Original research article, History and energy transitions

Up in the air: Barriers to greener air traffic control and infrastructure lock-in in a complex socio-technical system

Graham Spinardi

Science, Technology and Innovation Studies, University of Edinburgh, Edinburgh EH1 1LZ, United Kingdom

ARTICLE INFO

Article history:
Received 29 September 2014
Received in revised form 19 November 2014
Accepted 20 November 2014

Keywords:
Green aviation
Lock-in
Socio-technical systems

ABSTRACT

Greater automation of air traffic control (ATC) could reduce aviation's climate change impacts, but improvements predicted long ago have been slow to happen. This resistance to ATC modernisation is framed as an issue of lock-in, and the detailed case study described here enables an analysis of the factors involved in slowing change. Although the classic lock-in effects of 'increasing returns' and 'network externalities' are important, a major barrier to modernisation is due to the political and organisational challenges of coordinating change across a large, complex socio-technical system. However, lock-in effects are crucial with respect to the perceived increasing returns accrued from experience with manual ATC operations, and the difficulty of quantifying the risks of automation (particularly as regard the use of complex software) is a major barrier to further improvements. Overcoming this obstacle to further automation depends on finding ways to test and operate new ATC software and procedures without compromising safety.

© 2014 The Author. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/3.0/).

1. Introduction

In early 2010 the Obama Administration set out a new vision for NASA, with, amongst other changes, a refocusing towards 'green aviation', and a budget request that allocated $20 million per year 'to support NASA's environmentally responsible aviation program' [31]. The green aviation label encompassed many activities, including work on aerodynamics, engine technology and biofuels, as well as on air traffic control (ATC).

Although ATC improvements only offer modest environmental benefits, they have the advantage that they could be implemented without the need to replace current aircraft. Such ATC-driven reductions in aviation's climate change impact could thus in principle be implemented relatively quickly. The two main alternative approaches will take decades to have significant impacts. Aviation biofuel could be developed to make air travel sustainable, but the sheer amount required, along with the problematic nature of most current sources, make a rapid transition unlikely. Likewise, although greener aircraft with markedly better fuel-efficiency are feasible [33], their commercial viability is less certain [50]. Moreover, even if radically greener airliners can be built that are socially acceptable, it would be many years before the current aircraft inventory is completely replaced, whereas improvements in ATC efficiency would provide environmental benefits with both current and future aircraft.

Advances in ATC technology have long been predicted. A 1981 RAND report [70, p.2] noted that:

The prospect of almost total automation is no longer only science fiction. Computers are powerful and fast enough to project aircraft flight paths far into the future, to automatically correct them when they conflict with the anticipated flight profiles of nearby aircraft, and to digitally transmit the revised clearances up to the aircraft. Machines can continuously compute and update delay predictions, so that aircraft can be slowed at fuel-efficient higher altitudes when airports are operating at peak capacities.

However, over three decades later such levels of automation are still not implemented. Drawing on interviews with key personnel at NASA, and on analysis of NASA documentation, as well as of the trade and secondary literature, this paper describes the first major use of automation in the US ATC system, the Traffic Management Advisor (TMA) developed by the NASA Ames Research Center. Set in the context of broader US ATC developments, this case study is used to address three questions. What are the obstacles to the implementation of more automated ATC? Can these obstacles be understood in terms of 'lock-in' of the existing socio-technical system? And what measures could be used to overcome such lock-in?

* Tel.: +44 0131 650 6394.
E-mail address: g.spinardi@ed.ac.uk

http://dx.doi.org/10.1016/j.erss.2014.11.006
2214-6296© 2014 The Author. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/3.0/).
2. Aviation, climate change, and air traffic control

Aviation contributes to climate change by increasing levels of greenhouse gases such as carbon dioxide, ozone and water vapour, and by stimulating the creation of cirrus cloud around 'contrails' (condensation trails) that reflect more energy back to the earth than they reflect away from the earth. Total 'radiative forcing'—simply put, atmospheric warming—due to aviation is thought to be about three times that due to aircraft carbon dioxide emissions alone [60, p. 18]. Using biofuels (assuming that these can be sourced in a way that is both environmentally and economically desirable) would thus only be a partial solution.

More efficient ATC offers a ‘win-win’ outcome, with both environmental and economic benefits. For example, a 2000 report on aviation and the environment by the US Government Accounting Office (GAO) noted that:

Operational improvements in such areas as communications, navigation, surveillance, and air traffic management could also lead to reductions in aircraft emissions. . . . improvements in air traffic management worldwide could reduce the annual consumption of aircraft fuel by 6 to 12 percent over the next 20 years [28, p. 22].

The potential environmental benefits lie in two areas. First, better ATC could make air travel more fuel-efficient. The ideal flight path for an aircraft would be for it to be able to operate as if there were no other aircraft around, and so no potential conflicts that would force it to take a sub-optimum flight path. An optimum flight path, in terms of fuel efficiency, would not only take the shortest route, and involve no ‘stacking’ while waiting for a landing slot, but it would also fly at the optimum (high) altitude for most of the journey, with a continuous ‘cruise climb’ after take-off, and land following a continuous ‘idle thrust’ descent. In particular, enabling aircraft to utilise idle thrust descents, with low engine power, is a key challenge for ATC because of the complexity and time pressure of many different types of aircraft converging on a limited landing space at an airport. ATC automation enables trajectories to be predicted, and adjustments made, further from the airport so that the final descent can be smooth and energy efficient.

