb first. This is an example of a strategic

---

opinions and stifled debate over short-term interest rates (Meade, Stasavage, and London School of Economics and Political Science. Centre for Economic Performance. 2004).

<sup>22</sup> Marie Curie (1867 - 1934), Lecture at Vassar College, May 14, 1921.

<sup>23</sup> Robert Oppenheimer (1904-1967), “Physics in the Contemporary World”, lecture at M.I.T. November 25, 1947.

consideration mentioned earlier, national competitiveness. On the Marie Curie version of the Enlightenment model, neither this consideration nor scruples over how the scientific findings might later be used should be taken into account, at least not by the people doing the scientific work.

There are several general counterarguments against the view that the scope of our deliberations should be in some ways restricted.

First, ignoring considerations that are evidentially relevant to potential outcomes that we care about means ignoring relevant information. Ignoring relevant information might not be rational, and might impede our effectiveness in achieving our goals. The more relevant the information, and the more important the goals to which it is relevant, the greater the cost of such intentional ignorance.

Second, confining deliberations within a set of fixed constraints can yield power to those who determine what these constraints should be. This power can be misused. As rhetoricians and sophisticated technocrats are well aware, the framing of an issue – which implicitly determines what kinds of consideration will be seen as being appropriate and having a bearing – often effectively determines the conclusion a deliberation will reach. In political discourse, framings are often fiercely contested. To accept a set of discourse constraints might mean buying into someone else's agenda which tilts the deliberation in favor of some predetermined position.

Third, and related to the first two arguments, one fairly likely effect of adopting a narrow framework for our deliberations of the S&T enterprise is to create a bias in favor of a certain kind of “conservatism” – conservatism not in the sense of political ideology, but in the sense of a presumption in favor of business-as-usual. Scope-restrictions risk ruling out radical critiques, ones that challenge the fundamental assumptions behind the common way of thinking and doing things. In the context of S&T, the effect of this would not be to perpetuate the status quo, because change in science and technology is brought forth ineluctably by the intellectual advances generated by the enterprise itself. Rather, the effect could be to diminish the possibility of a *deliberate change of course*. Without recourse to radical critique, the locomotive will roll along its track, and the track might turn left or right; we might even be able to flick a switch here and there to select which branch of a bifurcation we take; but we exclude from our mental space a host of discontinuous possibilities, such as getting off the train and continuing our journey via another mode of transport.<sup>24</sup>

Radical critiques might challenge the metaphysical underpinnings of our world view. They might challenge our basic values or moral norms. They might undermine our confidence in the entire S&T project. Alternatively, they might suggest that our attempts to humanize the S&T enterprise will have the opposite effect from the one intended. They might argue that one particular anticipated technological breakthrough will have

---

<sup>24</sup> In the context of fundamental science, such a course change could be a Kuhnian paradigm shift.

consequences overshadowing all the rest, and that by failing to act accordingly we are grossly misdirecting our attention and our resources. They might contend that increased public engagement and increased efforts to anticipate the societal implications will have obnoxious consequences. They might argue for relinquishment of broad areas of technological research, or alternatively that everything should be put into accelerating some applications. They might suggest new ways of funding basic research that would sidestep expert panels and bureaucratic procedures. They might identify completely new ways of evaluating and measuring progress. The possibilities are myriad and impossible to specify in advance.<sup>25</sup>

Many radical critiques are utterly wrong, and some of them would be extremely dangerous if they became popular. Nevertheless, if we look back historically, and observe how many widely-held conventional wisdoms of the past are revealed as blinkered or deeply flawed by our current lights, we must surely admit that – by induction – it is likely that many of our own central beliefs, too, are deeply flawed. We need to build into our processes of individual and collective deliberation some self-correcting mechanism that enables us to question and rectify even our most deep-seated assumptions.

One way in which the scope of our deliberations can be broadened is by taking into account the kind of meta-level reflections that I have tried to illustrate in this essay. More specifically, one could argue that more work should be done on the normative dimensions of the S&T enterprise.

One normative dimension is ethics, and this is to some extent already part of the official programs, for example as the “E” in ELSI and NELSI, and as applied ethics more generally. One might have occasional misgivings about the quality, depth, or impact of this research, but at least there is some recognition of the significance and relevance of the questions it is supposed to address.

