Conferences in Research and Practice in Information Technology

Volume 151

Australian System Safety Conference 2013

Australian System Safety Conference 2013

Proceedings of the
Australian System Safety Conference (ASSC 2013),
Adelaide, Australia, 22nd-24th May 2013

Tony Cant, Editor.

Volume 151 in the Conferences in Research and Practice in Information Technology Series.
Published by the Australian Computer Society Inc.

Published in association with the ACM Digital Library.

Australian System Safety Conference 2013. Proceedings of the Australian System Safety Conference (ASSC 2013), Adelaide, Australia, 22nd-24th May 2013

Conferences in Research and Practice in Information Technology, Volume 151.

Editors:
Tony Cant

6 Landowne Tce. Walkerville SA 5081
E-mail: canttony@gmail.com

Series Editors:
Vladimir Estivill-Castro, Griffith University, Queensland
Simeon Simoff, University of Western Sydney, NSW
crpit@infoeng.flinders.edu.au

Publisher: Australian Computer Society Inc.
PO Box Q534, QVB Post Office
Sydney 1230
New South Wales
Australia.

Conferences in Research and Practice in Information Technology, Volume 151
ISSN 1445-1336
ISBN 978-1-921770-38-8

Printed May 2014 by Griffith University, CD proceedings.

The *Conferences in Research and Practice in Information Technology* series aims to disseminate the results of peer-reviewed research in all areas of Information Technology. Further details can be found at http://crpit.com/.

Table of Contents

Preface

The *Australian System Safety Conference 2013* was held at the Intercontinental Hotel, Adelaide, on 22-24 May, 2013. The conference, jointly sponsored by the Australian Safety Critical Systems Association (aSCSa) and the Australian Chapter of the System Safety Society, had the theme: "Automation and Resilient Systems" and was attended by more than 100 participants. The conference program was greatly enhanced by four keynote speakers:

- Prof Sidney Dekker (School of Humanities, Griffith University, Australia)
- David West (Science Applications International Corporation, USA)
- Dr Alan Hobbs (San Jose State University Foundation, USA)
- Prof John McDermid (Department of Computer Science, University of York, UK)

A program of tutorials was also held prior to the conference. Full program details are available from `asssc.org/conf2013`. More information on the aSCSa can be found at `www.safety-club.org.au`.

The Organising Committee is very grateful to the authors for the trouble they have taken in preparing their work to be included in these conference proceedings. The papers were peer-reviewed for relevance and quality by the Program Committee. Note, however, that the views expressed in the papers are the authors' own, and in no way represent the views of the editor, the Australian Safety Critical Systems Association, the System Safety Society, or the Australian Computer Society. The fact that the papers have been accepted for publication should not be interpreted as an endorsement of the views or methods they describe, and no responsibility or liability is accepted for the contents of the articles or their use.

The committee also wishes to thank the conference sponsors for their support: the Australian Computer Society; Adacore; BAE Systems; RGB Assurance; Nova Systems; Airservices Australia; and the Defence Materiel Organisation in the Australian Government Department of Defence. These organisations have all helped to make the conference a success.

I wish to thank all those involved in organising the conference (listed below). In particular, I would like to acknowledge the commitment and drive of my colleagues Clive Boughton, B.J. Martin, Kevin Anderson, Holger Becht and Anthony Acfield, who worked hard to make sure that the conference was a success. Special thanks are due to Michael Colsey (of Nova Systems), who assisted with local arrangements.

We are also grateful to Brian Clegg of the Australian Computer Society for his administration of the web-based registration system, as well as Alison Pitman (Events Manager - SA) and Mandy Watson (Branch Manager - SA) of the Adelaide Branch of the ACS for their assistance. Finally, our thanks to the Computer Systems and Software Engineering Board of the ACS for ongoing support.

Tony Cant,
May, 2014

Programme Committee

Programme Chairs

ASSC 2013 Organising Committee:

- Brett J. Martin (Chair)
- Clive Boughton (Program Chair)
- Holger Becht (Vice Chair)
- Kevin Anderson (Facilities and Operations)
- Michael Holsey (Local Arrangements)
- Anthony Acfield (Publicity)

ASSC 2013 Program Committee:

- Clive Boughton (Chair)
- Holger Becht (Vice Chair)
- Simon Connelly (Member)
- Derek Reinhardt (Member)
- George Nikandros (Member)
- Clive Boughton (Member)
- Tariq Mahmood (Member)
- Tim Kelly (Member)
- Paul Caseley (Member)
- Rob Weaver (Member)
- Brendan Mahony (Member)
- Tony Cant (Member)

Australian Safety-Critical Systems Association Committee:

- Clive Boughton (Chair)
- George Nikandros (Immediate Past Chair)
- Kevin Anderson (Secretary)
- Chris Edwards (Treasurer)
- Brett J. Martin (Member)
- Tony Cant (Member)
- Tariq Mahmood (Member)
- Anthony Acfield (Member)
- Luke Wildman (Member)
- Derek Reinhardt (Member)

Research Papers

ALARP and the Risk Management of Civil Unmanned Aircraft Systems

Reece A. Clothier[1] **Brendan P. Williams**[2*] **Neale L. Fulton**[3] **XunGuo Lin**[4]

[1] School of Aerospace, Mechanical, and Manufacturing Engineering
RMIT University,
PO Box 71, Bundoora, Victoria 3083, Australia
Email: reece.clothier@rmit.edu.au

[2] Australian Research Centre for Aerospace Automation
Queensland University of Technology,
22-24 Boronia Road, Brisbane Airport, Queensland 4008, Australia
Email: bp.williams@qut.edu.au

[3,4] Mathematics, Informatics and Statistics
Commonwealth Scientific and Industrial Research Organisation (CSIRO),
GPO Box 664, Canberra, Australian Capital Territory 2601, Australia
[3] Email: neale.fulton@csiro.au
[4] Email: xunguo.lin@csiro.au

Abstract

Key to the continued growth of the civil Unmanned Aircraft System (UAS) aviation sector is the development of a regulatory framework that will provide assurances in the management of the risks associated with their operation. Decisions in relation the evaluation and treatment of aviation risks need to be made in accordance with the As Low As Reasonably Practicable (ALARP) framework. There are a number of concerns in relation to the application of the ALARP framework to new technologies. This paper explores these concerns with respect to the risk management of civil UAS.

A review of the ALARP frameworks defined by the International Civil Aviation Organization (ICAO), the Civil Aviation Safety Authority (Australia), the Civil Aviation Authority (United Kingdom) and by the UK Health and Safety Executive is presented. This review identified subtle differences that can have a significant impact on how ALARP frameworks would be applied to UAS. A number of inconsistencies in the frameworks were also identified. These issues aside, it was found that a conceptual application of an ALARP framework can be made. However, significant difficulties were identified in the substantiation of a framework. In particular, the quantification of the decision criteria for UAS, the handling of uncertainty, and the identification, characterisation and representation of societal concerns within a framework. Guidance as to how the dimensions of societal concern and levels of risk can be jointly considered within an ALARP framework could not be identified within the literature. For new technologies such as UAS, these dimensions can be as significant a factor in decision-making as that of the quantified measures of the risk. Due to these deficiencies, there are significant difficulties in the application and substantiation of an ALARP framework to the risk management of new technologies such as UAS.

Keywords: ALARP, Unmanned Aircraft Systems, Risk Management, Regulation

1 Introduction

Much research is ongoing in the development of regulations to facilitate the safe, routine operation of UAS in civil airspace particularly over populated regions. In accordance with International Civil Aviation Organization (ICAO) Standards And Recommended PracticeS (SARPS), National Airworthiness Authority (NAA) policy, rule-making, and oversight activities should be governed by a systematic Safety Risk Management Process (SRMP) (ICAO 2009). A general description of the application of the SRMP to UAS can be found in Clothier & Walker (2013).

A number of frameworks can be used to support decision-making within the SRMP and in particular within the risk evaluation and treatment subprocesses. These decision-making frameworks include: As Low As Reasonably Practicable (ALARP), So Far As Is Reasonably Practicable (SFAIRP), As Low As Reasonably Achievable (ALARA), Globalement Au Moins Aussi Bon (GAMAB), Globalement Au Moins Equivalent (GAME), and Minimum Endogenous Mortality (MEM). Discussion on the differences between these decision-making frameworks can be found in Johansen (2009). ICAO SARPS stipulate that decision making within the SRMP should be made in accordance with the ALARP framework, Figure 1.