Second, greater ATC automation could enable aircraft to avoid areas most likely to produce contrails, thus minimising the radiative forcing resulting from cloud formation, although there would be trade-offs involved in not flying the most direct routes [64]. Aviation-induced cirrus cloud formation has a potentially large impact on radiative forcing that cannot be ameliorated by the use of biofuels or improved aircraft efficiency [72, p. 743]. Sophisticated ATC might also enable aircraft to fly below contrail-prone altitudes although again there would be a trade-off with fuel-efficiency, and maybe objections on grounds of comfort and safety.

3. The challenge of lock-in

There are many obstacles to the fundamental transitions in energy production and use that are required to limit climate change. Technological innovation is necessary, but it is not sufficient to bring about these transitions. As Sovacool [63, p. 1; see also [62]] notes, much research on energy has had too narrow a focus on technology and economics, while downplaying ‘the human dimensions of energy use and environmental change.’ At the level of economics it is clear that market forces alone will not bring about transitions quickly enough, and that economic instruments such as carbon taxes or emissions trading are often poorly focussed [23] and hard to implement without unintended consequences [39]. At the individual level, exhortations to adopt greener lifestyles have limited success, are inappropriate for much of the world’s population that live in poverty [3], and flounder due to inertia even when the financial payback is clear [65] or because of the lack of social acceptance of novel solutions.

With regard to US ATC technology the key question is why it has not made the transition to the greater levels of automation that would enable greener air travel. It is thus appropriate to address this question through the conceptual lens of lock-in theory. The idea that the adoption of more environmentally desirable technologies is prevented by lock-in builds on work by Arthur, David and others on ‘path dependency’. Arthur [2, p. 116] argues that technologies get locked in because ‘the more they are adopted, the more experience is gained with them, and the more they are improved’, and thus ‘a technology that by chance gains an early lead in adoption may eventually “corner the market” of potential adopters, with the other technologies becoming locked out.’

Alongside this ‘increasing returns’ effect, a second concept underpinning the idea of lock-in hinges on the role of ‘network externalities’. Although he did not use this term in his 1985 paper, this idea is central to David’s iconic, though contested (see [46]), QWERTY keyboard example. David [10, p. 334] argues that the history of QWERTY shows that what many consider an inferior technology remains locked in because of ‘technical interrelatedness, economies of scale, and quasi-reversibility of investment’ (his italics). In other words, there was a strong linkage between the typewriter keyboard design and the expertise to type on it quickly. Thus, the more that one keyboard design dominated, the more it paid to be skilful in its use, and once such a large stock of keyboards and of people skilled in their use existed, it became increasingly hard for a competitor to gain traction.

Increasing returns and network externalities provide mechanisms for understanding how an inferior technology might persist in the face of superior alternatives. Previous studies of technological lock-in have focussed both on particular artefacts – e.g., the light water nuclear reactor [7] and the gasoline car [8] – and on large technological systems [66]. Lock-in has been highlighted as a particular concern for infrastructure-dependent vehicle technologies because of ‘high infrastructure investment costs and the presence of network externalities’ [67, p. 98]. With automobiles or other vehicles, lock-in can be conceptualised as hinging on consumer or operator choice, and potential policy options include whether to support more R&D on new vehicle technologies or to support infrastructure development [67]. However, such an analysis focuses on the way that the infrastructure inhibits transitions in vehicles, but ignores the possibility that infrastructure improvements may themselves be inhibited by lock-in.

Unlike with vehicles, ATC technology cannot be understood as a single technical artefact, amenable to technological substitution based on consumer or operator choice; rather, it is part of the infrastructure that makes air travel possible. ATC is a large technological system with capabilities stemming from the interactions of many elements, including radars, communications systems, software, and human operators. Thus, there are not only lock-in issues to consider, but also the challenges of implementing change in large socio-technical systems [38] involving complex products and systems [36], often requiring large investments from both public and private sources [32]. The nature of the technology and its organisational and political context are thus likely to make radical innovation difficult, as is the emphasis in aviation on ‘high reliability’ [43] and risk minimisation [15]. These factors present challenges to the explanatory utility of an approach based solely on economics-derived lock-in theory. Accordingly an interdisciplinary approach is adopted here, because, as noted by Sovacool [63, p. 26], ‘the energy problems facing society cut across academic disciplines.’
Although components have been modernised over the years, the basic way that ATC operates has remained largely the same, with only limited automation. Automation matters because the traditional approach to ATC involving air traffic controllers manually guiding aircraft through airspace sectors cannot provide sufficiently fine control to enable the most fuel-efficient operation. The next section describes how traditional ATC developed, with most efforts at automation stymied, and this is followed in Section 5 by an account of the development and implementation of NASA’s TMA technology that enables aircraft arrivals to be more efficiently managed.

Section 6 then describes the potential for more fully automated ATC, and the obstacles to its implementation. In Section 7 the case study is analysed in terms of the economic lock-in concepts of increasing returns and network externalities. Other reasons for the lack of progress are discussed, and some policy implications set out in Section 8.

4. The development of US air traffic control

Reducing aviation’s environmental impacts has not historically been central to ATC developments: the main driver of change has been maintaining safety while increasing capacity. Because any collision could cause considerable (and highly visible) loss of life, it has long been seen as crucial to maintain sufficient separation between aircraft. The first moves towards a US system began in the 1920s as increasing numbers of flights raised concerns about congestion and safety. However, the modern system of air traffic management only took shape in the 1950s when civilian radar systems were installed (for the early history, see [42]). These ground-based radars could track aircraft, thus enabling air traffic controllers to plot their movements and provide guidance by radio links to help keep sufficient spacing between aircraft. A further development came in the 1960s with the implementation of secondary radar in which a small radar transponder carried by aircraft provides information as to its identity.