Another normative dimension that has been given rather less attention in these programs is that of applied (normative) epistemology. This encompasses a number of important problems. One such problem is to develop better higher-order epistemic principles for the conduct of scientific research. As said by E.T. Jaynes:

It appears to us that actual scientific practice is guided by instincts that have not yet been fully recognized, much less analyzed and justified. We must take into account not only the logic of science, but also the sociology of science (perhaps also its soteriology). But this is so complicated that we are not even sure whether the

---

<sup>25</sup> Although not necessarily intended as “radical critiques”, for a few recent examples see e.g. (Joy 2000; Bostrom 2003b, 2003a, 2005; Hanson 2003).

extremely skeptical conservatism with which new ideas are invariably received, is in the long run a beneficial stabilizing influence, or a harmful obstacle to progress.<sup>26</sup>

Yet the epistemological problems go beyond the challenge of how to maximize scientific and technological advancement. As we have seen, they also are also central to our thinking about the ethical and policy issues prompted by the S&T enterprise. Applied epistemology also lies at the heart of the problem of how to evaluate radical critiques of this enterprise.

There are many interesting approaches to these matters in addition to philosophical reflection and theoretical analysis. Here is a sample:

- Study the heuristics and biases affecting human cognition, and figure our ways of applying the findings to improve our judgment.<sup>27</sup>
- Study the correlates of true opinion, among experts and the public, and use this information as clues to who is right in cases of disagreement and to how we might improve our own epistemic situation.<sup>28</sup>
- Information technologies. Develop our information infrastructure in ways that will facilitate the collection, integration, and evaluation of information.
- Cognitive enhancement. Improve individual reasoning ability (e.g. concentration, memory, and mental energy) by educational, pharmaceutical, and other means.
- Study how vested interests, the mass media, and other social realities shape and bias (or facilitate) the processes collective deliberation.
- Public deliberation. Develop procedures, formats, or rhetorical standards to improve the quality of public debate by confronting “plebiscitory” reason.<sup>29</sup>
- Subsidize and implement institutional innovations such as information markets, which have been shown to outperform expert panels in many prediction tasks.<sup>30</sup>

The unifying theme is to explore how we could make ourselves smarter and wiser, both individually and as an epistemic community. The research could be slanted towards applications in S&T assessment but it is likely to have important spill-over benefits in other areas. Such a program could be combined with more narrowly-focused efforts to gather and analyze information in areas of particular concern, such as nanotechnology.

Finally, I want to call attention to one more “normative dimension”, except that it is not really a dimension but rather the space spanned by all the other vectors. I am referring to

---

<sup>26</sup> (Jaynes and Bretthorst 2003), p. 525.

<sup>27</sup> See e.g. (Kahneman and Tversky 2000; Gilovich, Griffin, and Kahneman 2002). For an attempt to apply this kind of information to a technology-related issues, see (Bostrom and Ord 2006).

<sup>28</sup> See e.g. (Tetlock 2005).

<sup>29</sup> (Chambers 2004).

<sup>30</sup> (Hanson 1995; Wolfers and Zitzewitz 2004; Leigh and Wolfers 2006).

the challenge of integrating all crucial considerations into some coherent unity that will let us determine what we have most reason to do all things considered. This might have to accommodate predictions about technology, social impacts, strategic considerations, value judgments, ethical constraints, and assorted meta-level thoughts about how all these things should fit together.

Such synthetic work is not in fashion at the present time. It is discouraged by the disciplinary structure of academia and by prevailing norms of academic publishing, and it does not appear to be strongly nurtured by the short-term grants funding available for interdisciplinary projects either. As a consequence, synthetic work is undertaken mainly by time-pressured journalists, popular science authors, retired scientists or senior scholars who are no longer willing or able to do serious work, crackpots, and miscellaneous eccentrics. The quality of the ensuing contributions is often poor, and even when something worthwhile comes out there is no guarantee that it will be used and built upon by others.

