Nesta...

Good and bad innovation: what kind of theory and practice do we need to distinguish them?

Geoff Mulgan

---

1. Introduction

Many firms, and many governments, are in favour of more innovation. They like to be seen to side with the new against the old. For the same reason, the majority of innovation agencies around the world promote innovation as a good in itself rather than as a means to ends, as do the vast majority of recent books about innovation. They may argue about the relative roles of states and markets. But nearly all are silent on the relative merits of different kinds of innovation.

Should they be? A moment's reflection shows that it's not altogether coherent (whether intellectually, ethically or in terms of policy) to simply be in favour of innovation. Some innovations are unambiguously good (like penicillin or the telephone). Others are unambiguously bad (like concentration camps or nerve gas). Many are ambiguous. Pesticides kill parasites but also pollute the water supply. New surveillance technologies may increase workplace productivity but leave workers more stressed and unhappy. Smart missiles may be good for the nations deploying them and terrible for the ones on the receiving end.

In finance, Paul Volcker, former Head of the US Federal Reserve, said that the only good financial innovation he could think of was the ATM. That was an exaggeration. But there is no doubt that many financial innovations destroyed more value than they created, even as they enriched their providers, and that regulators and policymakers failed to distinguish the good from the bad, with very costly results.

The traditional justification for a capitalist market economy is that the net effects of market-led innovation leave behind far more winners than losers, and that markets are better able to pick technologies than bureaucracies or committees.
But even if, overall, the patterns of change generate more winners than losers there are likely to be some, perhaps many, cases where the opposite happens. It would be useful to know.

But how should we judge? And what actions might flow from a more balanced assessment of technologies? No society delegates all decisions on new technologies to markets. Instead they tend to be highly regulated and constrained - whether to ensure that cars are not unnecessarily dangerous, or that foods don't contain excessive quantities of salt or sugar. We can argue about whether regulation is excessive or inadequate. But laissez-faire has not worked in any known society. Given that is the case, anyone in charge of directing public money to R&D or other innovation activities should want to ensure that more good rather than bad things are likely to result. The public, too, take a discriminating view of innovation. In countries like the UK, most are open to new ideas. But they want to know what outcomes these will achieve, and are sceptical of anyone who says that new things are inherently better, or that global competition is a sufficient justification for developing a particular new technology.

So what methods should be used to distinguish good from bad? Working at Nesta over the last few years, and engaging with a lot with innovation practitioners and academics, I've been surprised at the paucity of good answers. There is no shortage of approaches, some very impressive academic experts - Stilgoe, Callon, Wilsdon, Lövbrand, Stirling, Rayner, Hajer, Wynne, von Schomberg, to mention just a few - and some excellent overviews of the field, mainly focused on responsiveness and more open processes.

But the methods proposed are often hard to put into effect. This paper attempts to provide a possible framework that complements these. It recognises the inherent difficulties involved in assessing future possibilities, while arguing that intelligent judgements can guide allocations of money and the design of policies and regulations.

2. The challenges of assessing emerging technologies

How should a new field of technological advance be assessed - whether it's a new material like graphene or new methods in a field like medical robotics? There are no simple answers, though some agencies have committed themselves to better distinguishing good from bad innovation.

The European Commission, for example, recently promised to back 'responsible innovation' which is defined as an 'approach that anticipates and assesses potential implications and societal expectations… with the aim to foster the design of inclusive and sustainable research and innovation.' That's a worthy aim. But the definition is close to being tautologous, defining responsible innovation as innovation that takes account of its possible effects, without showing how this is to be done.

Part of the reason is that the disciplines that could have provided a more rigorous and useful approach have largely failed to do so. Economics has developed few coherent or comprehensive methods for analysing which kinds of innovation are good and which are bad. It can see when consumers do or don't want to buy something new, and the analysis of externalities can show in retrospect which innovations generate 'bads' (like pollution) as well as benefits (such as cheaper products). But economics offers no ways of doing so ex ante beyond traditional cost-benefit analysis. Economists from Marx to Erik Brynjolfsson have written about the distributional impact of technology, but on the whole it has been a minority interest, and has had more to say about technology in general (automation, ICT, robots) than about specific innovations.