A review of the different specifications of the ALARP framework is provided in section §2. An essential component of the ALARP framework are the decision criteria that demarcate the different decision making regions. The definition of these decision criteria and discussion on the issues associated with their specification are presented in section §3. Decision making within the ALARP framework requires

* Mr BP Williams is on secondment to the Queensland University of Technology from Boeing Research & Technology - Australia. Email: brendan.p.williams@boeing.com

consideration of the level of risk and the degree of societal concern associated with the operation of UAS. The societal concerns associated with UAS operations and their representation in the ALARP framework are presented in section §4.

2 The ALARP Framework

The ALARP risk decision-making framework is intended to reflect the types of safety decisions made in everyday life (HSE 1992, 2001*b*). These decisions are based on the Level of Risk (LoR) and the degree of societal concern associated with the particular technology, activity or situation under assessment. From Figure 1, a particular situation can be classified as either:

1. Unacceptable, intolerable or broadly unacceptable (§2.1);
2. Tolerable or requiring review (§2.2);
3. Acceptable or broadly acceptable (§2.3); or
4. Negligible.

Central to the classes of *tolerable* and *acceptable* risks is the meaning of ALARP. This is discussed in Section §2.4. The varying definitions and conditions associated with each class are described in the following sub-sections.

2.1 Unacceptable, Intolerable or Broadly Unacceptable Risk

There are certain activities where people are unwilling to accept a risk regardless of the potential benefits. Situations of this nature have been referred to as unacceptable, intolerable or broadly unacceptable. In an effort to simplify terms, this paper will use the single term *unacceptable* in place of the various terms defined in the ALARP frameworks. This will maintain consistency with Figure 1 and reduce possible ambiguity.

ICAO describe an unacceptable LoR as "unacceptable under any circumstances, where the probability and/or severity of the consequences of the hazards are of such a magnitude, and the damaging potential of the hazard poses such a threat to the viability of the organization, that immediate mitigation action is required" (ICAO 2009). Civil Aviation Safety Authority (CASA) guidance material defines the concept of an unacceptable LoR as being unacceptable regardless of the benefits associated with the activity (CASA 2012).

The conditions associated with a situation deemed as having an unacceptable LoR can vary. Under the ICAO specification of the ALARP decision-making framework, an unacceptable risk must be reduced to a tolerable level or the activity cannot be undertaken. Similarly, the ALARP framework defined by CASA states that risk reduction measures "are essential regardless of cost" (CASA 2009). The CASA framework includes the additional statement that an activity assessed as having an unacceptable LoR may still continue but only where there exists "exceptional reasons" or "extraordinary circumstances". The concept of unacceptable LoR defined by the Health and Safety Executive (HSE) in the UK also includes the additional condition that activities may be allowed to continue if there are "exceptional reasons" (HSE 2001*b*). Definitions of an "exceptional reason" or an "extraordinary" or "extenuating" circumstance are not provided.

2.2 Tolerable Risk or Risks Requiring Review

For most activities stakeholders are willing to tolerate risk in return for certain benefits associated with the activity. These situations have been described as tolerable or those "requiring review". As per Figure 1 and in keeping with our endeavour to simplify terms, the term *tolerable* also refers to the concept of "requiring review" as defined in existing frameworks.

The Civil Aviation Authority (CAA) in the United Kingdom (UK) describe a LoR that requires review as being a LoR where the consequence and/or likelihood is of concern (CAA 2010*a*). CASA and HSE UK describe a tolerable risk as a risk that people are generally prepared to tolerate in order to secure benefits (CASA 2012, HSE 2001*b*). Within these frameworks, a tolerable risk is further described as a risk that 1) has been properly assessed and appropriate measures to control the risk have been implemented, 2) where the residual risk is not unacceptable and is considered to be ALARP, and 3) will be periodically reviewed (CASA 2009, 2012, HSE 2001*b*).

The CAA state that measures to mitigate a risk to ALARP should be sought for all risks belonging to the review class (CAA 2010*a*). Similar conditions are mandated for tolerable risks in the CASA and HSE frameworks. Under the ICAO ALARP framework, risks initially assessed as being tolerable do not require further mitigation provided mitigation strategies already in place guarantee that, to the foreseeable extent, the probability and/or severity of the consequences of hazards are kept under organizational control (ICAO 2009). An explicit definition of "organizational control" is not provided.

With the exception of the ICAO framework, all of the reviewed decision frameworks explicitly associate the concept of ALARP with the tolerable decision region. Within the ICAO Safety Management Manual, it is stated that safety risks must be managed to ALARP. However, an explicit association between this requirement and the concept of a tolerable risk is not made.[1]

For the CAA, CASA and HSE ALARP frameworks, the concept of a tolerable risk requires more than the demonstration that the risk is ALARP (e.g., it has also been correctly assessed and reviewed, etc.).

2.3 Acceptable or Broadly Acceptable Risk

Often many of the risks in daily life are accepted as insignificant or trivial, or are accepted because we have no practical control over them. Situations of this nature have been referred to as acceptable or broadly acceptable. In this paper the term *acceptable* also refers to the concept of broadly acceptable as defined in existing frameworks.

The CAA describe an acceptable risk as one where the occurrence of a "consequence is so unlikely or not severe enough to be of concern" (CAA 2010*a*). CASA and the HSE define the concept of acceptable risks as those situations where the risks are generally regarded as sufficiently low, insignificant, and adequately controlled (CASA 2009, 2012, HSE 2001*b*). ICAO describes acceptable risks as those that are acceptable as they currently stand (ICAO 2009).

Under the ICAO framework an activity initially assessed as having an acceptable LoR requires no action "to bring or keep the probability and/or severity of the consequences of hazards under organizational

[1] A linkage between the concept of ALARP and tolerable risks is implied in a number of examples and illustrations throughout the document; see ICAO (2009).