In this traditional ATC, aircraft are passed from sector to sector, and guided through sectors under voice control by air traffic controllers using radar information displayed on screens. Between airports aircraft move along predefined air routes (known as ‘victor airways’ below 18,000 feet and ‘jetways’ above that). Having these routes meant that radars did not need to cover the whole of the US, and they also facilitate traffic management by air traffic controllers. Current rules applicable to commercial flights specify that in most cases aircraft must be kept at least 5 nautical miles apart horizontally and 1000 feet vertically.

These ATC procedures enabled growth in air traffic whilst maintaining safety (for exceptions, see [71], pp. 25–30). However, they do not provide the most fuel-efficient air travel, and have struggled to cope with congestion at busy airports. Many aircraft converging on an airport with limited landing capacity poses a particular challenge, and there is a clear trade-off between fuel efficiency and risks to safety. Maintaining safe separation with high levels of traffic can mean asking pilots to slow down or put their aircraft into holding patterns, but ‘holding traffic at low altitudes is not fuel efficient’ [73, p. 2].

Three developments spurred ATC improvement after declining investment during the 1970s [6, p. 4]. First, there was the 1973 ‘oil crisis’ that resulted from an embargo by Arab oil producers in protest at the US providing assistance to Israel in the Yom Kippur War. Greatly increased oil prices raised the cost of aviation fuel, and led to a concern about fuel efficiency. Second, the 1978 Airline Deregulation Act not only increased passenger numbers, but also allowed airlines to move to ‘hub and spoke’ operations producing increased traffic at hub airports [59, p. 50]. Finally, in 1981 there was a national air traffic controllers’ strike in the USA. Although unsuccessful (indeed catastrophic for the union and strikers), the strike raised awareness of the stressful nature of the job and the ageing nature of much of the equipment [49,52].

As a consequence the Federal Aviation Administration (FAA) established a ‘multibillion-dollar modernisation effort’ [25, p. 2] in its 1983 National Airspace System (NAS) Plan. Amongst the claimed benefits of the plan was that: ‘By permitting flight paths to be less circuitous, automating air route traffic control centres would save fuel’ [6, p. 16]. Assuming that Congress provided funding, the NAS Plan was expected to be completed by the mid-1990s, and would involve ‘the highest practical level of air traffic control automation’ [58, p. 10].

Central to this modernisation plan was the Advanced Automation System (AAS) which was ‘expected to include the new automated capabilities needed to cope with predicted increases in air traffic and to provide operational benefits to users, such as more fuel-efficient routes’ [24, p. 1]. Much of this effort comprised a range of technologies geared towards improving data processing, communications, displays, and radar systems and other devices for determining the location and speed of aircraft. It was thus intended to modernise the existing approach: ‘Each of the devices or systems mentioned above improves one aspect of the nation’s ATC system, but ATC authority remains firmly in the eyes, ears, and minds of human beings poised over radar scopes’ [70, p. 8].

However, the FAA also had a plan for more radical change in US ATC, one that would introduce full automation. A 1981 report prepared for the FAA set out this concept:

Suppose we could virtually replace these fallible human beings with a set of computer modules which could manipulate aircraft tracks so well that human intervention with individual aircraft would be necessary only in response to a major perturbation (e.g., a massive computer failure, or extensive storm-front passage). Suppose this computer system were able to automatically compute conflict-free clearances for aircraft under surveillance, to automatically transmit these clearances in a timely fashion, and to automatically monitor for compliance, taking corrective action as required [70, p. 8].

This fundamental ambition was at the heart of the FAA’s Automated En-Route Air Traffic Control (AERA) programme, described as ‘the most elaborate part of the FAA’s $32 billion plan to automate air traffic control’ [41, p. 77]. Developed by the Mitre Corporation, the AERA software was designed to compute aircraft trajectories based on radar data and local wind speed predictions, and thus project aircraft flight paths up to twenty minutes into the future [41, pp. 81 & 86]. Impending conflicts would be detected, and alterations made if necessary, according to certain priorities (such as whether an aircraft had already started to descend for landing, in which case it could receive priority to stay on its original course).

However, this attempt at ATC automation was largely unsuccessful – ‘sunk by unrealistic specifications and human factors difficulties, among other problems’ [55, p. 20]. In particular, ‘the AERA programme was never able to develop software that could do what was promised. Vastly exaggerated claims fell flat because the underlying science and mathematical algorithm simply did not exist at that time.’

As a result most of the AAS programme was cancelled in 1994 [55, p. 24]. According to the FAA’s Steven Zaidman: ‘We shot for the moon. We tried to do advanced technology, computer

---

1 Email from Heinz Erzberger, November 30, 2011.
replacements, new procedures, new software, and new decision support services all at once. We didn’t realise the full scope of human factors. We put too much risk in the programme in terms of pushing technology too fast. We underestimated the magnitude of the change’ (in [55], pp. 24 & 25).

However, similar concepts resurfaced in January 1995 when the Report of the RTCA Board of Director’s Select Committee on Free Flight was published. Formerly known as the Radio Technical Commission for Aeronautics, RTCA is a non-profit organisation involved in developing standards for the aviation industry. Its 1995 report defined free flight as a ‘safe and efficient flight . . . in which [pilots] have the freedom to select their path and speed in real time. Air traffic restrictions are only imposed to ensure separation, to preclude exceeding airport capacity, to prevent unauthorised flight through special use airspace (for example, airspace restricted for military operations), and (otherwise) to ensure safety in flight’ (in [55], p. 33).

