THE RISK OF NEGLIGENCE FOLLOWING THE FAILURE OF A HUMAN REPLACEMENT FUNCTION IN AN AUTOMATED SYSTEM

R. L. Maguire *, A. Banks †

* RS2A Limited, UK, rlm@rs2a.com
† Dstl, UK abanks@dstl.gov.uk

Keywords: Human replacement function, Negligence, Autonomous systems, RPAS, Caparo.

Abstract

In conventional human operated systems, unsafe failure of the system due to the behaviour of the operator can result in legal charges of negligence. However, if the human is replaced in the system, and a similar failure occurred, would similar charges be brought? And if so, against whom? In this paper, we review how the responsibility for correct system behaviour shifts, from the human controller to the design chain, when implementing human replacement functions. We do this using a pseudo-legal discussion of negligence tests in a variety of hypothetical scenarios.

It should be stressed that the authors are not legally trained and the contents of this paper should not be taken as legal advice. Further, the concepts and ideas described in this paper are solely those of the authors, and do not necessarily reflect the views of their parent organisations.

1 Introduction

Negligence is the area of Tort Law that involves harm through carelessness, not an intentional act. The modern UK law of negligence was established in Donoghue v Stevenson [1932] (AC 562)[3]; a relatively trivial case involving a dead snail in a bottle of ginger beer! Lord Atkin's ruling "You must take reasonable care to avoid acts or omissions which you can reasonably foresee would be likely to injure your neighbour.", has developed into a set of evidence requirements as a test of negligence.

Taking a technical, quasi-legal perspective, this paper discusses the legal tests for each of the proof stages and assesses the implications of the legal tests when judged against theoretical accident events that arise through the failure in performance of a human replacement function, carried out within an unmanned system. The paper proposes hypothetical alternative outcomes in a negligence case, based upon the options for the holder of the duty of care. Conclusions and recommendations are made for how a duty of care chain may be anticipated as part of programme risk management.

2 System and Accident Description

For the purposes of this paper, we describe an indicative non-real[1] fuel warning system for an RPAS that is being modified to implement a human replacement function. In the original system, which we will call System 1, (Figure 1) the controller enters initial mission data, via the Ground Control Station (GCS), that includes fuel-warning levels that will cause a warning to be triggered and provide an alert to prompt the controller to re-plan the route as required. This mission data is uploaded to the on-aircraft Vehicle Management System (VMS). The VMS uses it, along with information from the Flight Control System (FCS) (which in this system contains engine controls), and a Fuel Sensor to regulate fuel flow at the fuel pump as well as provide fuel data back to the controller who must re-plan the mission in response to warnings generated by the GCS. Failure of the controller to respond to the fuel warning in a timely manner could lead to the aircraft running out of fuel and crashing.

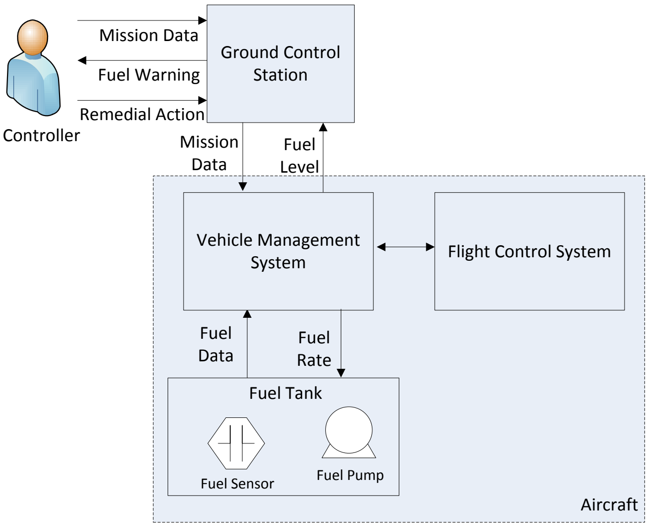

Figure 1. Human in the loop fuel system

In the non-real updated system, which we will call System 2, (Figure 2), an autonomous mission system is implemented in

[1] Non-real indicative systems are used to simplify the discussion and prevent the technical detail detracting from the main focus of the paper.

the VMS. This, *inter alia*, autonomously manages the fuel system. The controller now uploads the mission data including, minimum fuel levels and additional reversionary data (e.g. emergency landing sites) to the aircraft VMS via the GCS. In contrast to the original system, if the VMS detects a low fuel level in flight, it autonomously re-plans using the reversionary data and controls the aircraft accordingly without the need for controller intervention. However, the controller is informed of the re-planning via the mission status data sent to the GCS and can intervene if required. In the event of unexpectedly low-fuel, failure of the autonomous algorithms to re-plan the flight path appropriately could cause the aircraft to run out of fuel and crash.