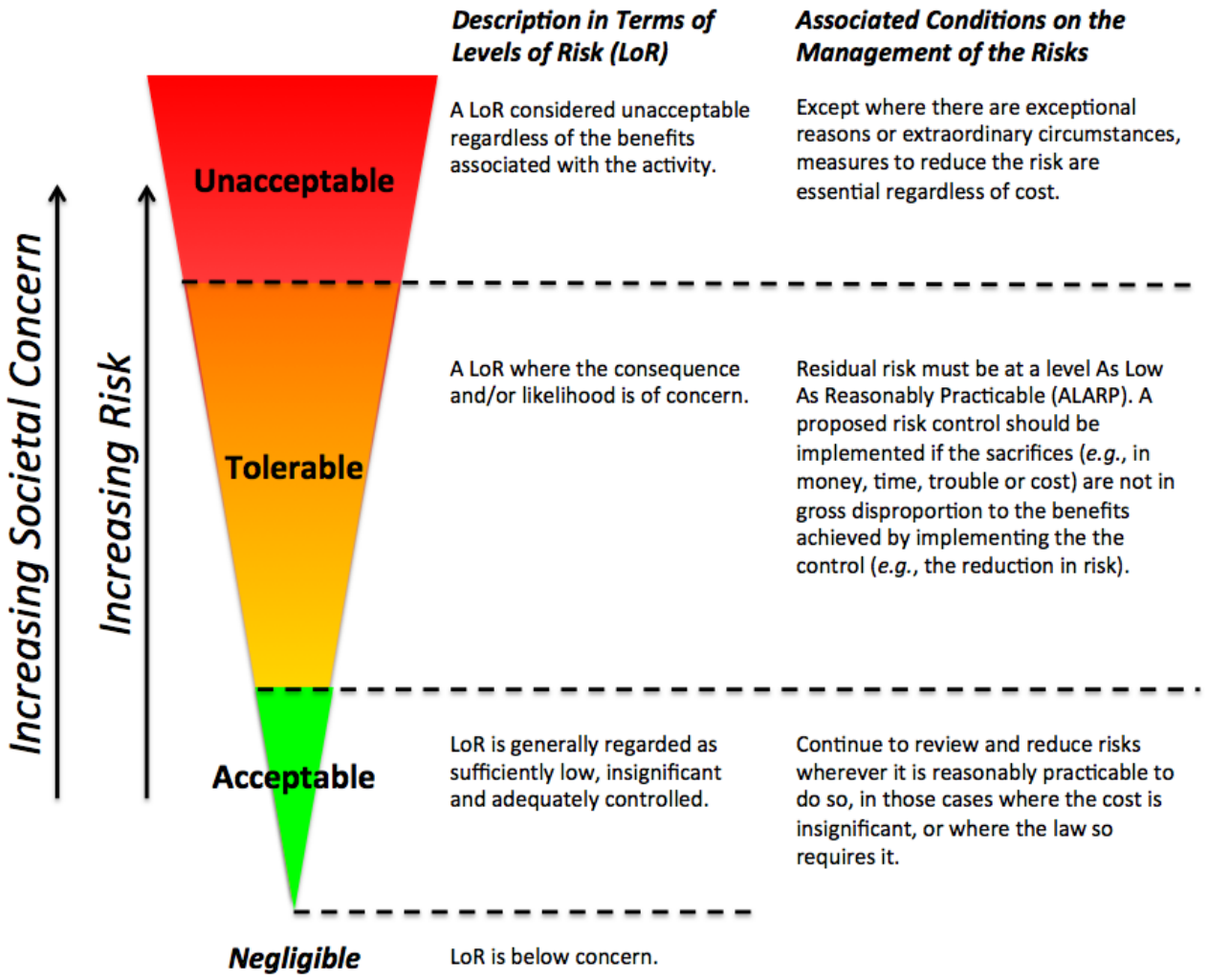

Figure 1: ALARP Risk Framework

control" ICAO (2009). Whereas, under the CASA, HSE and CAA decision-making frameworks, the reduction of an identified risk to ALARP is not abandoned; individuals and organisations should continue to review and reduce risks wherever it is reasonably practicable to do so, in those cases where the cost is insignificant, or where the law so requires it (CAA 2010*b*, CASA 2009, HSE 2001*b*). This approach is consistent with the goal of continual safety improvement wherever practicable (CAA 2010*a*).

CASA states that detailed working necessary to demonstrate that risks are ALARP may not be required for those risks initially assessed as being acceptable (CASA 2012).

2.4 The Concept of ALARP

Central to the definition of the tolerable decision region is the concept of a risk being reduced to ALARP. But what does this mean? CASA and the CAA define the concept of ALARP as where the risk is low enough that attempting to make it lower, or the cost of assessing the improvement gained in an attempted risk reduction, would actually be more costly than any cost likely to come from the risk itself (CAA 2010*a*, CASA 2009). ICAO describes ALARP as the point where it can be shown that any further risk reduction is either impracticable or grossly outweighed by the cost, which requires consideration of the technical feasibility of further reducing the safety risk, and the cost (ICAO 2009).

CASA (2009) and the HSE (1992, 2001*b*) explicitly relate the concept of ALARP with the concept of a gross disproportionality. Specifically, ALARP is the point where "the cost of reducing the risk is grossly disproportionate to the benefit gained" (CASA 2009) and determining that risks have been reduced to a level as ALARP involves an assessment of the risk to be avoided, of the sacrifice or costs (e.g., in money, time and trouble) involved in taking measures to treat that risk, and a comparison of the two to see if there exists a gross disproportion (HSE 2001*b*). The meaning of gross disproportion is further discussed in the next Section §2.5.

2.5 Gross Disproportionality

CASA (2009) and HSE (2001*b*) explicitly relate the concept of ALARP with that of a gross disproportionality between the benefit and costs associated with assessing and implementing measures to further reduce the risk. A finding of gross disproportionality should be supported by a cost benefit analysis. Guidance on the analysis process and the meaning of gross disproportionality can be found in HSE (2001*a*), Jones-Lee & Aven (2011) and CASA (2010). It is important to note that a finding of gross disproportion, on its own, is not sufficient for a determination of ALARP.

Gross disproportion is typically not represented as a single value but a range of values expressed on a finite scale (or ranking), which increases as a function of increasing risk. The higher the LoR the higher the ratio of the cost to benefit that is needed to be considered in gross disproportion; see (HSE 1992, 2001*a*,*b*).

In the context of the safety risk management of civil aviation, a quantified specification of the ratio of costs to benefit sufficient to constitute gross disproportion could not be identified in the literature. Even if existing scales were available, they may not be appropriate for UAS due to differences in the nature of the costs and benefits that need to be evaluated (e.g., greater uncertainty in their estimation, and differences in visibility and equity of the distribution of benefit to people exposed to the risks). A quantified specification of the condition for gross disproportionality is not necessary to substantiate the ALARP framework. A determination of gross disproportion can be made qualitatively, thus avoiding the quagmire of social and political issues associated with placing a value on different loss outcomes (e.g., a cost per life saved).

2.6 Summary of Findings

The review has identified subtle differences between the various specifications of the ALARP decision-making frameworks available in the literature. These small differences can have a significant influence on the risk management of UAS operations. For example, in contrast to other specifications, the ICAO specification of the ALARP framework does not mandate that all tolerable risks be reduced to ALARP. Consequently, the LoR associated with an UAS operation can be considered *tolerable* within the ICAO ALARP framework but may not be *tolerable* within other frameworks. Differences in the associated conditions will also impact the setting of quantitative safety criteria within the framework. For example, the ICAO framework may set more stringent risk criteria to demarcate the region of *unacceptable risk* from *tolerable risk* to account for less stringent conditions on the management of the risks (e.g., not all tolerable risks need to be mitigated to ALARP). Further, National Airworthiness Authorities will need to determine a common set of high level safety goals within their respective State aviation safety programs. For example, the goal for continual safety improvement will influence the specification of the ALARP framework and its application to UAS.

A number of deficiencies were also identified. Key terms are undefined or only loosely defined. For example, the concepts of "organisational control" or "exceptional reason". There are also conceptual difficulties encountered in the different frameworks. Within the ICAO Safety Management Manual, it is stated that safety risks must be managed to ALARP. This statement conflicts with the definition of the decision classes of *intolerable* and *acceptable* defined within the ICAO ALARP framework. Further, mandating that all risks should be managed to ALARP conflicts with other requirements defined within the ICAO ALARP framework (e.g., conflict with the condition that risks initially assessed as being tolerable do not require further mitigation, described in section §2.2).

Irrespective of which ALARP framework is to be adopted, it is clear that the application of the ALARP framework in the risk management of UAS will require the assessment of more than the safety risks. Estimates of the benefit and cost associated with UAS operations as well as those associated with the risk management activity itself (e.g., the cost in time and money to conduct further assessment and evaluation) need to be made. For new technologies such as UAS this is very much a "chicken and egg" scenario, where knowledge of the true benefits and costs may not be entirely known *a priori* the operation. There can be as much uncertainty in the assessment of the benefits and costs as there is uncertainty in the assessment of the risks. Whilst there is much research being undertaken to understand and quantifiably characterise the safety risks associated with UAS technologies, very little is being undertaken to characterise the associated benefits and costs to society.

For some classes of UAS, the primary consideration may not be the safety risks associated with their operation. The quantified risk analysis conducted by Clothier et al. (2010), Magister (2010) and Fraser & Donnithorne-Tait (2011) show that for some types of UAS and for some UAS operations there is negligible risk to people and property. For such UAS and UAS operations the dominating factor driving risk reduction (a determination of ALARP) are the costs and associated benefits. It is important to note that the assessment of the benefit and cost to society requires more than an assessment of the economic values associated with the UAS operation (e.g., the economic loss due to destruction of the unmanned aircraft or damage to third party property). For example, society can place value on a wide range of tangible and intangible objects (e.g., reputation, trust in technology/brand). The gain (benefit) and the loss (cost) registered to these objects of value need to be considered in a determination of the ALARP. Finally, one should also consider the hypothetical and/or actual costs of *not* using the technology.

Like risk, there can be a difference between the cost/benefit assessed by an expert and the cost/benefit perceived by the different stakeholders. Whilst there is a wealth of literature exploring the issues associated with how stakeholders perceive risk, there is very limited research on the factors influencing stakeholder perception of costs and benefits within a safety decision making context. This leads to a plethora of questions, for example:

1. What benefits and costs associated with UAS operations should be considered?
2. Will these need to change for different classes of UAS or UAS operations? E.g., defence UAS operations versus civil UAS operations, small "harmless" UAS versus large UAS.
3. For whom should the benefits and costs be measured? The operator? A member of the public exposed to the risk? Society in general?
4. What factors influence stakeholder perception of the costs and benefits and how should they be accounted for in the ALARP framework? Will the concept of gross disproportion change for different stakeholders or situations where there a significant differences between assessments and stakeholder perception of the cost/benefit?