A key capability planned for Free Flight was a ‘conflict probe’ based on the User Request Evaluation Tool (URET) that had been developed in AERA, and this was deployed at Indianapolis and Memphis in the late 1990s. This conflict probe ‘processes real-time flight plan and track data’ which are ‘combined with site adaptation, aircraft performance characteristics, and winds and temperatures from the National Weather Service in order to build four-dimensional flight profiles, or trajectories’ [5, p. xii].

URET thus helps air traffic controllers avoid conflicts, enabling ‘a greater number of user-preferred flight paths, and allows increased system capacity while maintaining the current level of safety’ [40, p. 5]. Amongst the benefits of this is that aircraft can remain at their preferred altitudes for longer, resulting in reduced fuel burn. A 2001 report noted that even the limited deployment of URET provided fuel savings that were ‘saving airlines approximately $950,000 annually’ ([40], p. 6; see, also [53]).

Nevertheless, the FAA’s major effort to modernise ATC up to the 1990s produced limited benefits, with the widespread perception that ‘success has been minimal’ [55, p. 19]. In particular, these efforts did not address one of the main challenges in ATC: the convergence of aircraft as they seek to land at a busy airport. Airport convergence is the key bottleneck for ATC, with the airline industry’s hub and spoke approach leading to severe over-crowding at certain airports. This meant delays, fuel-inefficient aircraft operation, and high levels of stress for air traffic controllers.

5. NASA and the algorithmic approach to ATC

This convergence problem was addressed by Heinz Erzberger and his colleagues at NASA Ames in California. Whereas others – such as in the FAA’s overly ambitious AERA – had conceptualised fully automated ATC ‘capable of flying an aircraft automatically from takeoff to landing’, Erzberger sought to tackle a more narrowly defined, but crucial part of the problem: ‘the problem of automating a segment of the flight during which pilot workload is particularly heavy . . . within the terminal area, roughly within a 50-mile radius of an airport’ [16, p. 2]. Initially, the primary concern of this work was with the heavy workload of both the pilot and air traffic controllers.

However, fuel efficiency became an important factor with the 1973 ‘oil crisis’. From 1973 to 1987 NASA as a whole focussed much of its aviation work on this issue in its Aircraft Energy Efficiency Programme [4]. Erzberger sought to contribute to this goal, though as fuel price concerns faded during the 1980s the work at Ames came to be justified more in terms of reducing congestion.

Erzberger’s speciality was trajectory algorithms; his earlier work on the Space Shuttle had focussed on idle thrust descent trajectories, which by their very nature are fuel-efficient. According to Erzberger, ‘we built these descent trajectories to be an idle thrust descent all the way to the bottom . . . you’re flying at thirty-five thousand and you know the route you’re going to fly. Now imagine you have to pick a time, while you’re flying, to land down there. You’re given the command, OK, close your throttles, never touch them again, but be assured that you’re going to get there without turning the throttles back on again.’

On its own the minimum fuel trajectory descent initially proved of little practical value for air traffic management, especially with the computational capability then available. As Erzberger later recalled: ‘The idea was just way ahead of its time. It was basically a bridge too far. Looking back, the whole work I did in the 1970s was twenty years too early.’ These trajectory calculations were trigonometry based using a ‘kind of closed form solution’ to avoid the need for ‘a lot of iteration and computationally intensive work.’ This minimised the computational demands but meant that the trajectories could not cope with the complexity of most real-world air traffic conditions. According to Erzberger: ‘We soon found out that controllers can’t work that because they have to interrupt the flight several times because of other traffic, the whole problem is so much more complicated than I really understood at the beginning. So this was useless. The idea was still correct, just that the implementation was complicated.’

Although not practical for ATC when first conceived, this work on idle thrust descent algorithms found application in the 1980s through its incorporation into aircraft Flight Management System (FMS) technology, which automates many in-flight tasks, including calculating the flight plan. This meant that the FMS could enable an optimum fuel descent, although in practice this would rarely be possible because ‘they get interrupted by the air traffic control process. So coming into a busy airport the FMS doesn’t help them even though they have the capability on board to fly very well from top down to close to an idle descent.’

When Erzberger returned to ATC in the mid-1980s he felt that to make further progress he needed to test his algorithms with real, live air traffic data – in contrast to the FAA’s then main contractor for ATC research, the MITRE Corporation, that had never sought access to live data. Getting access to this data for the work at NASA Ames was not easy:

A huge struggle began around 1985, going to the FAA myself, my managers, division chief, even the director of Ames, going to meetings with the FAA to convince them to give us a data-link to real-time air traffic radar output. That was a struggle that you cannot believe.

Erzberger argued that recorded data was not adequate to understand fully the challenges of ATC; with live data he could not only watch what air traffic controllers were doing, but also phone them to get near real-time explanations for their actions. Amongst those convinced were two FAA Administrators, Admiral James B. Busey and General Thomas C. Richards who visited Ames and found Erzberger persuasive. The first live data became available to Ames in 1989, with the full complement of data available by 1992. Others were also persuaded. According to Erzberger ‘we built
up support from industry, lots of airline people, vice-presidents, United, American, and they were all totally enthused by it.9

By the late 1980s the NASA team were able to put forward a suite of tools to aid air traffic controllers by providing accurate predictions of arrival times, and calculating optimum flight trajectories. Earlier attempts at ATC automation had failed, it was argued, because of the ‘tendency of developers to underestimate the complexity of automating even simple ATC functions’ [17, p. 2]. Significantly, the NASA system did not seek the full automation that had been the goal of earlier schemes. Rather the focus was ‘human-centred automation’ aimed at ‘developing tools that complement the skills of controllers without restricting their freedom to manage traffic manually’ [17, p. 2].