Let me try to indicate slightly more clearly what kind of synthetic task I have in mind. I am not referring to the erection of grand philosophical systems resting on foundations of indubitable first principles; nor do I refer to the painting of “visions” or the realization of comprehensive ideological “outlooks”; nor still am I pushing for some new kind of systematization of all knowledge, a universal taxonomy, or the creation of an encompassing database, or a library of commissioned studies. Rather, what I have in mind is the task of attempting to think through some of the big challenges for humanity in a way that does not leave out any *crucial* consideration – by which I mean a consideration such that if it were taken into account it would overturn the conclusions we would otherwise reach about how we should direct our efforts. For example, some of the strategic considerations related to the nanotechnology initiative that I mentioned earlier might be “crucial” in that they might plausibly, if they are sound and once taken into account, rationally deliver a practical conclusion pointing in the opposite direction from the one we might otherwise believe we ought to strive towards. But not only strategic considerations, but also other empirical, epistemological, axiological, and methodological considerations could be crucial in this sense.

Implicitly, we are confronting a challenge of integrating all crucial considerations every time we are attempting to make a reasoned decision about some matter which we think it is important to “get right”. When our goal is very limited, we might at least sometimes succeed in meeting the challenge (albeit not usually by relying on reasoned deliberation alone). When the goal is more open-ended, such as if we are attempting to decide what we have most reason to do with our own life all things considered, or if we are seeking to form an opinion on a topic such as what public policy ought to be with regard to some anticipated technological revolution, then the complexity of the synthetic challenge grows enormously. It is not clear that we ever manage to meet it in any robust sense. Instead, what answer we end up espousing might depend mostly on contingent factors such as the

political inclinations of our parents, the idiosyncratic views of our thesis advisor, the current cultural climate in the place we happen to live, or the mere fact that we failed to think of some crucially relevant consideration that would have caused us to come to a very different conclusion.

What this seems to amount to is that we can have very little rational confidence that our efforts, insofar as they are aiming ultimately at important long-term goals for humanity, are not entirely wrongheaded. Our noblest and most carefully considered attempts to effect change in the world might well be pushing things further *away* from where they ought to be. Perhaps around the corner lurks some crucial consideration that we have ignored, such that if we thought of it and were able to accord it its due weight in our reasoning, it would convince us that our guiding beliefs and our struggles to date had been orthogonal or worse to the direction which would then come to appear to us as the right one.

Other than becoming generally more agnostic, what can we do about this? It is not clear how effectively the synthetic task could be broken into subtasks which can be delegated to a team of researchers and then stitched back together. But it is also dubious that any single genius can accomplish the task alone. Perhaps room could be created for an open-ended enquiry by many minds to focus on the problems of integration. Considerable progress might be made over time, which would be interesting in its own right, whether or not it would succeed in creating a pragmatic synthesis of which one could be justifiably confident that it did not fail to recognize and do justice to any crucial consideration. A decisive breakthrough would mean that we would have the opportunity to make our choices in the full light of available facts and reasons, and with justified confidence that we are not pushing in entirely the wrong direction.

Whether things would be better that way, we do not know for sure. Nor do we know for sure what presently available actions would best encourage such an outcome, should we decide to strive for it, for this is of course one of the big open-ended goals to which the argument applies: even if we think hard and honestly about this issue, we are apt to neglect at least one crucial consideration. Unless and until we achieve a dramatic enlightenment in our capacity for pragmatic synthesis, then on this and on other revolutionary prospects, we will continue to stake out our ethics and policy positions in the dark.

## References

- Bainbridge, William Sims, and Mihail C. Roco, eds. (2006), *Managing Nano-Bio-Info-Cogno Innovations*. Dordrecht: Springer.
- Bostrom, Nick (2002), "Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards", *Journal of Evolution and Technology* 9.
- — — (2003a), "Are You Living in a Computer Simulation?" *Philosophical Quarterly* 53 (211):243-255.