Assessments of the costs and benefits and a measure of the uncertainty associated with the assessment will be needed in order to apply the ALARP framework. Most disconcerting is that none of the existing specifications of the ALARP framework provide substantive guidance on the management of uncertainty (i.e., its identification, representation, and consideration in decision making). This is a significant issue

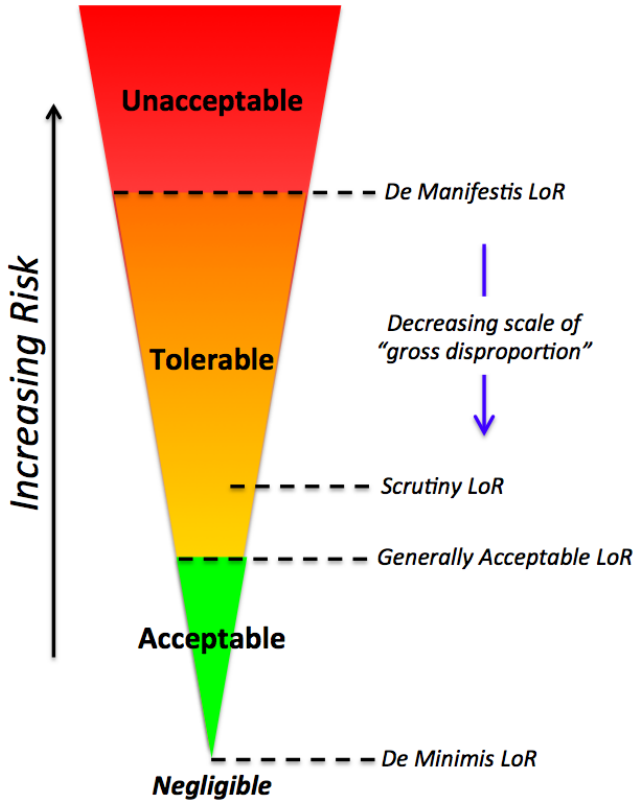

Figure 2: ALARP Risk Framework

for new technologies such as UAS where there is limited data and operational experience upon which to base assessments of the risk, cost and benefit. There are also uncertainties associated with the definition of the framework itself (e.g., the quantification of decision criteria within the framework).

3 Risk Criteria Within the ALARP Framework

High Level Safety Criteria (HLSC) governing the overall risk management and regulation of UAS operations have been defined. A detailed review of the different specifications can be found in Clothier & Walker (2013). An apparent consensus is that UAS, as a minimum, must demonstrate a level of safety equivalent to that of the safety performance currently demonstrated by Conventionally Piloted Aviation (CPA). This HLSC is commonly referred to as the Equivalent Level of Safety (ELoS) objective. In applying the ALARP framework to the safety risk management of UAS, a linkage between HLSC such as the ELoS objective needs to be established with the decision criteria defined within the ALARP framework.

The ELoS objective could be specified in relation to a number of decision criteria defined within the ALARP framework (Figure 2). These criteria include the *de manifestis*, *scrutiny*, *generally acceptable*, and *de minimis* criterion and the quantified concept of *gross disproportion*. In the next sections we explore the specification of the risk criteria within the ALARP framework. The specification of the *gross disproportion* criterion is the subject of a future research publication.

3.1 De Manifestis Risk Criteria

The concept of a *de manifestis LoR* stems from the legal definition of obvious risk and has been described as the LoR above which a person of ordinary level of intelligence intuitively recognises as being inherently unacceptable (Fulton 2002, RCC 2007). In the ALARP framework, the de manifestis LoR distinguishes an unacceptable LoR from a tolerable LoR (Figure 2).[2] It is important to note that demonstrating a LoR less than that of the de manifestis LoR is not sufficient for the risk to be deemed tolerable (refer to Section §2.2).

Except under extraordinary circumstances (refer to Section §2.1), de manifestis risk criteria can largely be considered as hard criteria (i.e., a rigid boundary on the decision space, one that must be satisfied). Therefore, de manifestis risk criteria essentially establish the minimum safety (or, equivalently, the maximum unsafe) expectation of civil UAS regardless of the potential benefit of the operation or whether the LoR can be practically achieved or not. One could logically conclude that the ELoS objective for UAS, which defines the overall *minimum* safety expectation for UAS operations, should be associated with de manifestis risk criteria within the ALARP framework. However, on deeper inspection, this association may not be appropriate.

Society tolerates the risk associated with CPA operations in return for the substantial and readily identifiable benefits of, for example, air transportation, aerial work and flying for recreation. However, society tolerates many activities with a higher LoR than that of passenger jet and turboprop fleet operations, which are referred to here as Conventional Airline CPA (CA-CPA). For example, it is widely recognised that passengers are more likely to die in a motor vehicle accident on their way to an airport than they are during their time on board a CA-CPA operation.

The level of Individual Risk (IR) of fatality associated with road accidents in the UK is estimated as 5.9×10^{-5} fatalities per member of the UK population per year (HSE 2001*b*). This LoR is two orders of magnitude greater than that associated with CA-CPA operations, which is estimated to be 2.3×10^{-7} fatalities per annum per worldwide population (Clothier et al. 2013). The point being made here is that society readily tolerates the risks associated with motor vehicles even though the LoR is greater than the LoR associated with CA-CPA transportation. It can be concluded that the LoR determined for CA-CPA operations does not reflect the critical LoR above which society broadly recognises as being unacceptable irrespective of its potential benefits (i.e., trade-offs between risk and benefit are still being made at a LoR two orders of magnitude greater than the LoR exhibited by CA-CPA operations). Further to this point, the HSE UK recommends a de manifestis IR of fatality of 1 in 1,000 per annum (1×10^{-3} per annum) for first parties and 1 in 10,000 per annum (1×10^{-4} per annum) for third parties (HSE 1992, 2001*b*). It is observed that these specifications of de manifestis criteria are more than two to three orders of magnitude greater than measures of risk determined for current worldwide CA-CPA operations.

The examples above indicate that society is willing to make a trade-off between the benefit, cost and risk for activities with a higher LoR than that currently exhibited by CA-CPA operations. The specification of the de manifestis HLSC for UAS should permit sim-

[2] Within the CASA framework, de manifestis criteria are indirectly referred to as basic safety limits (CASA 2009).

ilar trade-offs to be made for UAS for a comparable or higher LoR.

Associating the ELoS objective with de manifestis criteria within the ALARP framework would be inconsistent with the fundamental decision scenario that de manifestis criteria are meant to reflect (i.e., a demarcation between those situations where costs and benefits are not factored into the safety decision making process). Further, such an association would establish a minimum LoR requirement on UAS operations that is orders of magnitude above LoR already tolerated by society for other activities (including, for example, sport aviation operations or road transportation).

A more appropriate basis for specifying the de manifestis LoR for UAS are existing published regulatory limits or LoR determined from studies characterising the *upper limit* of public acceptability of manmade risks (e.g., CPA, power generation, etc.).

3.2 Generally Acceptable Risk Criteria

Unlike *de manifestis LoR* criteria, *generally acceptable LoR* criteria represent goal LoR. They are soft requirements that should be satisfied taking into consideration hard practical constraints on the available technology, on its operation and on the resources available to mitigate the risks.[3]

HSE guidelines suggest that the generally acceptable risk criteria should represent LoR comparable or lower than the background LoR members of the public are readily exposed to in day-to-day life (e.g., to annual risk of death due to a lightning strike). The HSE recommends the generally acceptable individual risk of fatality criterion of one in a million per annum (1×10^{-6} per annum) (HSE 2001*b*), stating that this LoR is extremely small compared to the typical background risk of fatality of one in a hundred per annum (1×10^{-2} per annum) averaged over a lifetime (HSE 2001*b*). There are a number of social and psychological factors that need to be taken into consideration when making a comparison between the risks associated with a technology (such as UAS) and the risks society readily accepts as an inescapable part of life. The key difference being that exposure to a technology is controllable. This difference and other social and psychological factors are discussed in Section §4.