Another potential source of methods is 'technology assessment', which became widespread in some places in recent decades, through formalised offices linked to governments or parliaments, and active networks of TA institutions. There are methods for 'Constructive Technology Assessment', Anticipatory Assessment, Real-time Technology Assessment, Value-Sensitive Design and others. But it's not clear how much influence these now have on key decisions – whether from public or private funding bodies. Much work has also been done on how individual scientists should think about their own responsibility, and there are some strong examples of institutions working to establish a balanced approach to complex areas of science, such as the Nuffield Bioethics Council.

As I show later, however, any attempts to assess technologies are fraught with difficulty - since we cannot know with much certainty how they may develop, let alone what effects they may have. As a result most innovation agencies primarily use innovation assessment tools that capture progress towards markets – and the perceptions of potential investors or consumers – or some rough and ready use of cost benefit analysis, rather than more comprehensive TA. Very little of the theoretical work that has been done on this question links to the practical needs of decision-makers.

3. Grounds for scepticism

There are at least four grounds for scepticism about *any* systematic attempts to assess emerging technologies:

The first is that no-one can predict how technologies will evolve. Conference speakers love recounting the many examples of people closely involved in key technology sectors - from computing and transport to energy - who dramatically misread how their field would develop. Being an expert is no guarantee of being able to make accurate predictions.

The second is that even if you can predict how a technology will evolve, it's very hard to predict who will benefit or suffer. In 1867, Marx wrote that the self-acting mule was threatening cotton spinners (a skilled job) and replacing them with unskilled kids. Serious industrial experts like Dr Ure agreed with him. In fact, the mule spinners weren't put out of jobs - changes in the way factories were managed meant these skilled cotton workers kept their highly paid jobs well into the 20th century. Forecasts on the effects of technology on jobs over the last 50 years have been equally inaccurate - indeed futurology has consistently exaggerated and misinterpreted the effects of automation on jobs.

The third is that it's impossible to define what the counterfactual to any given innovation is. A coal mine despoils nature and emits lots of CO_2. But if the alternative is to chop down and burn a large forest, the mine might be better both for nature and the climate. This concern presumably applies to other innovations too. If a nuclear war happens at some point, then we would probably be better off had nuclear weapons not been invented. But if you were considering whether to invent the atom bomb in Los Alamos in 1943, your choice wasn't between inventing nuclear weapons or nuclear weapons never existing, but between you inventing them now or someone else likely inventing them later.

The fourth is a general problem of anything future-oriented: the incumbents who may stand to lose most from an innovation are likely to be well-organised and powerful, while potential future beneficiaries may be powerless and lack a voice.

These are all reasons for caution and humility. In any system it's useful to allow a fair degree of freedom either for inventors or entrepreneurs. Excessive application of precautionary principles can inhibit very desirable progress.

But it's implausible to conclude that no scrutiny or debate is either useful or feasible. It's inherently unhealthy for any society to see technologies as things which emerge magically, and over which there is no possibility of control.

4. A framework for thought and action

So what are reasonable ways of assessing and then supporting innovations at different stages of their life-cycle that avoid these pitfalls? Here I briefly suggest some of the elements that would be needed.

i. A pragmatic and staged approach to knowledge, risk and uncertainty

It's very hard to know early on what innovations will turn into. This is the justification for some exploratory innovation relatively untrammeled by too many controls. Over time, however, the likely impacts become clearer. But this poses a big challenge. When a technology is young, it's hard to assess; when it's mature, it may be too late to reshape it.

So a rough compromise aims early on to explore possibilities and potential threats, and identify potential triggers, or irreversible steps, which could warrant more intensive scrutiny. For almost any technology, the moment of coming near to market, or being purchased for use by a government, should be a trigger. But for others - where the potential risks are particularly large - the critical moment may be an important technical breakthrough (AI falls into this category). A lot of work is underway on the governance of these kinds of risks. The key is to keep choices open, not to close them off. Many of the methods being suggested to support responsible innovation offer stage-gate models that ask what kinds of information are needed at each stage.

ii. Distinguish different types of good, and guide innovation to the ones with the greatest potential benefit

The value of innovation will depend on the kinds of good or service that it leads to. There are at least five very different kinds of good (though these are sometimes confused or conflated in economics), and policy should be designed to distinguish them.

The first category includes goods with **network effects or positive externalities** that become more valuable if others are also consuming them—like telephones and other network technologies. Public health would also fall into this category. Because of their positive externalities, there is a case for judging growth in consumption of these as more valuable to an economy than growth of other kinds of consumption. For similar reasons, innovation that contributes to goods of this kind should be a higher priority for any society.

