Neither Pollyanna nor Chicken Little
Thoughts on the Ethics of Automation

C. Michael Holloway
NASA Langley Research Center
Hampton, VA, USA

John C. Knight
University of Virginia
Charlottesville, VA, USA

John A. McDermid
University of York
York, UK

Abstract—The terms Pollyanna and Chicken Little are used as caricatures for those who are relentlessly positive and those who, in contrast, see problems in everything. Many opinions about automation in aviation tend to reflect these contrasting viewpoints. In this paper we propose the introduction of an ethical safety case as a means to reconcile these views by clearly articulating ethical issues involved in automation.

Keywords—ethics, automation, argument, safety case, aviation

I. INTRODUCTION

The use of automation is growing in many classes of system that can affect the public including medical devices, transportation systems, power generation, and defense systems. There are many ethical issues with all of these types of systems, and considering automation in general terms is not possible. In this paper, we focus on automation in aviation systems. We investigate in some depth general issues of automation in current aircraft and particular ethical issues arising from the growing call to introduce unmanned air systems (UAS) (which are necessarily highly automated) into controlled air space.

The notion of “autonomy” is neither absolute nor new in aviation. Modern aircraft engines are controlled by Full Authority Digital Engine Controllers (FADECs); the earliest of which were introduced around 1980. The introduction of FADECs enabled airlines to save money by removing the flight engineer from the flight deck. Here “full authority” means that the control system directly controls all the key engine parameters (fuel flow, for example); the pilot can request thrust but cannot “bypass” the FADEC to directly control many of the mechanical elements of the engine.

A remotely piloted air system (RPAS) is more autonomous than a FADEC. With an RPAS, a remote pilot will make “strategic” decisions, such as changing a route, but “lower level” automation can be used for many vehicle systems, not just the engine, e.g. ice detection and protection. A UAS is more autonomous still, able to choose routes and deal with some failure scenarios, perhaps even deciding to crash land or self-destruct if it cannot continue safe flight and landing.

These differences raise an obvious ethical question: “What level of automation is appropriate in aviation?” But this question is ill formed. Instead we should ask: “How do we design systems so that the level of automation is appropriate to the situation?”. This question leads us into the realm of engineering ethics, but first we make a brief terminological observation.

We have already used three terms that are related if not cognate: autonomy, authority and automation. The terms “authority” and “autonomy” both have human connotations and are perhaps the most obvious basis for an ethical debate. Specifically, “how much authority shall we give to the system?” and thus implicitly remove from the human pilot. In our view, the use of morally loaded terms does not really help the discussion and we choose to use the more neutral term automation. This approach has the advantage of allowing clear separation of the engineering concerns of automation from the ethical concerns of “how much automation is appropriate.” Thus we use the term “automation” for the rest of this paper, except where we wish to make specific ethical points.

Recently, in the UK the Royal Academy of Engineering (RAEng) has undertaken an extensive study of engineering ethics, and has articulated four principles [1, 2]:

1. Accuracy and rigour, including ensuring that others are not misled;
2. Honesty and integrity, including preventing corruption;
3. Respect for life, law and public good, including risks to public safety;
4. Responsible leadership, including promoting public awareness of issues, for example relating to risk.

These principles might be paraphrased by saying that there exists an obligation to “do no harm” (protect the public). Each has a bearing on automation in aviation systems, and we return to them later in the paper.

In the rest of this paper, we present a summary of levels of automation in current (“modern”) aircraft in section II to provide a reference, and then discuss trends in aviation autonomy in section III. We review different attitudes towards risk, which we refer to as “worldviews”, in section IV, reflecting the stereotypes of the title. Then, in section V, we set out the ethical issues which are raised before considering a way of reconciling these views, via an approach that we refer to as “ethical safety cases”, in section VI. Finally we present our conclusions and make suggestions for further development in section VII.