The potential of this work stimulated sufficient interest at the FAA for it to establish a joint programme with NASA [11, p. 1]. Erzberger’s algorithms were meshed with databases ’consisting of several hundred aircraft performance models, airline preferred operational procedures and a three-dimensional wind model’ to enable trajectory projection [11, p. 1]. This meant it was possible to compute ‘aircraft trajectories approximately 40 min into the future, starting at the aircraft’s current position, then factoring in the aircraft’s intent as contained in its flight plan, along with a model of the performance capabilities of individual aircraft, including lift, drag, and thrust, and atmospheric conditions’ [55, p. 30]. According to Erzberger, ‘with new computer technologies and fast algorithms . . . now we can do all of that stuff . . . to be able to get this fuel optimum trajectory and to avoid traffic at the same time or make only very small deviations to avoid loss of separation against other traffic’.10

Previously this task has been handled through the ability of controllers to visualise evolving aircraft movements so as to seek the optimum sequencing of aircraft, and to provide appropriate directions for changes in direction, altitude and speed. The task was further complicated by ‘the necessity to compensate for strong variations in ground speed during the descent as a result of altitude-dependent winds and atmospheric effects’ [17, p. 5]. The NASA tools could assist controllers because they allowed the evolving situation to be predicted further into the future, enabling prediction up to around 25 min away rather than the 5–10 min possible manually [17, pp. 5 & 6]. The key tool was the Traffic Management Advisor (TMA) with its primary function ‘to plan the most efficient landing order and to assign optimally spaced landing times to all arrivals’ [17, p. 3]. Working with aircraft at between 150 and 200 miles from an airport, the TMA plans arrival times ‘such that traffic approaching from all directions will merge on the final approach without conflicts and with optimal spacing’ [17, p. 3]. The TMA software drives a screen that displays efficient aircraft arrival scheduling, enabling air traffic controllers (should they choose) to quickly advise pilots of necessary delays before they begin their descents.

TMA was introduced following protracted lobbying. Initially, the FAA was unconvinced by the need to adopt Erzberger’s approach, but the enthusiasm of the airlines proved crucial, as did the involvement of key Congressional committees. According to Erzberger: ‘They [the airline companies] came and talked to me and to others here and I demonstrated systems in the laboratory to them and then they put the pressure on the FAA to implement this and eventually the FAA very much embraced the concept’.11 The airline companies were ‘the key advocate at that time; they went to the Congress and asked the FAA to explain why this system should not be deployed. And that gave a lot of incentive for the FAA to ask for monies and they provided the monies to do that.’12

The success of TMA led to its widespread use, and it is now used not only for arrival traffic, but has also been extended to handle departures from nearby airports:

For example, in the Los Angeles traffic area you have this traffic from Vegas and from San Francisco and so TMA is able to determine the best take off time slots for traffic that is waiting to go so that they don’t have to go and circle and overcrowd the airports. So its made a huge difference over there . . . I was actually stunned by it, the amount of time and fuel savings.13

NASA’s TMA has improved US ATC efficiency in recent years, aiding air traffic controllers, reducing congestion, and providing fuel savings. However, progress in ATC automation falls far short of that predicted at the start of the 1980s. Further improvements are underway in the FAA’s NextGen programme, including deployment of Automatic Dependent Surveillance-Broadcast (ADS-B), a technology that can ‘update activity on air traffic controller displays more frequently and with greater accuracy, providing information such as aircraft type, call sign, heading, altitude and speed’ [20, p. 11]. The introduction of ADS-B will thus benefit ATC algorithms by enabling more accurate trajectory predictions.

6. Further US ATC modernisation

As a whole, though, NextGen plans are progressing slowly [37,28,61]. Some of the problems stem from the difficulty of funding and coordinating improvements across a large, complex socio-technical system, with a key example being the ground-air data link system known as ‘Data Comm’, a relatively straightforward technology that could be implemented quickly if there were sufficient will on the part of key actors. Other significant advances depend on increased automation, where barriers to progress hinge on fundamental concerns over safety (a key concern because of the potential loss of life and bad publicity that could result from aircraft collisions).

Improvements in data links between ground controllers and aircraft have long been considered desirable. In 1993 aviation commentator John Nance complained that: ‘Voice communications are archaic’ [41, p. 86]. For the purposes of ATC automation, voice communications are inadequate to transmit detailed flight plan information. ATC automation technology can predict conflicts and produce revised flight plans, but without a high bandwidth data link such as Data Comm, air traffic controllers have to transmit these revisions verbally. Direct electronic transmission would both eliminate misunderstandings, and be more efficient in not requiring the aircrew to translate verbal instructions into a new flight plan: ‘The value of the data link is that you can send a complex trajectory up directly into the aircraft FMS or autopilot. That would involve half a dozen numbers, such as coordinates of waypoints, speeds, etc. There’s no way a controller could issue those precisely by voice.’14

Air traffic controllers could thus work much more efficiently, though they would still be in control. It would not be a fully automated system: ‘The controller and the pilots are still in the loop, but there’s lots of numerical communications going on under the hood which humans need not know about, and then for the final decision

9 Erzberger 2008 Ferguson interview.
10 Erzberger 2010 interview.
11 Erzberger 2010 interview.
12 Erzberger 2010 interview and email from Heinz Erzberger, November 30, 2011.
13 Erzberger 2010 interview and email from Heinz Erzberger, November 30, 2011.
14 Erzberger 2010 interview, and 2011 email.
a controller can still be involved.\textsuperscript{15} The automation would replace routine activities, but not the air traffic controllers themselves: You would still have probably as many controllers as you have today but they would be more handling exceptions. . . . The routine and repetitive stuff would be done by the automation in a more precise way to ensure flights are safely separated. And then when there’s weather or emergency or the pilot asks for deviations for a variety of reasons, the controller still has to mediate that.\textsuperscript{16}

