CHAPTER 5
Decisive strategic advantage

A question distinct from, but related to, the question of kinetics is whether there will there be one superintelligent power or many? Might an intelligence explosion propel one project so far ahead of all others as to make it able to dictate the future? Or will progress be more uniform, unfurling across a wide front, with many projects participating but none securing an overwhelming and permanent lead?

The preceding chapter analyzed one key parameter in determining the size of the gap that might plausibly open up between a leading power and its nearest competitors—namely, the speed of the transition from human to strongly superhuman intelligence. This suggests a first-cut analysis. If the takeoff is *fast* (completed over the course of hours, days, or weeks) then it is unlikely that two independent projects would be taking off concurrently: almost certainty, the first project would have completed its takeoff before any other project would have started its own. If the takeoff is *slow* (stretching over many years or decades) then there could plausibly be multiple projects undergoing takeoffs concurrently, so that although the projects would by the end of the transition have gained enormously in capability, there would be no time at which any project was far enough ahead of the others to give it an overwhelming lead. A takeoff of *moderate* speed is poised in between, with either condition a possibility: there might or might not be more than one project undergoing the takeoff at the same time.[1]

Will one machine intelligence project get so far ahead of the competition that it gets a *decisive strategic advantage*—that is, a level of technological and other advantages sufficient to enable it to achieve complete world domination? If a project did obtain a decisive strategic advantage, would it use it to suppress competitors and form a *singleton* (a world order in which there is at the global level a single decision-making agency)? And if there is a winning project, how "large" would it be—not in terms of physical size or budget but in terms of how many people's desires would be controlling its design? We will consider these questions in turn.

Will the frontrunner get a decisive strategic advantage?

One factor influencing the width of the gap between frontrunner and followers is the rate of diffusion of whatever it is that gives the leader a competitive advantage. A frontrunner might find it difficult to gain and maintain a large lead if followers can easily copy the frontrunner's ideas and innovations. Imitation creates a headwind that disadvantages the leader and benefits laggards, especially if intellectual property is weakly protected. A frontrunner might also be especially vulnerable to expropriation, taxation, or being broken up under anti-monopoly regulation.

It would be a mistake, however, to assume that this headwind must increase monotonically with the gap between frontrunner and followers. Just as a racing cyclist who falls too far behind the competition is no longer shielded from the wind by the cyclists ahead, so a technology follower who lags sufficiently behind the cutting edge might find it hard to assimilate the advances being made at the frontier.[2] The gap in understanding and capability might have grown too large. The leader might have migrated to a more advanced technology platform, making subsequent innovations untransferable to the primitive platforms used by laggards. A sufficiently pre-eminent leader might have the ability to stem information leakage from its research programs and its sensitive installations, or to sabotage its competitors' efforts to develop their own advanced capabilities.

If the frontrunner is an AI system, it could have attributes that make it easier for it to expand its capabilities while reducing the rate of diffusion. In human-run organizations, economies of scale are counteracted by bureaucratic inefficiencies and agency problems, including difficulties in keeping trade secrets.[3] These problems would presumably limit the growth of a machine intelligence project so long as it is operated by humans. An AI system, however, might avoid some of these scale diseconomies, since the AI's modules (in contrast to human workers) need not have individual preferences that diverge from those of the system as a whole. Thus, the AI system could avoid a sizeable chunk of the inefficiencies arising from agency problems in human enterprises. The same advantage—having perfectly loyal parts—would also make it easier for an AI system to pursue long-range clandestine goals. An AI would have no disgruntled employees ready to be poached by competitors or bribed into becoming informants.[4]

We can get a sense of the distribution of plausible gaps in development times by looking at some historical examples (see Box 5). It appears that lags in the range of a few months to a few years are typical of strategically significant technology projects.

Box 5 Technology races: some historical examples

Over long historical timescales, there has been an increase in the rate at which knowledge and technology diffuse around the globe. As a result, the temporal gaps between technology leaders and nearest followers have narrowed.