- — — (2003b), "Astronomical Waste: The Opportunity Cost of Delayed Technological Development", *Utilitas* 15 (3):308-314.
- — — (2005), "The Fable of the Dragon-Tyrant", *Journal of Medical Ethics* 31 (5):273-277.
- — — (2006), "Infinite Ethics", *Working manuscript*. Available from <http://www.nickbostrom.com/ethics/infinite.pdf>.
- Bostrom, Nick, and Toby Ord (2006), "The Reversal Test: Eliminating Status Quo Bias in Bioethics", *Ethics* 116 (4):656-680.
- Chambers, S. (2004), "Behind closed doors: Publicity, secrecy, and the quality of deliberation", *Journal of Political Philosophy* 12 (4):389-410.
- Crow, Michael M., and Daniel Sarewitz (2001), "Nanotechnology and Societal Transformation", in Albert H. Teich, Stephen D. Nelson, Celia McEnaney and Stephen J. Lita (eds.), *AAAS Science and Technology Policy Yearbook*, Washington, DC: American Association for the Advancement of Science, 89-101.
- Drexler, Eric (2003), "Nanotechnology Essays: Revolutionizing the Future of Technology (Revised 2006)", *AAAS EurekAlert! InContext* April.
- Drexler, K. E. (1992a), *Engines of Creation*. Oxford: Oxford University Press.
- — — (1992b), *Nanosystems: Molecular Machinery, Manufacturing, and Computation*. New York: John Wiley & Sons, Inc.
- Elam, Mark, and Margareta Bertilson (2002), "Consuming, Engaging and Confronting Science: The Emerging Dimensions of Scientific Citizenship", in Peter Healey (ed.), *STAGE ( HPSE-CT2001-50003) Final Report*, 121-158. Available from [http://www.stage-research.net/STAGE/documents/STAGE\\_Final\\_Report\\_final.pdf](http://www.stage-research.net/STAGE/documents/STAGE_Final_Report_final.pdf).
- Gilovich, Thomas, Dale W. Griffin, and Daniel Kahneman (2002), *Heuristics and biases: the psychology of intuitive judgement*. Cambridge, U.K.: Cambridge University Press.
- Gosseries, Axel (2005), "Publicity", in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*.
- Grice, Paul (1975), "Logic and Conversation", in Peter Cole and Jerry L. Morgan (eds.), *Syntax and Semantics, Vol. 3, Speech Acts*, New York: Academic Press, 41-58.
- Hanson, R. (1995), "Could Gambling Save Science? Encouraging an Honest Consensus", *Social Epistemology* 9:1:3-33.
- Hanson, Robin *Shall We Vote on Values, But Bet on Beliefs?* 2003. Available from <http://hanson.gmu.edu/futarchy.pdf>.
- Jaynes, E. T., and G. Larry Bretthorst (2003), *Probability theory: the logic of science*. Cambridge, UK: Cambridge University Press.
- Jones, C. I., and J. C. Williams (1998), "Measuring the social return to R&D", *Quarterly Journal of Economics* 113 (4):1119-1135.
- Joy, B. (2000), "Why the future doesn't need us", *Wired* 8.04.
- Kahneman, Daniel, and Amos Tversky (2000), *Choices, values, and frames*. Cambridge, UK: Cambridge University Press.

- Latour, Bruno (1987), *Science in action: how to follow scientists and engineers through society*. Cambridge, Mass.: Harvard University Press.
- Leigh, A., and J. Wolfers (2006), "Competing approaches to forecasting elections: Economic models, opinion polling and prediction markets", *Economic Record* 82 (258):325-340.
- Lenman, James (2000), "Consequentialism and Cluelessness", *Philosophy and Public Affairs* 29 (4):342-370
- Meade, Ellen E., David Stasavage, and London School of Economics and Political Science. Centre for Economic Performance. *Publicity of debate and the incentive to dissent evidence from the US Federal Reserve*. Centre for Economic Performance, London School of Economics and Political Science 2004. Available from <http://cep.lse.ac.uk/pubs/download/dp0608.pdf>.
- Rawls, John (1999), *The law of peoples*. Cambridge, Mass.: Harvard University Press.
- Roco, Mihail C., and William Sims Bainbridge (2003), *Converging technologies for improving human performance: nanotechnology, biotechnology, information technology and cognitive science*. Dordrecht; Boston, Mass.: Kluwer Academic Publishers.
- Salter, A. J., and B. R. Martin (2001), "The economic benefits of publicly funded basic research: a critical review", *Research Policy* 30 (3):509-532.
- Tetlock, Philip (2005), *Expert political judgment: how good is it? how can we know?* Princeton, [N.J.]: Princeton University Press.
- Vandermolen, Thomas D. (2006), "Molecular Nanotechnology and National Security", *Air & Space Power Journal* 20 (3):96-106.
- Wolfers, J., and E. Zitzewitz (2004), "Prediction markets", *Journal of Economic Perspectives* 18 (2):107-126.