In addition, a data link would enable aircraft data to be downloaded to the ground systems to help fine-tune trajectory predictions. Erzberger’s original algorithms provided a prediction standard deviation of around plus or minus 75 s, and although this has improved, the error range needs to be brought down to plus or minus 20 s in order to enable the most energy-efficient continuous descents.\textsuperscript{17} A data link to download aircraft data would enable this. However, despite these advantages, data link implementation has been delayed. As Perry [\textsuperscript{55}, p. 32] noted over a decade ago: ‘While nearly everyone agrees data link is a good idea, implementation is not as obvious’. From NASA’s perspective, the problem is that ‘the value of this data link is not fully understood by both FAA and airlines.’\textsuperscript{18} With a wide range of actors involved, there is ‘a situation where the decisions are made by many different people with totally different objectives . . . you have lots of players and they all have their own function to optimise which is not necessarily the best for the optimisation of the system.’\textsuperscript{19}

The data link technology currently in operation is the Aircraft Communications Addressing and Reporting System (ACARS) that was introduced in the late 1970s to replace radio as the main means that airline companies could use to communicate with their aircraft. A wide variety of data can be transmitted through ACARS including data on aircraft and engine performance, but although ACARS is used for air traffic control, the limited data size precludes the transmission of complex flight plan changes.

ACARS had competitive benefits to airline companies because it provided a mechanism for rapid reporting of their aircraft data, ‘but when it comes to data link for air traffic control purposes, that’s not a direct business case for them. They use their data links for competitive advantages very directly, but data link used by air traffic control doesn’t have the same parochial value to them.’\textsuperscript{20} Although more advanced ATC could offer benefits in enabling fewer delays and fuel savings, it would do so equally to all airlines. Moreover, Data Comm would not provide much benefit to early adopters because they would have to wait until adoption was widespread before its ATC use became standard. There is thus what has been termed a ‘NextGen Equipage Paradox’ in which ‘those operators who are last to equip with NextGen avionics gain the greatest financial benefit, while those operators first to adopt the new technologies will pay a much higher price at far greater risk’ (\textsuperscript{31}, p. 3; see also \textsuperscript{9}).

While it appears that a suitable data link could be deployed relatively easily were sufficient funding and commitment available, implementing greater automation faces more fundamental obstacles. As well as enabling greater fuel-efficiency, the goal of such an approach would be ‘to achieve a significant increase in capacity and throughput while providing even higher levels of safety than today’s system’ [\textsuperscript{18}, p. 355]. The challenge, however, is making a convincing case that safety would not be compromised, never mind improved.

Automation could completely eliminate the current reliance on air routes, facilitating fuel-efficient ‘free flight’ whilst ensuring adequate separation, but there are two safety concerns. First, the system must be implemented in such a way that the failure of any key component would not endanger aircraft. Thus Erzberger and Paielli [\textsuperscript{18}, p. 357] argue that:

> The most important technical and operational challenge . . . lies in providing a safety net to ensure the safety of operations in the event of failures of primary system components such as computers, software, and data-link systems. This includes defining procedures for reverting to safe, though less efficient, back-up systems.

In addition, there is a second safety issue that appears more fundamental, and more difficult to address, because it is centred on concerns about the extent to which software can encompass all possible events, and to what degree safety can be demonstrated in large, complex programs. As Erzberger and Paielli [\textsuperscript{18}, p. 361] note:

> Automation software . . . is inherently limited to the solution of problems that fall within the operational envelope determined by the finite parameterization of solutions built into the software. Unfortunately, for complex software comprising several hundred thousand lines of code, the controllable problem set cannot be determined, because of the extremely high dimensionality of the input conditions that would have to be evaluated. Therefore, the boundary between the set of solvable and unsolvable problems is unknowable. Although the envelope of problems that controllers can solve is also limited, it is much larger . . . Moreover, human controllers excel at adapting their control strategies to completely new situations, a capability that is beyond existing software design.

The software of an automated system may perform well under expected conditions, but ‘unplanned and unpredictable events, such as equipment failure or severe weather conditions, may produce conditions that fall outside that envelope’ with the result that ‘traffic flow could become inefficient and chaotic, risking the loss of separation’ [\textsuperscript{18}, p. 361].

Alongside this concern over the ability of the software to cope with unexpected conditions, there is general concern applicable to large, complex software programs. Such software is difficult to verify because it cannot realistically (in terms of time and money) be run in all possible states, and so it cannot be proven empirically. Although mathematical techniques can be used to carry out ‘formal verification’ of software, providing deductive proof that the software satisfies its specification, such techniques are currently feasible only for software of modest size and complexity, and provide no defence against specifications that are inadequate or wrong [\textsuperscript{47}]. In the case of advanced ATC automation, ‘the complexity of the algorithms embedded in the software presents another obstacle to the system passing a certification test. Establishing the robustness and operational envelope of the algorithms and even documenting the design will be difficult’ [\textsuperscript{18}, pp. 361 & 362]. These challenges have led to uncertainty as to the degree of automation that will be attempted in NextGen.

15 Erzberger 2010 interview, and 2011 email.
16 Erzberger 2010 interview, and 2011 email.
17 Interview with Heinz Erzberger, April 9, 2012.
18 Erzberger 2010 interview.
19 Interview with Banavar Sridhar, 8 November 2010.
20 Erzberger 2010 interview.
7. Lock-in and ATC

This condensed history of US ATC developments shows slow progress in modernisation. The long predicted automation has not happened, and even apparently modest technologies such as Data Comm have not been implemented. Can this lack of progress be explained as lock-in due to increasing returns and network externalities? Or do other factors limit progress in such complex infrastructure systems?