---

1 The specific ethical issues that arise with autonomous (the use of the term is deliberate) military aviation systems are outside the scope of this paper.
II. ISSUES WITH CURRENT AVIATION SYSTEMS

Modern aircraft are highly automated. For example, many aircraft are fitted with autopilots that can fly the aircraft along designated routes then land the aircraft at a (suitably equipped) airport. It is often said that modern pilots “monitor and supervise the automation” rather than “fly the plane”; see, for example, a cockpit video of an A380 approach to San Francisco (SFO) [3]. Note that this is a difficult approach, referred to by pilots as a “slam dunk” due to the steepness of the approach, and one where a Korean 777 crashed with a comparatively experienced pilot flying (manually) into the airport for the first time [4].

This contrast tends to suggest that “automation is good”, and some people take this view (see below). Of course the situation is much more subtle than such a simplistic view, and we now set out some of the issues in modern aviation which relate to the degree of automation and the contextual issues of how airlines operate, and how pilots are trained.

First, automated can mean “unpractised”. Even if pilots do one landing in ten manually, due to the rosters under which they operate, even experienced pilots can find themselves carrying out activities, such as landing at a tricky airport, for the first time under difficult weather conditions. For example, the pilot of the Korean aircraft coming into SFO had 10,000 hours flying experience but only 43 on the 777 and this was the first time he had ever flown manually into SFO. (Of course, it is impossible to avoid “first times” and the training process is meant to ensure that the previous experience means that pilots can successfully cope with new airports and situations.)

Second, much of a pilot’s training is now done in simulators. Whilst the simulators are good and the training includes dealing with emergencies, there is much less experience of “seat of the pants” flying. It should be noted that Captain Sullenberger who famously landed his aircraft on the Hudson after total engine failure [5, 6] is “old school” and an ex-military pilot. His experience and decision-making almost certainly were the critical factors in the successful handling of the emergency; one might also speculate that he chose the river also on the basis that a (failed) landing on terra firma was likely to put many third parties on the ground at risk (an ethical decision, of course).

Third, there is the now widely recognised phenomenon of “automation surprise”. The automated systems are very complex, and often highly moded. This means, for example, that one button on a control column may do one thing at one time, and something different at another. On Airbus aircraft pushing the stick forward disengages the autopilot except in go around mode; this exception is said to have been a causal factor in the crash at Nagoya [7]. Thus a pilot can be surprised at what happens in response to his commands and some of these surprises are well documented, but may not be rectified by the manufacturers [8, 9, 10].

Fourth, pilots often get back control from the automation when something bad has already happened and the computers can no longer cope. For example, the A380 incident near Singapore [11] which resulted from an uncontained engine failure led to thousands of alarms going off, which was successfully handled by five crew in the cockpit. On long-haul flights two complete flight crews are carried, and there was a fifth (training) pilot on board.

Perhaps the key ethical concern underlying these issues is one of eroding professional skills, and leaving pilots with responsibilities that they are not fully equipped to discharge. In terms of the principles articulated by the RAEng, principle 1 (accuracy and rigour) is an issue for pilot training, including the fidelity of the simulators. Principles 3 and 4 are particularly pertinent to the system design, and the willingness (or otherwise) of aircraft manufacturers to rectify known problems, including automation surprises. In this context, the effective design of the automation – the balance between pilot and technology – is the critical issue.

Airworthiness authorities have reported on the underlying concerns for almost two decades [8, 9, 10]. The topic is complex, but the following [8] (p. 35) serves to illustrate the issues: “The unexpected pilot behavior evidenced in recent accidents and incidents appears to be the result of many factors, including the increased capability, reliability, and authority of the automated systems, increased flight crew use of and reliance on such systems, protective features of these systems (real or imagined), automation philosophy (or lack thereof) of the operator, and cultural differences. An additional factor may be that flight crews are becoming less confident in their own airmanship skills relative to the capabilities they perceive to be present in the automation, particularly in a stressful situation.”