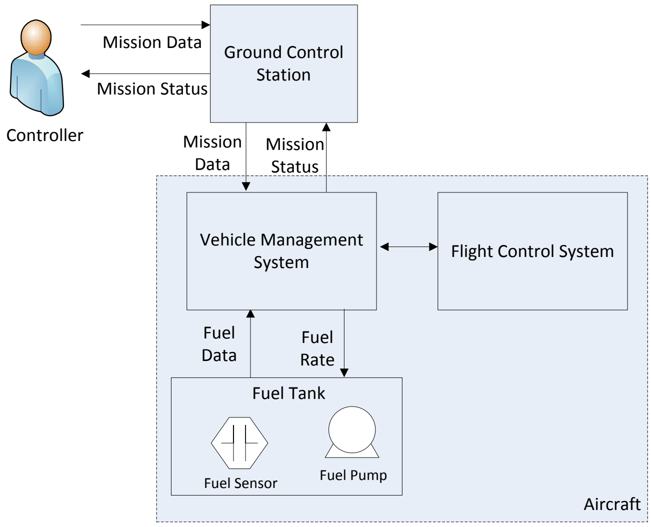

Figure 2. Human on the loop fuel system

Having described the basic system, we now describe a small fuel leak event that will allow us to consider the failure of the human replacement function. Here, the original fuel system will continue to operate and sends the correct fuel level to the VMS. Beginning with System 1, in our original Human in the loop system a warning is presented to the controller. Given the existing fuel state and leak rate, the controller has only one Reversionary Landing Site (RLS) in range, we will refer to this as RLS1. However, despite training that required the controller to compare the expected and actual fuel level to deduce a fuel leak and its rate, he fails to deduce the leak and elects not to select that landing site, choosing instead an RLS further afield (RLS2). Due to insufficient fuel this is now out of range and before the aircraft reaches its destination it loses power and crashes in a populated area.

Switching now to System 2, using the same scenario, the autonomous mission planner is sent the correct fuel data and re-plans a route to an RLS. However, due to algorithms that have the potential to re-plan sub-optimally (in this case because they fail to account for the leak rate), it too re-plans away from RLS1 preferring the out of range RLS2 and like System 1 it causes the aircraft to crash in a populated area.

To summarise, for the purposes of the following discussions we have described a basic scenario where an event occurs in which a human controller could, potentially, be found negligent for failing to respond appropriately to a low fuel warning. We then described a second scenario which was the same as the first except an autonomous mission management system fails to respond appropriately in the same way as the human. However, a machine cannot be negligent so we will now theorise who, if anyone, in the design and operational chain of responsibility could be found negligent.

3 Demonstrating Negligence in Law

In the aforementioned Donoghue v Stevenson case, Lord Atkin's ruling has developed into a four-step evidence requirement for the demonstration of negligence [3]. In order for a negligence claim to be demonstrated, the claimant must provide evidence to demonstrate the following:

1. The defendant owed them a duty of care;
2. The defendant was in breach of that duty;
3. The breach of duty caused damage; and
4. The damage was not too remote.