A second category encompasses more **normal commodities** like clothing or tins of baked beans. Whether or not I consume these doesn't have much impact for better or worse on other people. These are the types of good around which most economics is shaped. Their profitability can be improved by reducing inputs or increasing the extent to which they are reused or recycled. But their external effects are modest.

The third category contains **goods that destroy value for some while creating it for others**. These include cars (which create pollution, noise, and dislocation for those who don't own them), airlines (which disproportionately worsen climate change) and many other industries. Economics recognizes that they produce 'negative externalities'. It measures these when doing exercises in cost-benefit analysis, and policymakers try to internalize them through taxes or regulations. But only the most obvious and material externalities are recognized in economics; and even the ones that are recognized aren't measured in GDP or company accounts. Later on I look in more detail at these distributional questions.

Fourth, there are what the economist Fred Hirsch called '**positional goods**', whose value comes from their exclusivity; stately homes and tropical islands developed for luxury tourism are classic examples as is getting on the guest list for the best parties or membership of the most exclusive golf clubs. Their scarcity can be physical—meaning that a good is scarce in some absolute or socially imposed sense (such as land used for pleasure and personal enjoyment), or the scarcity can be social—meaning that it can be subject to congestion or crowding through more extensive use (as in the case of a privileged education). It's hard to see why innovation policy should ever prioritise goods of this kind - though Concorde, the flagship of UK and French innovation policy in the 1960s, was arguably as much about positional value as absolute value.

Finally there are **goods whose very value comes from the negative externalities** created for others. At the extreme are weapons: teenagers buy knives and nations build nuclear missiles to frighten others. Their negative impact on lived value is not an unfortunate by-product but rather integral. Roughly half of all direct public spending on innovation in the US, Russia, UK, France and other countries goes on military technologies.

It should be obvious that innovation spending should treat these goods very differently. Yet most funding rules make no distinction between them, and this is particularly a flaw of tools such as R&D tax credits.

iii. Consider distributional effects - even though these are hard to predict

In retrospect it should be fairly straightforward to look in a rounded way at the distributional effects of new technologies, even if this is very difficult in advance. You would expect to look at the following categories:

- Gains to consumers (who presumably won't buy the new thing unless they perceive gains, though in the longer term they may suffer losses, for example if a new kind of food harms their health)
- Gains (and sometimes losses) to investors (who are presumed to be able to look after their own interests)
- Gains to some workers who get better jobs
- Losses to other workers who lose their jobs or see a cut in pay
- A mix of gains and losses to natural capital

Such assessments can be comparative static or dynamic: e.g. will this cluster of technologies create new jobs, business clusters &c, or have knock-on effects on many other sectors in the manner of GPTs? They're rarely easy because of the sheer number of possible factors involved. But they can at least in principle be reasonably objective and avoid cultural judgements ('smartphones/satellite TV/ Snapchat will destroy our culture…'). They can also be done in real-time, tracking the actual effects of technologies (e.g. what effects are different national choices over nuclear power and solar having distributionally?)

iv. Assess technologies ethically - and apply the golden rule test

The ethical assessment of innovations is more complex. A useful starting point is to look for compatibility with the golden rule: do unto others what you would have them do unto you. Good innovations are ones that we would want for ourselves and those we love. Bad ones (like many of the financial innovations of the 1990s and 2000s) clearly breach the golden rule in that the providers would not want themselves to be consumers.

Taking a broad look at technology, it's obvious that some innovations are clearly compatible with the golden rule while others support predation, making it easier to control, exploit, or conquer. Technologies for war, or for surveillance, are by their very nature contrary to the spirit of the golden rule. There is no missile system, directed energy weapon, or security camera for which it makes sense for others to do unto you as you would do unto them. Computer viruses are very obvious predators, secretly stealing your credit card details, often to support organized-crime syndicates.

New fields of technology bring new patterns of predation as well as empowerment. The internet of things with its arrays of sensors is a good example. There are great potential advantages to the efficiency of transport, energy and security. But almost any action may now be generating data to be matched, mined, and commercialised, without your knowledge or consent. You cannot help but generate data for them, and you may not mind since there is no immediate harm to you. But something of you is being taken without your permission and without any reciprocal benefit to you.