Network externalities clearly matter for initiatives such as Data Comm because there may not be an immediate return on investment made by an individual airline until most other airlines have also made similar investments. A NASA interviewee noted that: ‘We have to make a lot of infrastructure improvements and its not always possible to have benefits immediately unless everybody has adopted it. And that’s one of the biggest problems.’ Similarly, even though ADS-B infrastructure is being deployed, users have been sceptical about equipping their aircraft, ‘fearing early adopters will have to upgrade equipment one or more times before ADS-B is fully operational’[69, p. 38]. The highly networked nature of ATC inhibits uncoordinated change, with the FAA having limited authority to compel such coordination.

In addition, there is also lock-in due to increasing returns because of the learning effects that imbue confidence in existing technology and practices, and militate against changes that might undermine safety. The essence of this lock-in lies not in any particular technological artefact or even network of artefacts (although many are involved), but rather in judgments about what is sufficiently safe given the accumulated knowledge and expertise of those involved. The effectiveness of the ATC system, particularly as regard safety, is widely regarded as hinging on ‘human factors’. Every change, and particularly every move towards increased automation, must thus be considered with regard to effects on ATC users and operators, who are central to handling any potential events because:

Humans are thought to be more flexible, adaptable, and creative than automation and thus better able to respond to changing or unpredictable circumstances. Given that no automation technology (or its human designer) can foresee all possibilities in a complex environment, the human operator’s experience and judgment will be needed to cope with such conditions [71, pp. 241 & 242].

Objections to increased automation thus stem from concern about changing a system that works. The current ATC system (with controllers managing aircraft along predefined airways) is essentially the same as established over fifty years ago. Recent years have seen improvements such as the NASA-developed TMA system, but full automation would be a radical departure requiring ‘a change in culture, and there’s always fear of the unknown.’ In particular, the challenge of certifying safety for complex automated ATC software is considered a fundamental barrier. Existing ATC technology is known to work safely, it is argued, whereas certification of fully automated ATC software is not possible, given the difficulty of quantifying the risk involved.

However, the challenges to ATC modernisation go beyond overcoming lock-in due to network externalities and increasing returns. Timely and radical change is also difficult because of the broader political and organisational factors that trouble many infrastructure projects. It has long been argued that ‘bounded rationality’ within organisations leads to routinised behaviour and a tendency to ‘satisfice’ that favours incremental over radical solutions, and others have noted what has been termed ‘institutional lock-in’[22,56] or entrapment [68] in which organisational relationships limit rapid change. The US National Airspace System is not just large and technically complex, it is also socially and organisationally complex, making change difficult to achieve in ‘a situation where the decisions are made by many different people with totally different objectives.’ A potential obstacle to further automation could be union concerns that this would involve job losses. While NASA has stressed that automation is not a threat to the jobs of air traffic controllers, the European ‘Single European Sky’ automation initiative has led to strikes centred on this concern.

Moreover, the key US government actor in this process – the FAA – is limited in its ability to direct developments in areas other than safety:

FAA has the overall control from the point of view of certifying aircraft and providing the infrastructure for air traffic management. That’s where their responsibility ends. … they can ask airlines to put equipment in if it is necessary for safety, but they don’t have any power to say, if everybody behaved this way the system will run very efficiently. They have no way of enforcing that.

In addition, there are specific political problems in the USA in an era of fiscal conservatism. The FAA’s funding must be voted through Congress, and is prone to the political wrangling that typifies the annual appropriations process. Funding for ATC modernisation has thus only been approved ‘in dribs and drabs’ which ‘makes realistic long-term planning for air traffic control modernisation nearly impossible’[57]. For example, in July 2011 the FAA instructed many of its contractors, including some working on NextGen, to stop work due to the failure of Congress to pass a FAA bill [21].

8. Policy implications: overcoming lock-in

This analysis indicates that overcoming ATC lock-in will be difficult because of the many factors involved. Network externalities and increasing returns effects are exacerbated by the political and organisational limits to driving radical change in a large infrastructure project comprising a complex socio-technical system. Gil et al. [32, p. 453] conclude that infrastructure decisions involving many partners tend to ‘freeze adoption decisions on proven technologies’ so as to reduce risk, and it is therefore important, as van der Vooren et al. [67, p. 115] recommend, that ‘policymakers should allocate substantial financial resources to the public support for infrastructure development’.

The risk-averse attitudes that favour proven technologies are a particularly significant barrier to ATC automation because of safety concerns that hinge on the perceived challenge of certifying the reliability of complex software. However, as Downer [13–15] convincingly argues, current civil aviation technology (his work deals with aircraft and not ATC) is certified even though its reliability cannot be readily quantified. The FAA’s approach to safety certification requires that aviation technologies and procedures should meet specified performance requirements, but this relies on judgement rather than objective measurement alone.

Measuring the performance of new technologies depends on testing, but this can differ from operational use in significant ways. Aero-engines, for example, cannot be tested to destruction in large

---

21 Sridhar interview.
22 Erzberger 2010 interview.
23 Sridhar interview.
25 Sridhar interview.
numbers because of the expense of so doing. The data that is collected from engine tests with regard to their ability to withstand bird strikes is therefore necessarily limited, and contingent on the particular way that the tests are designed and carried out [13]. There is thus an unavoidable element of judgement involved in aircraft certification [27], but all involved – regulators, manufacturers and airline companies – have a strong, shared interest in a conservative approach that prioritises high safety standards.