This may be represented in a GSN structure as follows in Figure 3:

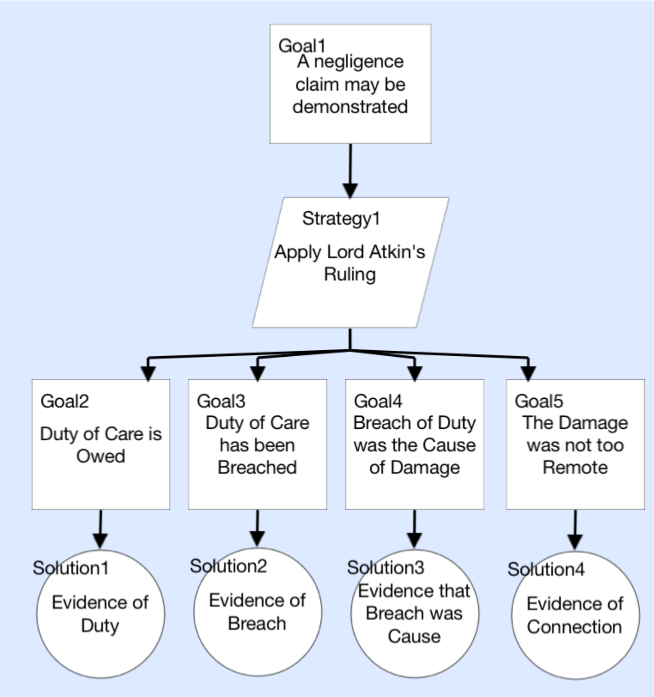
Goal1
A negligence claim may be demonstrated
Strategy1
Apply Lord Atkin's Ruling
Goal2
Duty of Care is Owed
Goal3
Duty of Care has been Breached
Goal4
Breach of Duty was the Cause of Damage
Goal5
The Damage was not too Remote
Solution1
Evidence of Duty
Solution2
Evidence of Breach
Solution3
Evidence that Breach was Cause
Solution4
Evidence of Connection

Figure 3. GSN representation of Lord Atkin's ruling

The phrase 'Duty of Care' refers to the circumstances and relationships that the Law recognises as giving rise to a legal duty to take care. A failure to take such care can result in the defendant being liable to pay damages to the neighbour who has suffered injury or loss as a result. The existence of a Duty of Care depends on the type of loss that has been suffered e.g., injury, reputation, monetary; and there are different legal tests to apply to different types of losses.

The 'neighbour' test taken in its widest sense could be very broad allowing liability in a whole range of situations. However, in cases subsequent to the Donoghue vs. Stevenson snail case, the scope was narrowed down in the application to cases where a consumer was suing a manufacturer.

In 1978 Lord Wilberforce sought to resurrect an all embracing test for duty of care in judging the case of Anns v Merton London Borough Council (1978) AC 728 [1]. Lord Wilberforce's constructed a two stage test:

1. Examine whether the loss was reasonably foreseeable and there existed a relationship of proximity. If so a prima facie duty of care arises.

2. The defendant may put forward policy considerations to negate liability.

The first stage was essentially the elements of the original neighbour test, however in order to address the fears of floodgates of claims that might arise, this was subject to the second stage which provided a get out clause for defendants where there existed policy or explicit legal reasons for denying the imposition of a duty of care.

4 The Caparo Test for Negligence

In novel situations (perhaps covering our RPAS human replaced actions), the question of whether there is a duty of care is now subject to the Caparo test, from Caparo Industries vs. Dickman (1990) 2 AC 605 [2]. In this case, Lord Bridge established a three-stage test for imposing a duty of care, known as the Caparo test:

Under the Caparo test the claimant must establish:

1. That harm was reasonably foreseeable
2. That there was a relationship of proximity
3. That it is fair, just and reasonable to impose a duty of care

This may be represented in GSN as follows in Figure 4.

It can be seen that the first two stages are taken directly from the original neighbour test. Fair, just and reasonable relates to the same policy considerations under the Anns test. However, the principle difference is that in the Caparo test, it is the claimant that has to demonstrate the policy or explicit legal reasons for imposing a liability, rather than the defendant showing there was no imposition of duty of care.

Lord Bridge's statement on the case is as follows;

"What emerges is that, in addition to the foreseeability of damage, necessary ingredients in any situation giving rise to a Duty of Care are that there should exist between the party owing the duty and the party to whom it is owed, a relationship characterised by the law as one of "proximity" or "neighbourhood" and that the situation should be one in which the court considers it fair, just and reasonable that the law should impose a duty of a given scope upon the one party for the benefit of the other.” [*ibid*].