China managed to maintain a monopoly on silk production for over two thousand years. Archeological finds suggest that production might have begun around 3000 BC, or even earlier.[5] Sericulture was a closely held secret. Revealing the techniques was punishable by death, as was exporting silkworms or their eggs outside China. The Romans, despite the high price commanded by imported silk cloth in their empire, never learnt the art of silk manufacture. Not until around AD 300 did a Japanese expedition manage to capture some silkworm eggs along with four young Chinese girls, who were forced to divulge the art to their abductors.[6] Byzantium joined the club of producers in AD 522. The story of porcelain-making also features long lags. The craft was practiced in China during the Tang Dynasty around AD 600 (and might have been in use as early as AD 200), but was mastered by Europeans only in the eighteenth century.[7] Wheeled vehicles appeared in several sites across Europe and Mesopotamia around 3500 BC but reached the Americas only in post-Columbian times.[8] On a grander scale, the human species took tens of thousands of years to spread across most of the globe, the Agricultural Revolution thousands of years, the Industrial Revolution only hundreds of years, and an Information Revolution could be said to have spread globally over the course of decades—though, of course, these transitions are not necessarily of equal profundity. (The *Dance Dance Revolution* video game spread from Japan to Europe and North America in just one year!)

Technological competition has been quite extensively studied, particularly in the contexts of patent races and arms races.[9] It is beyond the scope of our investigation to review this literature here. However, it is instructive to look at some examples of strategically significant technology races in the twentieth century (see Table 7).

With regard to these six technologies, which were regarded as strategically important by the rivaling superpowers because of their military or symbolic significance, the gaps between leader and nearest laggard were (very approximately) 49 months, 36 months, 4 months, 1 month, 4 months, and 60 months, respectively—longer than the duration of a fast takeoff and shorter than the duration of a slow takeoff.[10] In many cases, the laggard's project benefitted from espionage and publicly available information. The mere demonstration of the feasibility of an invention can also encourage others to develop it independently; and fear of falling behind can spur the efforts to catch up.

Perhaps closer to the case of AI are mathematical inventions that do not require the development of new physical infrastructure. Often these are published in the academic literature and can thus be regarded as universally available; but in some cases, when the discovery appears to offer a strategic advantage, publication is delayed. For example, two of the most important ideas in public-key cryptography are the Diffie–Hellman key exchange protocol and the RSA encryption scheme. These were discovered by the academic community in 1976 and 1978, respectively, but it has later been confirmed that they were known by cryptographers at the UK's communications security group since the early 1970s.[20] Large software projects might offer a closer analogy with AI projects, but it is harder to give crisp examples of typical lags because software is usually rolled out in incremental installments and the functionalities of competing systems are often not directly comparable.

Table 7 ***Some strategically significant technology races***

	United States	Soviet Union	United Kingdom	France	China	India	Israel	Pakistan	North Korea	South Africa
Fission bomb	1945	1949	1952	1960	1964	1974	1979?	1998	2006	1979?
Fusion bomb	1952	1953[11]	1957	1968	1967	1998	?	—	—	—
Satellite launch capability	1958	1957	1971	1965	1970	1980	1988	—	1998?[12]	—[13]
Human launch capability	1961	1961	—	—	2003	—	—	—	—	—
ICBM[14]	1959	1960	1968[15]	1985	1971	2012	2008	—[16]	2006	—[17]
MIRV[18]	1970	1975	1979	1985	2007	2014[19]	2008?			

It is possible that globalization and increased surveillance will reduce typical lags between competing technology projects. Yet there is likely to be a lower bound on how short the average lag could become (in the absence of deliberate coordination).[21] Even absent dynamics that lead to snowball effects, some projects will happen to end up with better research staff, leadership, and infrastructure, or will just stumble upon better ideas. If two projects pursue alternative approaches, one of which turns out to work better, it may take the rival project many months to switch to the superior approach even if it is able to closely monitor what the forerunner is doing.

Combining these observations with our earlier discussion of the speed of the takeoff, we can conclude that it is highly unlikely that two projects would be close enough to undergo a fast takeoff concurrently; for a medium takeoff, it could easily go either way; and for a slow takeoff, it is highly likely that several projects would undergo the process in parallel. But the analysis needs a further step. The key question is not how many projects undergo a takeoff in tandem, but how many projects emerge on the yonder side sufficiently tightly clustered in capability that none of them has a decisive strategic advantage. If the takeoff process is relatively slow to begin and then gets faster, the distance between competing projects would tend to grow. To return to our bicycle metaphor, the situation would be analogous to