Downer’s work points to how ATC lock-in due to increasing returns can be overcome. Complex automation software can never be fully tested, of course, but it can be tested extensively without safety risks because it can be run in parallel with current ATC operations. Thanks to Erzberger’s initiative in the 1980s, NASA Ames has such a capability:

On the other side of this hallway is our laboratory and we have the live data, live radar data coming in from all the centres and some TRACON’s as well. We can drive our system with this and see what the decisions are and watch it. Of course they have their benefits for operation, but we can monitor that we can shadow that. And when we have better algorithms, we can see the difference, we can check out new software and systems.26

This means that the developers of new ATC software have the ability to carry out real-time simulation, to use live traffic data, and to interact with air traffic controllers and pilots in this process. This facilitates a highly iterative development process in which ‘the requirements, design, simulation, and operational tests are conducted concurrently with a high level of interaction’ and new developments can be introduced gradually with an approach of ‘design a little, test a little’ [11, p. 1].

Thus, although in principle the risks involved with ATC software cannot be quantified, in practice this software can be tested extensively. As one of the computing’s pioneering figures, J.C.R. Licklider, put it in 1969, ‘even quite complex software subsystems can be “mastered” (which is not the same as “perfected”) and made to provide useful and effective service if they can be developed progressively, with the aid of extensive testing of systems (as well as subsystems and components), and if they can be operated more or less continually in a somewhat lenient and forgiving environment’ [45, pp. 126 & 127]. ATC software meets Licklider’s requirements in this regard. By shadowing live air traffic operations, new ATM software can be operated before being introduced into use. And when introduced, more than one software system can be used to provide redundancy for separation assurance [19].

Finally, it should be noted that although the potential fuel savings with more efficient ATC for any particular flight may be of the order of 5–10%, overall environmental benefits cannot be guaranteed to accrue due to the ‘jevons paradox’. Lower fuel costs could mean cheaper flights, thus increasing demand and leaving total climate change emissions no lower (see [34]). While better ATC can have other benefits for climate change because of the potential to reduce cloud formation by rerouting aircraft, gaining the full benefits of fuel efficiency may require other policy measures to limit customer demand.

9. Conclusions

This case study of US ATC goes beyond most previous studies of lock-in by providing a detailed account of the mechanisms that inhibit change. This analysis shows that the barriers to US ATC modernisation feature the classic lock-in characteristics of increasing returns and network externalities. In particular, the centrality of safety to the functionality of ATC means that new approaches – notably automation – are judged against the learning effects of increasing returns perceived to have accrued in the traditional manual approach to aircraft control. Overcoming this type of lock-in requires particular attention to building trust in automation (through the provision of redundancy, piecemeal introduction in semi-automated operation, and laboratory testing of a shadow system).

This analysis suggests that classic lock-in theory has two limitations, at least for this type of technology. First, whereas traditional lock-in theory posits economic obstacles to change, this analysis of ATC lock-in suggests that barriers to change can involve perceptions of risk that do not comprise readily calculable costs or benefits. Instead, the more sociological analysis done here indicates that apparently quantifiable entities such as reliability are not only constructed by the actors, but also filtered through organisational lenses. Just as trust in aircraft performance depends crucially on the relationship of the FAA to aviation engineers rather than on estimates of reliability [14], so too trust in ATC performance cannot be generated solely through calculation. Similar concerns are likely to arise in other software-intensive technologies where risk minimisation is central (for example, in nuclear power operation).

Second, over-emphasis on lock-in as an explanation can mean that more mundane, but equally important, factors are ignored. In the case of infrastructure, large amounts of investment need to be obtained and coordinated. ATC technology may be an extreme example, but many infrastructure projects face the challenge of driving change across a complex network of heterogeneous organisations and interests. In the case of ATC, achieving further advances in automation will require more than just good technology, it will also require ‘heterogeneous engineering’ [44], combining technical developments with efforts to overcome significant political and organisational barriers (just as Erzberger and his NASA colleagues had to do to lobby an initially reluctant FAA to adopt the TMA tool).

Understanding and promoting environmental transitions is one of the most important challenges we face [1]. In recent years the conceptual orthodoxy requires that such transitions be understood in terms of the Multi-Level Perspective (MLP) approach (e.g. [34,29]). In contrast, the other side of the coin, so to speak, has been neglected, with relatively few recent studies of the lock-in of existing sociotechnical systems. However, understanding why transitions do not occur is important to provide some methodological ‘symmetry’ to MLP case studies that overwhelmingly describe successful transitions. As leading MLP proponents [30, p. 79] acknowledge, there is a need to ‘correct the bias towards winners and novelty’ in our understanding of transitions. In some cases (e.g. [12]) ‘de-alignment’ of existing regimes appears to happen relatively easily, whereas in others the ‘path dependence’ effects of lock-in are stronger. Detailed historical case studies [35] can help us understand why some regimes are more locked-in and others more easily de-aligned. The implications of this study of ATC are that lock-in effects can constitute a significant hurdle to even modest transitions, and that detailed study of specific cases are important because targeted rather than generic solutions may be needed.

Acknowledgements

This research was supported by the UK Economic and Social Research Council (Grant no. RES-062-23-2379). I am grateful to those at National Aeronautics and Space Administration (NASA) who helped with this research, particularly Heinz Erzberger, Banavar Sridhar and Glenn Bugos. I am also grateful to Donald MacKenzie and Rebecca Slayton for their helpful comments.

---

26 Erzberger 2010 interview, and 2011 email.