As a final observation, pilots are often “blamed” for accidents. But how many of these accidents involved pilots dealing with very complex situations that were exacerbated, not aided, by the automation? (Fortunately the number of cases where pilots were a clear cause of an accident, such as [12], are very rare.) There are many cases, such as the Sullenberger and A380 examples above, where pilots deal with very difficult circumstances; it is not obvious that these situations are adequately recorded and factored into design-time decision-making. Thus it is perhaps no surprise that there is a trend towards greater automation despite the warnings from the authorities over almost two decades.

III. ISSUES WITH TRENDS IN AUTONOMY

The trend towards autonomy can perhaps be characterised (if it is not too much of a caricature) as “if people are the problem, take them out of the loop”. This is clearly the Google car motivation [13], although, interestingly, having got the car on the road Google is now doing a classical safety analysis for the vehicle!

This pressure is apparent in many aviation sectors, even without considering the military. There are already some UAS cleared for use in restricted airspace, such as the ScanEagle, which is certified for use in a part of Alaska [14]. The FAA has been instructed by Congress to develop rules to allow commercial UAS in controlled airspace by September 2015 [15]. In the UK an RPAS has flown in controlled airspace with a pilot in the cockpit for take-off, landings, and emergencies [16]. Amazon has said it wants to operate UAS [17], and Deutsche Post in Germany has done a limited operation with
an RPAS [18]. In the latter two examples the motivation is economic; like the cost-savings that came from removing flight engineers from cockpits, Amazon and other businesses see potential cost-savings from using automation.

From an engineering and ethical perspective we view RPAS and UAS as very similar. An RPAS may lose its communication links with the remote pilot; if it does, then the reliance on the automation is exactly the same as with a UAS. The details may be a little different, but for the purpose of this paper we can assume that the core problems are largely the same.

Humans are naturally adaptive; computers are not. Thus the difficulty is either:

- To predict and program for every situation that might be encountered;
- To make the computer system adaptive.

The first case we shall refer to as the “closed world” assumption; here we generally do not know that the system will behave appropriately (be safe) if a situation is encountered which had not been anticipated (it is outside the computer’s model of the world). This obviously relates to principle 1 (accuracy and rigour), but it is very strongly influenced by the other three principles, which relate to what we should say and claim about our ability to achieve the necessary integrity and rigour.

Building adaptive systems involves simulating mental capacities in decision-making, situation assessment (including image analysis), and possibly speech recognition and synthesis. All of these sub-problems are really hard technically; solving them all together and producing a system that can mimic human behaviour is extraordinarily difficult. Yet, because of the perception of human fallibility, there is a desire to do this, and a belief that it can be done – and done safely and ethically – which leads us to a discussion of worldviews.

IV. RISK AND WORLDVIEWS

It is possible to build highly automated systems (modern aircraft illustrate this); it is possible to build adaptive systems (see for example [19]), and to address some of the other sub-problems such as illustrated by IBM’s Watson playing Jeopardy [20]. Because of these (in some cases hugely impressive) successes and (perhaps) a belief in technological progress and prowess, there are those that believe automation is the answer; these we have characterised as the Pollyannas who are relentlessly positive about what can be achieved.

On the other hand, there are those who point out (as we have hinted) the enormous technical challenges of either building systems on the closed-world assumption with a good enough world model that we can trust them with human life, or making systems which are adaptive, but where we are confident that the adaptations will be desirable. We have characterised this group as the Chicken Littles, who can see the problems in everything (but rarely the benefits). The Chicken Littles probably also think that the Pollyannas are “risk blind”; conversely, the Pollyannas probably think that the Chicken Littles are excessively “risk averse”.

Both these worldviews have merits. The point of this paper is not to label either as necessarily right or wrong, but rather to consider how to reconcile them. We set out our proposed approach below, but expand on the notion of risk a bit before doing so.