An alternate approach for specifying generally acceptable risk criteria is through reference to LoR currently accepted in society for a similar technology or industry. In the context of UAS, the generally acceptable LoR could be specified in terms of the current LoR for CPA operations. The CPA LoR can vary significantly depending on the historical period of assessment and the type of CPA operation (Clothier & Walker 2006). For example, the measures of individual fatality risk due to a midair collision determined by Fulton et al. (2009) clearly show the variation between the LoR for the aviation sectors of general aviation, sport, regular passenger transport aviation. This variation in LoR also illustrates differences in society's appetite for risk and the different risk-cost-benefit-feasibility trade-offs that exist for different sectors of the CPA industry (e.g., acceptability of risks associated with sport aviation versus those associated with regular public transportation).

Society is becoming increasingly risk aware (Slovic 1987, Kates & Kasperson 1983, Slovic 1999). Further, the public believe they are exposed to more risks today than they were in the past and that these risks will continue to increase (Slovic 1999). With this increasing awareness comes increasing opposition to new sources of risk (Slovic 1999), particularly those associated with new technologies (Kates & Kasperson 1983). In addition, there is the increasing expectation for assurances in the safety of systems that were previously considered as of lower societal concern. This shifting social climate is a challenging environment for both the proponents of new technologies and the authorities charged with their regulation.

Therefore, when considering historical CPA LoR it is critical to understand that those LoR may only have been acceptable only in that period and would be considered very differently today. The acceptable LoR for aviation that bracketed the start of CPA at Kitty Hawk was higher than the acceptable LoR that existed for the USA space program, which is different to the acceptable LoR for today's regular public transportation and general aviation sectors. An argument that UAS as a new technology should be afforded a grace period with LoR similar to that of the early CPA era would not be accepted. Society has evolved generally acceptable risk criteria across the aviation sector that are independent of new technology introductions.

Finally, one must also note the difference between a goal LoR (i.e., what is aspired to or designed for) and the actual safety performance demonstrated by a system. A LoR based on historical CPA accident and incident data reflects the latter of these measures and not the goal safety performance for CPA. The demonstrated safety performance of CPA may far exceed or fall short of the goal LoR for CPA (i.e., a LoR deemed generally acceptable for the CPA). CPA accident and incident data may not provide a suitable basis for defining the *goal* LoR (i.e., the generally acceptable LoR) for UAS within an ALARP framework. It is more appropriate to specify generally acceptable risk criteria for UAS in relation to other criteria reflecting *goal* LoR. For example, it would be more appropriate to specify generally acceptable risk criteria for UAS through reference to the generally acceptable LoR (a goal LoR) used within an ALARP framework for a CPA category. If such criteria were not available, generally acceptable LoR as specified in the ALARP frameworks for other socio-technical risks could be used.

3.3 De Minimis Risk Criteria

The *de minimis LoR* stems from the legal principle de minimis non curat lex: "the law does not concern itself with trifles" (Fulton 2002, RCC 2007, Pate-Cornell 1994). The de minimis LoR can be used within the ALARP framework as a guide for determining when risks have been managed to a level that could be considered below general concern, i.e., that a LoR is approaching negligible risk. Like the generally acceptable LoR, the de minimis LoR is a goal LoR. The HSE UK proposes a de minimis risk of individual fatality of one in ten million per annum (1×10^{-7} fatalities per annum) (HSE 2001*b*).

In accordance with the CASA, HSE and CAA-UK ALARP frameworks, the requirement to continue to reduce the risks applies regardless of whether the LoR is considered generally acceptable or not. This is consistent with the overarching goal to continually pursue safety improvement in civil aviation. Therefore, it is not mandatory that de minimis risk criteria be defined in the ALARP framework for UAS as mechanisms to reduce risk should always be undertaken until a gross disproportion can be demonstrated.

[3] Within the CASA framework, the generally acceptable LoR is indirectly referred to as the basic safety objective (CASA 2009).

3.4 Scrutiny Risk Criteria

A reference or scrutiny level is sometimes used to put newly assessed risks in context with risks that have been tolerated or broadly accepted in the past (Clothier & Walker 2013). Scrutiny LoR are not decision criteria but points of reference that allow decision makers to contrast newly assessed risks against the management of similar or familiar risks. Scrutiny lines can exist anywhere in the tolerable or acceptable regions of the ALARP framework and are often specified in terms of LoR determined for a similar activity or industry.

The current levels of risk exhibited by CPA provide a good reference point against which to contrast the safety performance of civil UAS. UAS are a viable alternative to CPA in many applications. Such alternatives should be evaluated in accordance with the general principles of ALARP. Further, the media and members of the public are likely to use the current safety performance of CPA as a "litmus test" for UAS (Clothier & Walker 2013), thus a reference LoR may be a useful tool in communicating the risks to different stakeholder groups. For these reasons it is recommended that the ELoS objective be represented as scrutiny criteria within the ALARP decision making framework.

3.5 Impact of Risk Exposure

The preceding discussion on risk criteria has assumed an equal risk exposure level when comparing different criteria. This is a natural tendency resulting from biases such as "worst case thinking" and "availability heuristic" (Evans 2012), where risk is assessed from a perspective that the worst outcome will be more likely and that our measure of risk is biased by recalling recent similar experiences. The applicability of CPA risk thresholds to UAS may be an obvious social starting point, but a deeper assessment would reveal that the exposure of people and property is very much dependent on the specific mission. This point is clearly illustrated in the quantitative risk analyses conducted by Weibel (2005), Clothier & Walker (2006), Clothier et al. (2007), and Dalamagkidis et al. (2009), amongst others. For CPA, the same variability in exposure is not encountered as there is always at least one person exposed to the primary hazards (i.e., the pilot). Different criteria may need to be substantiated within ALARP framework for UAS.

3.6 Summary

The review of risk criteria within the context of the ALARP framework has identified that concepts such as *de manifestis LoR*, *generally acceptable LoR*, and *de minimis LoR* would adequately define the divides between unacceptable, tolerable, and acceptable risk classifications. However, moving beyond the conceptual application of the ALARP framework to UAS, reveals many complications in the actual specification of these criteria. This complexity is driven by the variability of risk tolerance in society both in time and via the inherent cost/benefit analysis undertaken for each hazard. This is further complicated when taking into account the impact of risk exposure, which indicates that the criteria are not static and that direct comparisons between particular CPA categories and UAS may not be appropriate.

4 Representing Societal Concern

Discussion thus far has focussed on the decision dimension of risk. As illustrated in Figure 1, the ALARP decision-making framework has an additional dimension describing societal concern. The dimension of societal concern reflects the degree of "socio-political response" (HSE 2001*b*) to the realisation of an hazard. It has been stated that the origin for societal concern is public aversion to certain characteristics of the hazards concerned (HSE 2001*a*). Some characteristic features attracting a higher degree of societal concern include:

1. Lack of familiarity of the hazardous activity/technology
2. Scale of the detrimental outcomes (e.g., multiple fatalities or widespread detriment)
3. Prolonged effects
4. Vulnerability of the people impacted by the hazard (e.g., children and the elderly)
5. Lack of equity of the distribution of risks or benefits associated in the activity
6. Involuntariness of exposure
7. Inspiration of dread

(HSE 2001*a*)

With the exception of the above, the existing literature provides very little guidance on how these characteristic 'features' can be measured or on how they can be incorporated into the ALARP framework (e.g., as criteria to be balanced alongside measures of risk). General guidance on the consideration of the dimension of societal concern specific to the safety risk management of civil aviation could not be identified in the literature. The following sections provide a brief exploration of these factors in the context of UAS.

4.1 Representing Societal Concern Due to Scale of Detrimental Outcomes

Societal concerns arising due to the occurrence of multiple fatalities in a single event can be characterised through a measure of the Societal Risk (SR). The representation of the ALARP framework using SR measures is provided in Figure 3. Horn et al. (2008) discussed some of the mathematical foundations of SR.

As illustrated in Figure 3, the risk decision criteria (e.g., de minimis, de manifestis criteria) are not represented as a single value but a function of the potential magnitude of loss for a single event. The barrier functions are monotonically-decreasing, reflecting society's apparent aversion towards those accidents with the potential for greater levels of harm (although there is some debate as to whether it is controllability or voluntariness that is causal for this difference in tolerability; see (Reid 2000)).