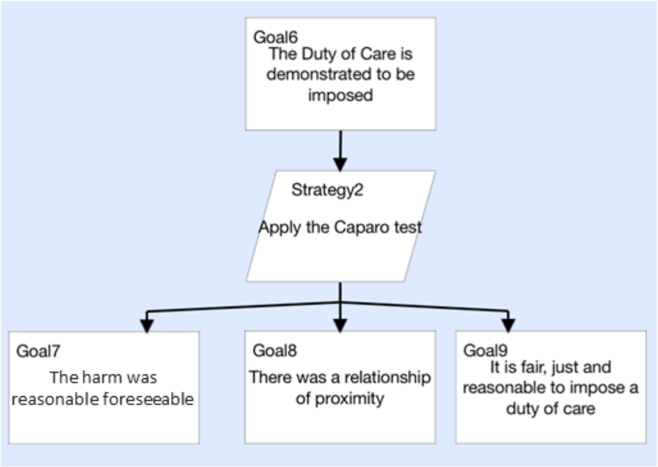
Goal6
The Duty of Care is demonstrated to be imposed
Strategy2
Apply the Caparo test
Goal7
The harm was reasonable foreseeable
Goal8
There was a relationship of proximity
Goal9
It is fair, just and reasonable to impose a duty of care

Figure 4. GSN representation of the Caparo Test.

5 Other applicable legislation

In what might be considered high-risk industry, the policy reasons for imposing a liability for duty of care are often mandated through policy articles, legislation and statute. For example, the Health and Safety at Work Act (HASAW); the Supply of Machinery (Safety) Regulations; or the Provision & Use of Workplace Equipment Regulations (PUWER).

Within the aviation domain, EASA has assumed responsibility for the type-certification and continued airworthiness of a large number of UK registered aircraft. The list of specific EASA and non-EASA aircraft types is contained in CAP 747 - "Mandatory Requirements for Airworthiness" [4]. CAP 747 also provides a statement of the general categories of aircraft that are excluded from European Regulations and so remain subject to National rules.

As of 16 July 2008, Commission Regulation (EEC) No. 3922/91 Annex III (EU-OPS) applied to all aeroplanes operated by European Community Member States' operators [5]. The EU-OPS requirements apply to all EU operators of aeroplanes flying for the purpose of commercial air transportation. The EU-OPS requirements supersede the Air Navigation Order (ANO) in these relevant areas and the objective is to reflect these changes so far as necessary in the ANO. EU-OPS does not apply to helicopters.

The Regulation aims in particular to enhance aviation safety and promote a level playing field in commercial air transportation within the European Union. The Regulation transposes the non-binding Joint Aviation Requirements (JAR-OPS) established by the Joint Aviation Authorities at non-binding inter-authority level into a binding Community legislative act.

The aim of enhancing safety ultimately leads to a suite of derived requirements to satisfy industry good practices in the development of aircraft and aviation systems. In our case, this would lead to the application of guidance such as RTCA DO-178 [8], RTCA DO278A [9] and RTCA DO254[7] for example.

There is also further legislation that exists to cover the 'everything else we haven't thought of' case, or what might even be thought of as 'emergent situations' e.g., autonomous systems perhaps. SI2005:1803 implementing 2001/95/EC titles as The General Product Safety Regulations, is a statutory requirement for all Producers. These regulations apply to a product in so far as there are no specific provisions with the same objective in rules of Community Law governing the safety of the product. It applies to any products that are intended for consumers or are intended for professional use, which it can be reasonably foreseen, may be used by consumers, and for products supplied in the course of a service.

Regulation 3 of SI2005:1803 provides that the Regulations apply except where there are no other specific provisions in rules of Community law other than the Directive. Regulation 5 requires producers only to place safe products on the market. Regulation 6 provides that a product, which complies with certain safety standards, is presumed to be safe unless there is evidence to the contrary. Regulation 7 requires producers to inform customers about the risks of products and to monitor the risks their products pose. Regulation 8 requires distributors to act with due care so as not to supply unsafe products and to co-operate in monitoring the safety of products.

6 Typical Supply Chain – Who could be in the Dock?

From the range and breadth of legislation and guidance that is within the legal framework, there are multiple, separate individuals and organisations that have to take cognisance of the potential for them to have a duty to take care on behalf of their neighbour. For the analysis in this paper, these groups are collated into three clusters;

1. Designer Organisations and Producers. These would be the design companies, Prime and sub-contractors – the groups that design and produce unmanned systems, the VMS, GCS or the hardware and software thereof.