In safety engineering, the use of formalised, quantitative risk assessment (QRA) is common in some domains. The QRA models focus on aleatory uncertainty, which is more commonly referred to as random. A QRA represents known failure models such as wear out usually based on underlying models of physics of failure. Whilst this approach is well established, there is evidence that the results are not accurate numerically [21], and that many real-world QRA results are flawed. One of the reasons for the flaws is that the models are limited by epistemic uncertainty. Put simply, the QRA may leave out significant factors, but we do not know what they are.

The risk blind Pollyannas do not see the epistemic uncertainties and assume that the QRA (and other analyses) are effective and appropriate. The Chicken Littles know that uncertainties must exist; whilst it cannot be said that they exaggerate them, it is perhaps fair to say they are swayed more than perhaps they should be by the lack of, or limitations in, knowledge. This is a point on which we wish to focus the ethical spotlight – but there is one further issue to consider before we do so.

The key capabilities of UAS and RPAS are realised in software. Software doesn’t have a physics of failure. It is logic: it does what it does. If a system containing software fails (other than as a result of physical failures of the hardware) it fails for epistemic reasons. This might be because of an inadequate model of the real world, such as the designers forgetting about the existence of the International Date Line [22]. It might be because of an error in the development of the software leaving a bug that can cause an incident or accident at some time, such as [23] where pilots eventually got control back despite some bugs that caused uncommanded altitude excursions of around 16,000 feet.

Software in modern aircraft (and thus potentially UAS) amounts to many 10s of millions of lines of code (MLoC). Whilst it is hard to get accurate data there is evidence to suggest that good safety critical code contains about one bug per thousand LoC (kLoc) and the best is probably about 1 per 10 kLoC [24]. Thus a modern aircraft will contain thousands of bugs in its software – this might seem to support the Chicken Little view, but what little evidence there is shows that the dangerous failure rate of avionics software is remarkably low [25], which supports the Pollyanna view.

So why does this situation pose ethical issues, not just technical ones?
V. ETHICAL ISSUES

Whilst it might be thought that the most extreme ethical issues are brought about by UAS, there are also serious issues with regard to levels of automation in conventional aircraft. As the issues are rather different, we highlight some of the issues in the two classes and consider RPAS which, in some respects, combine the challenges of both!

We note that the airworthiness authorities (the FAA in the United States and EASA in Europe) have important roles in that they act as surrogates for the flying public (and others affected by aviation) in making decisions about allowing aircraft to operate. Thus there are four entities we need to consider: pilots, airworthiness authorities, operators, and designers / manufacturers.

A. Automation in Piloted Civil Aircraft

We consider ethical questions in terms of the different entities identified above; this is intended to be illustrative, not exhaustive:

- Pilots – under what conditions, related to the automation of the aircraft systems, is it ethical for pilots to accept responsibility for the lives of the aircraft’s occupants (and people on the ground)?
- Operators – under what conditions, related to training and experience, is it ethical for operators to place pilots in situations where they are responsible for the lives of the aircraft’s occupants (and people on the ground)?
- Designers – under what conditions, related to the intrinsic uncertainties in developing complex automated systems, is it ethical to design aircraft so that the pilots may neither have the information nor the authority to control (aspects of) the aircraft?
- Authorities – under what conditions, related to the automation of the aircraft systems and the intrinsic uncertainties in developing complex automated systems, is it ethical for the authorities to approve aircraft types (designs) on behalf of those directly exposed to the risks?

These questions are, of course, subtle and complex. Some of the context for these questions has been set out in the previous sections. In a sense the “Pollyanna vs Chicken Little” dichotomy (we abbreviate this PvCL from here on) is most sharply focused for the authorities who have to act for the community, despite the fact it holds conflicting views.

B. UAS

For UAS the pilot view is no longer relevant; for the other legal individuals we consider the issues related to the closed world assumption versus adaptation issues, as these are distinct, additional, concerns for UAS:

- Operators – under what conditions, related to the operating environment, is it ethical for operators to deploy UAS where they can endanger the lives of people on the ground?
- Designers – under what conditions, related to the operational assumptions and adaptive capabilities of complex automated systems, is it ethical to design UAS without the opportunity for remote operation?
- Authorities – under what conditions, related to the operating environment and the operational assumptions and adaptive capabilities of complex automated systems, is it ethical for the authorities to approve aircraft types (designs) on behalf of those directly exposed to the risks?