Society's perception of a LoR is largely driven by the nature and magnitude of the potential consequences, more so than the associated likelihoods of occurrence. Society tends to be more adverse towards those potential accidents that have a higher degree of loss. The magnitude of the potential loss primarily depends on the nature of the exposure relationship between the hazard and the population of people at risk. For example, accidents involving small general aviation aircraft tend to result in a smaller number

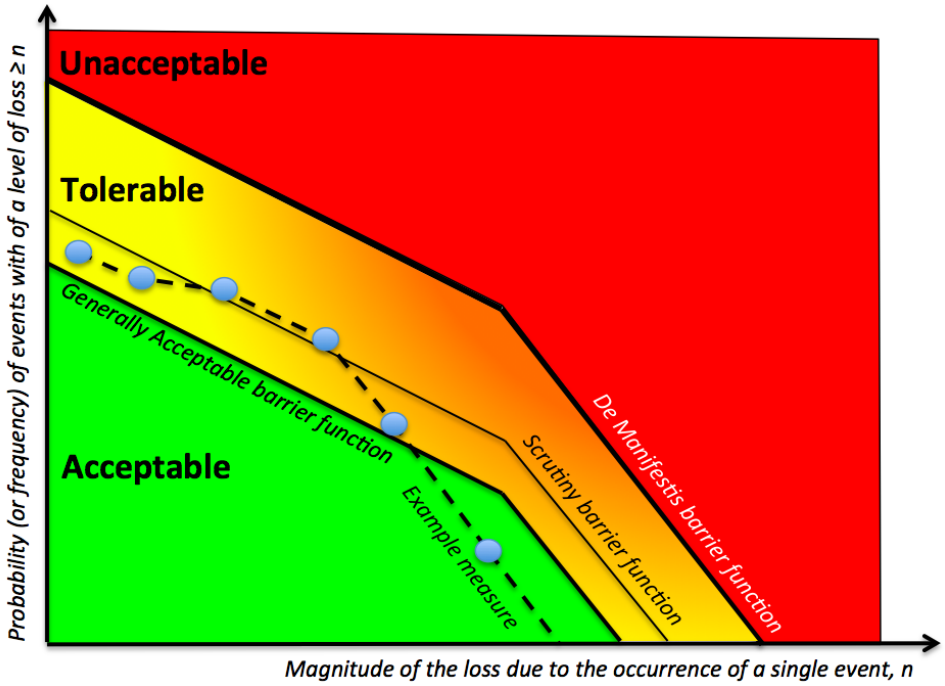

Figure 3: ALARP Risk Framework Represented Using Measures of Societal Risk

of fatalities than those accidents involving commercial passenger aircraft operations. This is because a smaller number of people are exposed given the occurrence of the hazard. Measures of SR provide an important tool for accounting for differences in the "risk portfolio" within an industry or aggregation of activities (Horn et al. 2008).

Safety criteria expressed in terms of SR are widely used in the regulation of a diverse range of industries (a review on the use of SR in the European Union for a range of industries can be found in Trbojevic (2005)). The HSE states that the "proper regulation of risks requires that both the individual risks and societal concerns engendered by a hazard must be addressed" (HSE 2001*b*).

The risk profile associated with UAS operations will be different to that associated with CPA. There are significant differences between the CPA and UAS fleets (Palmer & Clothier 2013). In particular, the significant diversity in the UAS fleet compared to that of the CPA fleet. Due to this diversity, the risk profile associated with the UAS fleet is likely to be very different from that of the CPA fleet. For example, almost 70% of the current UAS fleet have a maximum take-off mass of less than 150 kg (Palmer & Clothier 2013). Subsequently, the risk profile for the UAS fleet for the impact of ground collision is most likely to be characterised by accidents involving a smaller number of casualties. The SR profile associated with the hazard of a midair collision involving UAS will also be different to that of CPA. This is largely due to the absence of a population of people onboard the unmanned aircraft who would be exposed to the hazard of a midair collision. For these reasons, measures of the SR based on CPA accident and incident data should only be used to define scrutiny barrier functions within ALARP framework for UAS (e.g., for use as guidance only). Due to differences between classes within the UAS and CPA fleets (Palmer & Clothier 2013), the scrutiny barrier function for UAS should be based on a fleet-level aggregation of the safety performance of CPA.

4.2 Representation of Other Aspects of Societal Concern

Measures of SR do not provide a comprehensive representation of the dimension of societal concern. Other factors influencing societal concern, as mentioned in §4, are not taken into account (e.g., familiarity, dread, vulnerability, etc.) by measures of SR.

Some aspects are indirectly captured within the existing safety risk management framework. The aspects of controllability of exposure, voluntary and involuntary exposure, and the distribution of benefits to those exposed are indirectly taken into account when making the distinction between the risks posed to first, second and third parties. More stringent safety criteria are typically imposed for those hazards that pose a risk to people whom have limited controllability over their own exposure to the hazard, are involuntarily exposed, or to those who receive no immediate or readily perceived benefit from the hazardous activity.

The HSE makes the distinction between workers (first parties) and members of the public (third parties), proposing different safety criteria for each as summarised in Table 1. Similar distinctions are made in the management of safety on defence ranges in the USA (RCC 2007). In the context of UAS operations, first parties can be identified as those people directly associated with the operation of the unmanned aircraft, i.e., remote pilots and field personnel. Third parties are those people over flown by the UAS who have no direct connection to the UAS operation (e.g., members of the public). In the Common Risk Management Framework for Airspace and Air Traffic Management in Australia (DOIT 2012) it is stated that "safety criteria must be premised on the basis of the effect on aircrew, other safety critical staff, the travelling public and the community". As evident in this statement, there is the additional category of people at risk that must be considered in the risk management of aviation hazards, secondary parties. In the context of UAS, secondary parties would be those who are somehow involved or receive benefit

through aviation operations but are not directly associated with the operation the UAS. An example of secondary parties would be pilots, crew and passengers on board other aircraft who accept some level of risk in return for certain benefits (e.g., employment, transportation, etc.).

It is important to note that for UAS, the primary risks are to second and third parties, whereas for CPA the primary risks are to first and second parties (Clothier & Walker 2013). The difference in the populations at risk must be taken into consideration when specifying safety criteria for UAS. For example, it would be inappropriate to use existing safety criteria defined for risks to CPA first parties as a basis for defining safety criteria for UAS.

An accident can have an impact beyond that of injury to people. Some of these broader losses are indirectly captured in the assessment of the costs as part of the cost-benefit analysis undertaken for a determination of gross disproportion. Specifically, ICAO states that the following cost factors should be considered as part of the cost-benefit-analysis process: loss of business, equipment, productivity, managerial, legal, cultural, market, political and public (ICAO 2009). For UAS, loss of the unmanned aircraft, damage to the environment or property, the loss in earnings and loss of public or client confidence may be particularly significant for smaller UAS where the LoR to people is low. Such aspects would be considered as a secondary concern in the risk management of CPA. The potential impact on the efficiency of the existing airspace system is also important consideration.

It is important to note that the scope and magnitude of loss associated with an UAS accident, particularly in the early phases of their operation, can be amplified. The concept of the social amplification of risk is described in Kasperson et al. (1993).

4.3 Differences in Social Concern between Manned and Unmanned Aviation

There will be differences in stakeholder perceptions of the risks associated with the operation of UAS compared to that of CPA. This will in turn affect the acceptability/tolerability of the risks and consequently, the specification of safety criteria for UAS within the ALARP framework. Clothier & Walker (2013) describe a number of factors that may lead to differences in the perception and in turn acceptability of UAS operations compared to that of CPA operations. These factors include the visibility of the benefits, voluntariness, control, and the knowledge and information available to stakeholders. Other general factors that can influence public perception of risks are described in Slovic (1987). Directly adopting existing risk criteria (e.g., generally acceptable criteria) for CPA may not reflect differences in stakeholder perceptions and preferences in relation to the risks associated with UAS technologies.

As discussed previously, it is important to note that stakeholder perception and preferences towards UAS technologies will also change with time. It is likely that stakeholders will have a heightened sensitivity towards the risks while UAS remain a new and unfamiliar technology. This is a common situation for new technologies, as described by Melchers (2001):

> *History suggests that a new technology will only survive if it has no major catastrophes early in its development.*

Community attitudes towards the safety of UAS technologies are likely to change as stakeholders become more accustomed with the technology, more familiar with its associated risks and benefits, and as more information becomes available to regulators (i.e., trust). Further, the nature of the risks associated with the UAS industry will change as the sector grows and as new applications for the technology are identified and exploited.