Other technologies are more obviously compatible with the golden rule - like mobile phones that only become valuable if others have them, oral rehydration therapy, yellow fever vaccines, or new crops enriched with vitamins. Others sit in between, like cars that simultaneously provide value to their owners but also take away clean air, space, and peace from people who don't have them. An interesting current case study is the attempt by Facebook to introduce free internet access in India, which shows just how different perceptions of value can be.

Then, too, there are technologies of predation that benefit people but leave nature worse off. How you view these depends on just how human-centric your world-view is. To some eyes, large-scale mining, whether of the land or the oceans, is by its nature predatory (even when it doesn't come with the messy combination of displacement, abuse, and occasional windfall pay-offs to indigenous communities). To others, it's just the good fortune that humans enjoy thanks to their evolutionary superiority. There has been great advance in methods for understanding the impact of technologies, or whole industries, on natural capital. These don't quite fit with the human-centric golden rule and provide an additional lens for any future TA. Such a combination of golden rule-based evaluations, and natural capital evaluations, would be an obvious option for public innovation agencies. At the very least this is a healthy conversation for any society to have.

5. Agency and assessment

Who might make these assessments? And who might act on them? Interest in this field is currently limited to a few big funders (like the European Commission and research councils). But there are many other potential user/creators of more systematic assessments of innovations at different stages of development:

- **Governments**, innovation agencies and with them regulators and policymakers, at national or transnational levels. Most TA is attached to governments or parliaments.
- **Businesses and investors** - including ones committed to corporate reporting on environmental, social and governance issues, or employee organisations. This is a very underdeveloped field.
- **Universities and research centres** - which have more capabilities, and the whole panoply of science and technology studies to draw on
- **Citizens, and NGOs, social movements, media** - aiming to represent a civic interest, though these usually campaign for or against technologies rather than playing an overt role in assessment.

Each of these might judge innovations in different ways. But there may be some benefit in common frameworks, language and analytical technique, for example around descriptions of risk and opportunity, or design of experiments to improve knowledge about these. And there should be great advantages in more truly society-wide processes that debate in a rounded way the possibilities and threats of emerging technologies.

6. Outcomes and actions

What effect should an assessment regime seek to have? How might it be acted on? Such methods are only worthwhile if they lead to action:

- To *encourage* – where a major potential is diagnosed but without adequate economic or other support to drive its development. Drugs for rare or poor country diseases are a good example.
- To *block* – where the harms are serious and the priority is to slow down, redirect or stop R&D and technology development
- To *redirect* – where they may aim to guide R&D or commercial investment, e.g. drones for social good, or regulation.

This obviously raises the question of how the arbiters of spending in fields like R&D could be more systematically held to account for their use of assessment tools to guide decisions. This would be a useful outcome from any work in this field – a more systematic combination of 'golden rule based evaluations' and natural capital evaluations.

Next steps

This note is a sketch and doesn't pretend to be definitive. It aims to contribute to the conversation about how we fill the space between two undesirable, and intellectually incoherent, poles:

- On the one hand a hard precautionary principle which tries to stop any discovery or invention that might bring with it risks and losses;
- On the other hand the view that, because any assessment is difficult, we should just let technologies develop according to the push from scientists and inventors, and pull from markets

Neither position adds up. But finding a sensible path between these extremes is complex, even though it's clearly important for innovation ministries, agencies and funders.

I'd welcome comments and suggestions on who is doing this well and more generally on better approaches than the ones I've suggested (I'm less interested in comments which just reiterate the difficulties).

Geoff Mulgan is Chief Executive of Nesta. He has worked in governments (as head of policy for Prime Minister Tony Blair and head of the UK government Strategy Unit, and as an adviser for many other governments around the world); in telecoms (in which he has a PhD, and has been an investor, funder and researcher on digital economies); as a social entrepreneur (for example, establishing the global social innovation exchange and a network of new schools); and as an author of books translated into dozens of languages. He has been a visiting professor at University College London, LSE, Melbourne University and is currently senior visiting scholar at Harvard.

info@nesta.org.uk @nesta_uk facebook.com/nesta.uk www.nesta.org.uk

Nesta is a registered charity in England and Wales with company number 7706036 and charity number 1144091.
Registered as a charity in Scotland number SCO42833. Registered office: 1 Plough Place, London, EC4A 1DE.

July 2016