As set out here the operating environment should be taken to include weather, routes, cargo, and the like. For now it is assumed that the UAS are completely unmanned; of course similar questions could be posed for pilotless passenger carrying systems.

C. RPAS

RPAS have characteristics of highly automated piloted aircraft, except that they are unlikely to carry passengers. They also have some of the characteristics of UAS as they need to operate autonomously if the communication links to the remote pilot are lost. Thus RPAS can be viewed as an amalgam of the conventional, but highly automated, case and UAS, with the added issue of when/how the RPAS “decides” it has become isolated and should operate autonomously. There are also changes in the notion of “pilot control” which again have an ethical dimension – the pilot is reduced to what the system “tells” him or her; there is no “feel” for the aircraft, no view out of the window etc., so the remote pilot is much more dependent on the automation which changes the emphasis in the above questions.

D. General Issues

There are a number of other issues identified above which we return to here as ethical issues. They are posed in general terms, in the expectation that many of them have to be dealt with either by the specialist community or, more generally, by society and/or governments.

- What weight should be placed on QRA and calculations of risk, given the growing evidence that the analyses are often flawed, but the systems have been remarkably safe, despite these limitations?
- What weight should be placed on the intrinsic uncertainties related to software, especially as the software gains more authority (used deliberately here) over the operation of the systems?
- What weight should be placed on potential security vulnerabilities, which might mean that third parties could gain unauthorised control over RPAS (and possibly other forms of aircraft)?

There are also legal issues related to design and operation of UAS (and RPAS) but they are outside the scope of this paper.

VI. ETHICAL SAFETY CASES?

The notion of a safety case originated in Europe, particularly the UK, as a “structured argument, supported by evidence, which provides a comprehensive and compelling case that a system is safe to operate, in a given scenario”.

Whilst the concept is quite general, the practice (admittedly more outside aviation than within) is to base the argument on philosophical principles, such as Toulmin’s analysis of natural language “argumentation” [26], and encoded into special-purpose notations, such as the Goal Structuring Notation (GSN) [27, 28] and Claims Argument Evidence (CAE) [29].

Safety cases are hard to construct and may be developed or used in a way that is inappropriate or misleading. Whilst not expressed in these terms, the review of the Nimrod safety case [30] shows that it violated the four principles set out by the RAEng. Others have set out a more broad brush indictment of safety cases [31], but the basic principle of articulating the reasons why a system is believed to be safe, and also identifying the attendant risks, seems to be sound both in engineering and ethical terms. We take this as a premise in our proposed way forward which we term “Ethical Safety Cases” or ESC for short. We do not believe that an ESC can resolve the differences in views between the Pollyannas and Chicken Littles, but we do believe it can expose the issues so that ethical decisions can be made that take into account and try to balance the different views appropriately.

We focus on two separate aspects of safety cases, which we believe are steps towards producing ESCs: the acceptance process and the articulation of the arguments.

A. Acceptance

Currently, safety documentation (although not formally a safety case) is communicated between the manufacturer and the authorities and subject to review before acceptance (or otherwise); in practice there is interaction during the development process, but acceptance occurs at the end. We propose some changes to move towards ESC:

- Stakeholders – all the stakeholders discussed above should be represented and have an input to decisions:
  - Pilots – perhaps through trade associations;
  - Operators (airlines) – perhaps represented by the lead customer;
  - Manufacturer – perhaps supported by the designers of the most critical systems;
  - Authority – as now, acting as a surrogate for the (flying) public.

- Formal review and challenge including production of counter-arguments. Legal arguments include a case for and a case against a charge; this is not suggesting such an extensive counterargument, but the selective construction of challenges should help see how robust the argument is for novel aspects of the design (this directly addresses the PvCL issue).