The specification of safety criteria for UAS will need to reflect the initial sensitivity of stakeholders to new and unfamiliar technologies and the changing nature of the risks presented by UAS. Safety criteria will also need to reflect the objective for continued safety improvement as stipulated in the ICAO States' Safety Plan (SSP) (ICAO 2009). It can be concluded that the substantiation of the ALARP framework for UAS will need to be periodically revised to account for differences in stakeholder perceptions and preferences, changes in the risk profile associated with UAS operations, and to satisfy the objective for continued safety improvement as defined in the SSP.

Finally, different stakeholders will have different concerns. It cannot be assumed that the set of safety criteria defined in the ALARP framework are necessarily representative of all stakeholder perspectives.

4.4 Summary of Societal Concern

This section has only briefly touched on the dimension of societal concern within the ALARP framework. The most significant finding of this review is that the literature provides very little guidance as to this dimension and its incorporation into the ALARP framework.

With the exception of some high level discussion presented in Clothier & Walker (2013), no existing literature could be identified which specifically addressed the impact of societal concerns on the safety risk management and in turn regulation of UAS technologies. Addressing societal concerns will be a significant factor in the safety risk management and decision making for UAS, particularly in the early phases of the introduction of the technology. A number of factors influencing societal concern have already been identified but there is almost no guidance as to how these factors can be taken into consideration in the ALARP framework. Measures of SR and accounting for differences in the voluntariness of the exposure do not account for all of these factors. Further research to identify, characterise and incorporate societal concerns in the risk management of UAS is needed.

5 Conclusions

This paper has set out to identify and explore the issues of applying the As Low As Reasonably Practicable (ALARP) decision-making framework to the risk management of UAS. It was found that there are subtle differences between the different specifications of the ALARP framework made in safety literature. Inconsistencies in the existing frameworks were also identified. These subtle differences and inconsistencies can have significant impact on how the ALARP frameworks are to be substantiated for the risk management of UAS. A single, consolidated framework, should be adopted by the aviation safety community.

A conceptual application of the ALARP framework can be made using *de manifestis LoR*, *generally acceptable LoR*, and *de minimis LoR* as the boundary definitions between *Unacceptable*, *Tolerable*, *Acceptable*, and *Negligible*. However, significant difficulties were identified in the substantiation of the ALARP framework. In particular, in the specification of the

Decision Criteria	Population	Individual Risk of Fatality per Annum
De manifestis	Workers	1×10^{-3}
De manifestis	Members of the public	1×10^{-4}
Broadly acceptable	Workers and members of the public	1×10^{-6}

Table 1: Individual Risk (IR) of fatality criteria for different populations at risk (HSE 2001*b*)

ALARP decision criteria and the identification, characterisation, and representation of societal concerns.

Difficulties in relation to the specification of the ALARP decision criteria arise due to a number of factors, including:

1. Differences in the primary populations at risk, which in turn creates a difference in the nature of the exposure and the acceptability of the risks.
2. Unique systems and missions for which there are no CPA equivalents.
3. Differences in how society perceives the risks and benefits associated with new technologies such as UAS compared to that for established and familiar technology such as CPA. This in turn influences society's appetite for risk.
4. The time sensitivity of the acceptability of the risks associated with new technologies.
5. Additional uncertainties that need to be incorporated in the cost benefit analysis for UAS (to support a finding of gross disproportion)

Guidance as to how the dimensions of societal concern and levels of risk can be jointly considered within the ALARP framework could not be identified in the literature. In the case of new technologies, such as UAS, the dimension of societal concern can be as significant a factor in decision making as that of the quantified measures of the risk. Further research on the impact of social dimensions on risk thresholds, beyond the quantification of risk, is required. Decision making requires not only an appreciation of the risk but how individuals and society respond to that risk.

Finally, none of the existing specifications of the ALARP framework provide substantive guidance on the management of uncertainty. This is a significant issue for new technologies such as UAS where there is limited data and operational experience upon which to base assessments of the risk, cost and benefit.

In considering these deficiencies it is concluded that there are significant difficulties in the application and substantiation of the ALARP framework to the risk management of new technologies such as UAS.

6 Acknowledgement

The research presented in this paper was conducted under Project ResQu led by the Australian Research Centre for Aerospace Automation (ARCAA). The authors gratefully acknowledge the support of ARCAA and the project partners, Queensland University of Technology (QUT); Commonwealth Scientific and Industrial Research Organisation (CSIRO); Queensland State Government Department of Science, Information Technology, Innovation and the Arts; Boeing and Insitu Pacific.

References

CAA (2010*a*), CAP760 guidance on the conduct of hazard identification, risk assessment and the production of safety cases. for aerodrome operators and air traffic service providers., Technical report, Safety Regulation Group, Civil Aviation Authority, Norwich, United Kingdom.

CAA (2010*b*), Safety management systems - guidance to organisations, Technical report, Civil Aviation Authority, London, UK.

CASA (2009), CAAP SMS-1(0) safety management systems for regular public transport operations, civil aviation advisory publication, Technical report, Civil Aviation Safety Authority, Canberra, Australia.

CASA (2010), Cost benefit analysis procedures manual, Technical report, Civil Aviation Safety Authority, Canberra, Australia.

CASA (2012), Booklet three - safety risk management. SMS for aviation–a practical guide, safety management systems (SMS) resource kit., Technical report, Civil Aviation Safety Authority, Canberra, Australia.

Clothier, R. A., Lin, X. & Fulton, N. L. (2013), Quantification of high level safety objectives for civil unmanned aircraft systems, Technical Report EP13967, Division of Mathematics, Informatics and Statistics, Commonwealth Scientific and Industrial Research Organisation (CSIRO).

Clothier, R. A., Palmer, J. L., Walker, R. A. & Fulton, N. L. (2010), Definition of airworthiness categories for civil unmanned aircraft systems (UAS), *in* '27th Congress of the International Council of the Aeronautical Sciences (ICAS 2010)', Nice, France.

Clothier, R. A. & Walker, R. A. (2006), Determination and evaluation of UAV safety objectives, *in* '21st International Unmanned Air Vehicle Systems Conference', Bristol, UK.

Clothier, R. A. & Walker, R. A. (2013), *Safety Risk Management of Unmanned Aircraft*, Springer Science and Business Media B.V., Dordrecht, Netherlands.

Clothier, R. A., Walker, R. A., Fulton, N. L. & Campbell, D. A. (2007), A casualty risk analysis for unmanned aerial system (UAS) operations over inhabited areas, *in* 'Twelfth Australian International Aerospace Congress Conference, 2nd Australasian Unmanned Air Vehicles Conference', Melbourne, Australia.

Dalamagkidis, K., Valavanis, K. & Piegl, L. (2009), *On Integrating Unmanned Aircraft Systems into the*

National Airspace System. Issues, Challenges, Operational Restrictions, Certification, and Recommendations, Vol. 36 of *International Series on Intelligent Systems, Control, and Automation: Science and Engineering*, Springer Science and Business Media B.V.

DOIT (2012), Common risk management framework for airspace and air traffic management, Technical report, Department of Infrastructure and Transport Department of Defence, Civil Aviation Safety Authority, Airservices Australia, Canberra, Australia.

Evans, D. (2012), *Risk Intelligence: How to Live With Uncertainty*, Atlantic Books.

Fraser, C. S. R. & Donnithorne-Tait, D. (2011), An approach to the classification of unmanned aircraft, *in* 'Proceedings of the 26th International Conference on Unmanned Aerial Vehicle Systems (UAVS) 2011', Curran Associates, Red Hook, USA, pp. 157–211.

Fulton, N. L. (2002), Regional airspace design: A structured systems engineering approach, PhD thesis, School of Aerospace and Mechanical Engineering, the University of New South Wales, Canberra, Australia.

Fulton, N. L., Westcott, M. & Emery, S. (2009), 'Decision support for risk assessment of midair collisions via population-based measures', *Transportation Research Part A: Policy and Practice* **43**(2), 150–169.

Horn, M. E., Fulton, N. & Westcott, M. (2008), 'Measures of societal risk and their potential use in civil aviation', *Risk Analysis* **28**(6), 1711–1726.

HSE (1992), The tolerability of risk from nuclear power stations, Technical report, Health and Safety Executive (HSE).

HSE (2001*a*), 'Principles and guidelines to assist HSE in its judgements that duty-holders have reduced risk as low as reasonably practicable'. **URL:** *http://www.hse.gov.uk/risk/theory/alarp1.htm*

HSE (2001*b*), Reducing risks, protecting people. HSE's decision-making process, Technical report, Health and Safety Executive, Norwich, United Kingdom.