2. Assurers and Independent Evaluators. These would be the subject matter experts and specialist engineers with legal standing that offer independent advice to the design organisations or operators about the satisfaction of legislation, levels of assurance/integrity and certification requirements.

3. Operators and Users. These would be considered as the humans in or on the loop; they fly and have 'control' over the unmanned system's flight – including operational duty holders, flight authorisers, maintainers and the remote pilots.

In many cases, it has been the pilot in charge of the aircraft that has faced the negligence charge. Typically, the charge is that the pilot failed to ensure the mission was executed in a manner, which minimised the risks to the aircraft, its occupants or the wider public over whom the aircraft was flown. The pilot may also be charged with failing to advise the aircraft commander accordingly and if necessary, to offer specific guidance to avoid hazardous situations [6]. We suggest that this may be somewhat different in an autonomous or automatic platform, which includes multiple human replacement functions.

7 Accident and Pre-Accident Scenarios

To show how subtly different but very realistic outcomes can arise, it is necessary to allow for some sensitivity analysis in the System 2 situation (described in Section 2) – the situation with the human-replacement function in place. We now provide some additional context around the potential conclusion to our scenarios.

In our hypothetical event there were no fatalities, but several members of the public were injured by debris and the £1m airframe and equipment was damaged beyond economic repair. The inquiry reviews the integrity and assurance of the VMS software that is responsible for the human-replacement fuel management function. Three pre-accident scenarios are developed and then further evaluated with regard to negligence in Section 8.

1. The fuel management function was allocated a Design Assurance Level (DAL) -C assurance level under RTCA DO-178C [8]. The independent evaluators assessed the design processes of the designers and sampled the test and verification processes in accordance with the FAA recommended audit processes. The design process and solution was given independent assurance to DAL-D, with a proposal that there should be additional mitigation e.g. additional verification and validation activities or a duplex arrangement to achieve DAL-C. The designer corroborates and accepts the DAL-D level that has been achieved and introduces the duplex design. The independent evaluators re-assess the process for introducing the change and give endorsement that the DAL-C functional requirement has been satisfied. The designer passes on the independent assessment and recommendations to the Operator, highlighting the residual risk via a software accomplishment summary and a safety assessment report. The Operator accepts the reports.

2. The fuel management function was allocated a DAL-C assurance level under RTCA DO-178C [8]. The independent evaluators assessed the design processes of the designers and sampled the verification and validation processes in accordance with the FAA recommended audit processes. The design process and solution were given independent assurance to DAL-C, however they (the designer

organisation) do not fully take into account the full range of human behaviour the new fuel management function is replacing and does not implement leak deduction algorithms. This is not detected by the independent assurance activity which as stated takes a sampling approach. The Designer passes on the independent assessment and recommendations to the Operator, highlighting the known residual risk via a software accomplishment summary and a safety assessment report. The Operator accepts the reports.

3. This fuel management function was allocated a DAL-C assurance level under RTCA DO-178C [8]. The design process and solution was given independent assurance to DAL-D, with a proposal that there should be additional mitigation e.g. additional verification and validation activities or a duplex arrangement to achieve DAL-C. The designer disagrees that DAL-C has not been met, but passes on the independent assessment and recommendations to the Operator, highlighting the potential increased risk via a software accomplishment summary and a safety assessment report. The Operator accepts the reports but fails to introduce additional monitoring, by the Controller, of the Mission Status reports and hence the incorrect re-planning was not detected.

8 Liability Outcomes Discussion

In all scenarios, we suggest that several parts of the Caparo test [2] are already partially met – particularly the identification of 'who is my neighbour?' Due to proximity, the Designer has a duty of care to the Operator; there will be a contract of supply between them and an agreed set of legislation to be met. The Independent Assurers also have a duty of care to (perhaps) both the Designer and the Operator – to be careful in their advice and endorsement. The Operator has a duty of care to the Designer – to operate the design in accordance with the safety limitations. The Operator will also have a duty of care to the public, as this would be reasonably expected for a flying authority in public airspace.