- Formally agreed acceptance criteria to help focus the acceptance activity (see the discussion of articulation below).

- Monitor and feedback: collection of operational data to assess the safety of the aircraft, as predicted in the safety case to proactively identify safety issues.

With regard to the last point, at present accidents and incidents lead to review of the type approval and, if appropriate, the authorities impose mandatory design changes. What is proposed here is an extension to try to use data to prevent or avoid incidents and accidents; there are complexities here that space does not allow us to treat.

B. Articulation

At present there is little guidance on the production of good safety cases. Implicitly it is assumed that a design is good if the risks to life are adequately managed in the system design; this now seems limited and a richer approach is needed. The following are suggested as approaches particularly for novel aspects of design such as automation.

- Set out both the advantages and disadvantages of the design: Identify why the design gives advantage over a “standard” approach, as well as saying what the risks are (this directly addresses the PvCL issue).

- Define criteria (also used for acceptance) that force articulation of arguments beyond risk control, addressing the PvCL dichotomy, such as the following:
  - Where a function is completely automated, the result is “at least as good as” a human – this position is already taken by the UK Civil Aviation Authority (CAA) and it is challenging technically, but seems to be “correct” ethically – if we are removing authority from humans (moving responsibility from operators to designers) then the result should be no worse (ideally better).
  - Where a function is partially automated, the pilot is left in a situation where “authority and knowledge match responsibilities” and the result is “at least as good” as “manual” flying (as mentioned earlier, no aircraft is really flown “manually” these days, so this has to be a comparison with current design practices).

- Explicitly set out the assumptions and boundaries of modelling and analysis used to underpin the safety case, which enables challenges regarding the uncertainty in the design (again implicitly a PvCL issue) as noted in the process discussion above.

- Explicitly set out the resilience of the design, perhaps adopting the notions of anti-fragility. This is intended to be an alternative or addition to the use of risk-based analysis, explicitly dealing with the limits to knowledge (uncertainty) and complements the definitions of assumptions and boundaries (again implicitly a PvCL issue).

The last point is relatively radical, and would need work to flesh it out. The ideas of resilience have been receiving some attention from people traditionally working in safety engineering [32]. The notion of “anti-fragility” stems from Taleb’s work [33]; it is adopted as a means of articulating the ability of humans to gain from disorder and it is perhaps the ultimate criterion to assess the adaptability of systems as giving human levels of capability. Also, recognising the short
timeframes on which aviation systems change – as compared to the much longer timeframes for human development – might make us more cautious about levels of automation (the PvCL issue again).

VII. CONCLUSIONS

This paper has raised issues concerning the ethics of automation in aviation systems, and outlined ways of thinking about the issues that may help in ethical decision making. It is very easy to be carried along by technology and the Pollyanna view, but just because we can do something, doesn’t mean we should – which is perhaps a little milder than the Chicken Little view. Both views have merits, and we would view ethical decisions as ones that more appropriately balance or reconcile these conflicting viewpoints.

We have set out some of the background to the problems of automation in aviation systems, but are aware that there is much more that could be said (considering military UAS, for example). We hope, however, that the brief introduction provides a foundation for the ethical questions that we have set out.

The underlying aim in proposing ESCs is to make understanding ethical issues easier so that ethically-informed decisions can be made. Whilst we have not linked the discussion directly back to specific ethical decisions, we believe that making explicit those issues on which such judgments are based is a contribution to ethically informed decision making. We also believe that the four principles set out by the RAEng are reflected in this approach.

We acknowledge that what we have set out, especially the ideas of ESC, goes some way beyond current practice and principles and there are significant technical issues to resolve before such an approach could be implemented. It is hoped, however, that the ideas will help improve the production and presentation of safety cases in a range of industries not just aviation – a Pollyanna view, of course!

ACKNOWLEDGEMENTS

Thanks are due to Prof Sidney Dekker for his input on the problems faced by pilots in dealing with automation on current generation aircraft.

REFERENCES