In pre-accident scenario 1, we suggest that in this case, the Designers and Independent Assurers are in breach. The key point is that the final DAL-C assurance has been based on the design process to change to the duplex system. Neither the designer nor the independent assurers have carried out additional reasonable safety or airworthiness analysis of the new system design and its common cause failure modes. We suggest that the Operator is not in breach; they followed correct procedures against the airworthiness and safety evidence given to them. From their point of view, the harm was not reasonably foreseeable; therefore, it would not be fair and reasonable for them to hold the duty in this area.

In pre-accident scenario 2, we suggest that in this case, the Designer is in breach. The Designer has clearly failed to implement behaviours that the human controllers were trained to perform (i.e. to deduce a fuel leak from unexpectedly low fuel levels) and in producing the human replacement function they had a duty of care to implement all behaviour that the human could have been reasonably expected to undertake. The Independent Assurers are not in breach – it would not be fair and reasonable to expect them to check that all reasonable human behaviour had been implemented. They had undertaken an appropriate and recognised sampling strategy and in doing so had discharged their duty of care. Furthermore, the Operators, acted in good faith based on the information presented to them and could be reasonably expectant that the function fully replaced the human behaviour.

Finally, in pre-accident scenario 3, we suggest that the Operator is in breach. The Designer and Independent Assurers have identified and communicated a lower assurance level and reasonably foreseeable harm to the Operator. There is an express recommendation for additional mitigation. The failure to take care through introducing effective additional operational mitigation fulfils part of the Caparo test [2], and we suggest that it would be fair and reasonable for a duty to take care to be imposed.

9 Conclusions

As well as providing an overview of the principle of negligence, we have discussed three potential scenarios that each produces a different possible negligent party. In many ways this is can be thought of as being no different to any conventional system, where people or organisations can engage in negligent behaviour. However, the subtle difference here is that the previous system had a human in the loop that could respond to the situation and was clearly negligent in accident scenario 1 described in Section 2. In accident scenario 2 though, the negligence was not clear until we presented the pre-scenario discussion and whilst we have tried to be as obvious as possible with respect to the act of negligence for this paper, such behaviour might not be so obvious. Even so there are a multiple of complexities that could cloud where negligence occurred, particularly where a system has been independently assessed. The purpose of the discussion in this paper though, is to draw attention to the hidden feature of human replacement functions: that machines do not have responsibility for their behaviour and so negligence in the case of a machine failing to behave in a reasonable manner could lie earlier in the design chain.

We therefore conclude that implementations of human replacement functions have the potential to shift responsibility for ensuring correct platform behaviour away from the immediately obvious neighbour (the operator) and into the design and assurance chain. We suggest that the Caparo tests concerning fairness and reasonableness may increase the liability risk to those parties and hence, arguably, increases the need to place greater emphasis on performing and recording correct individual and organisational behaviour.

References

[1]. British and Irish Legal Information Institute. " Anns and Others (Respondents) V. London Borough Of Merton (Appellants)", Lords' Journals, 12th May (1977)

[2]. British and Irish Legal Information Institute. " Caparo Industries Plc. Respondents and Dickman And Others Appellants", The Law Report Appeal Cases, 8th Feb (1990)

[3]. British and Irish Legal Information Institute.", M'Alister (or Donoghue) (pauper) appellant; and Stevenson respondent. ", Lords' Journals, May 26 (1932).

[4]. CAA "CAP 747 Mandatory Requirements for Airworthiness", Civil Aviation Authority, 25th November (2014).

[5]. European Commission. " As Regards Common Technical Requirements and Administrative Procedures Applicable to Commercial Transportation by Aeroplane" Commission Regulation (EC) No 859/2008 of 20 August (2008).

[6]. Evening Gazette. “Crash Pilot Admits Negligence”, The Evening Gazette, 15th November (2011).

[7]. RTCA, "Design Assurance Guidance for Airborne Electronic Hardware", RTCA DO-254, RTCA Incorporated, Washington, USA (2000)

[8]. RTCA, "Software Considerations in Airborne Systems and Equipment Certification", RTCA DO-178C, RTCA Incorporated, Washington, USA (2011)

[9]. RTCA, "Software Integrity Assurance Considerations for Communication, Navigation, Surveillance and Air Traffic Management (CNS/ATM) Systems", RTCA DO-278A, RTCA Incorporated, Washington, USA (2011)