ICAO (2009), Safety management manual (SMM), doc 9859, Technical report, International Civil Aviation Organization (ICAO), Montréal, Canada.

Johansen, I. L. (2009), Foundations and fallacies of risk acceptance criteria, Technical Report ROSS (NTNU) 201001, Norwegian University of Science and Technology, Trondheim, Norway.

Jones-Lee, M. & Aven, T. (2011), 'ALARP—what does it really mean?', *Reliability Engineering & System Safety* **96**(8), 877–882.

Kasperson, R. E., Renn, O., Slovic, P., Brown, H., Emel, J., Goble, R., Kasperson, J. X. & Ratick, S. (1993), 'The social amplification of risk: A conceptual framework', *Risk Analysis* **8**(2), 177–187.

Kates, R. & Kasperson, J. (1983), Comparative risk analysis of technological hazards, *in* 'Proceedings of the National Academy of Sciences of the United States of America', Vol. 80, pp. 7027–7038.

Magister, T. (2010), 'The small unmanned aircraft blunt criterion based injury potential estimation', *Safety Science* **48**(10), 1313–1320.

Melchers, R. E. (2001), 'On the ALARP approach to risk management', *Reliability Engineering and System Safety* **71**(2), 201–208.

Palmer, J. L. & Clothier, R. A. (2013), Analysis of the applicability of existing airworthiness classification schemes to the unmanned aircraft fleet, *in* '15th Australian International Aerospace Congress (AIAC 15)', Melbourne, Australia.

Pate-Cornell, E. (1994), 'Quantitative safety goals for risk management of industrial facilities', *Structural Safety* **13**(3), 145–157.

RCC (2007), Standard 321-07: Common risk criteria for national test ranges, Technical report, Safety Group Risk Committee, Range Commanders Council, US Army White Sands Missile Range, New Mexico.

Reid, S. G. (2000), 'Acceptable risk criteria', *Progress in Structural Engineering and Materials* **2**(2), 254–262.

Slovic, P. (1987), 'Perception of risk', *Science* **236**(4799), 280–285.

Slovic, P. (1999), 'Trust, emotion, sex, politics, and science: Surveying the risk-assessment battlefield', *Risk Analysis* **19**(4), 689–701.

Trbojevic, V. M. (2005), Risk criteria in EU, *in* 'European Safety and Reliability Conference ESREL', European Safety and Reliability Association, Tri City, Poland.

Weibel, R. E. (2005), Safety considerations for operation of different classes of unmanned aerial vehicles in the national airspace system, Masters thesis, Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA, USA.

A Review of the Concept of Autonomy in the Context of the Safety Regulation of Civil Unmanned Aircraft Systems

Reece A. Clothier[1] **Brendan Williams**[2]* **Tristan Perez**[3]

[1] School of Aerospace, Mechanical, and Manufacturing Engineering
RMIT University,
PO Box 71, Bundoora, Victoria 3083, Australia
Email: reece.clothier@rmit.edu.au

[2] Australian Research Centre for Aerospace Automation
Queensland University of Technology,
22-24 Boronia Road, Brisbane Airport, Queensland 4008, Australia
Email: bp.williams@qut.edu.au

[3] School of Engineering
The University of Newcastle,
Callaghan, New South Wales 2308, Australia
Email: tristan.perez@newcastle.edu.au

Abstract

Civil aviation safety regulations and guidance material classify Unmanned Aircraft Systems (UAS) as either Remotely-Piloted Aircraft Systems (RPAS) or Autonomous Aircraft Systems (AAS). This distinction is based on the premise that the effective safety risk management of UAS is dependent on the degree of autonomy of the system being operated. However, it is found that there is no consensus on the concept of autonomy, on how it can be measured, or on the nature of the relationship between Levels of Autonomy (LoA) and the safety-performance of UAS operations.

An objective of this paper is to evaluate existing LoA assessment frameworks for application in aviation safety regulations for UAS. The results from a comprehensive review of existing concepts of autonomy and frameworks for assessing LoA are presented. Six case study UAS were classified using the published LoA frameworks. The implied LoA of UAS for existing modes of operation (e.g., teleoperation, semi-autonomous) were also assessed using the published frameworks.

It was found that the existing LoA assessment frameworks, when applied to the case study UAS, do not provide a consistent basis for distinguishing between the regulatory classes of RPAS and AAS. It was also found that the existing regulatory definition of an autonomous aircraft is too broad, covering UAS of significantly different levels of capability and system complexity. Within the context of aviation safety regulations, a new LoA assessment framework for UAS is required.

Keywords: Autonomy, Unmanned Aircraft Systems, UAS, Regulation

1 Introduction

A classification scheme establishes the foundation for a regulatory framework for civil Unmanned Aircraft Systems (UAS). The components of the International Civil Aviation Organization (ICAO) regulatory classification framework relevant to the regulation of civil UAS are shown in Figure 1. ICAO (2011) defines an UAS as "an aircraft and its associated elements which are operated with no pilot on board." As illustrated by Point A in Figure 1, ICAO further classifies UAS as being either Remotely-Piloted Aircraft Systems (RPAS) or Autonomous Aircraft Systems (AAS)[1]. Where a Remotely-Piloted Aircraft (RPA), a component of the RPAS, is defined as:

> *An aircraft where the flying pilot is not on board the aircraft.*

and, an *Autonomous Aircraft* as:

> *An unmanned aircraft that does not allow pilot intervention in the management of the flight.*

Implicit to this regulatory distinction is the premise that the effective safety risk management of UAS is dependent on the degree of autonomy of the system being operated. However, there is no consensus on the meaning of "autonomy", on how it can be measured, or on the nature of the relationship between levels of autonomy and the safety-performance of UAS operations. Given the foundational role classification plays in the development of a regulatory framework, further clarity on the meaning and definition of autonomy is needed in order to progress the development of regulations and standards for UAS.

This paper presents a review of existing concepts of autonomy with the view to evaluate their use in the refinement of existing regulatory classifications of

* Mr BP Williams is on secondment to the Queensland University of Technology from Boeing Research & Technology - Australia (Email: brendan.p.williams@boeing.com).

[1] ICAO does not explicitly define Autonomous Aircraft System (AAS). A definition of an *Autonomous Aircraft*, a component of an AAS, is provided implying the existence of the AAS class of UAS.

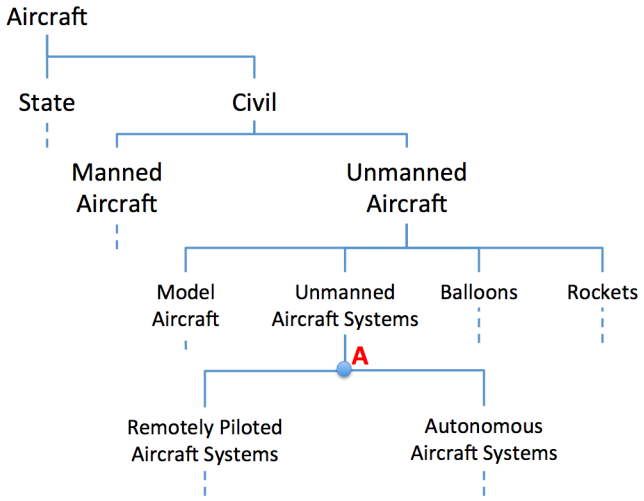

Figure 1: ICAO Regulatory Classification Framework (focused on UAS component) (ICAO 2011)

UAS. A comprehensive review of the literature is undertaken. The review identifies a number of definitions for the general concept of autonomy. The differences between these definitions and those of other related concepts are discussed in Section §2. The literature review also reveals numerous possible frameworks for assessing the degree or Level of Autonomy (LoA) of a system (presented in Section §3). In Section §4, the assessment frameworks are applied to six case study UAS. The LoA determined for the case study UAS are then compared against the LoA corresponding to the ICAO definition of an *Autonomous Aircraft*.

ICAO (2011) also defines an *Autonomous Operation*. A review of the literature reveals numerous definitions of the UAS modes of operation. The modes imply different LoA in the system being operated. The modes of operation and their corresponding LoA are briefly reviewed in Section §4.3.

Section §5 provides discussion and recommendations on a LoA framework for use in the regulation of UAS. A summary of the key outcomes from the study are presented in Section §6.

2 General Concepts of Autonomy

We start by exploring the common meaning of the concepts of *autonomy*, *autonomous* and the related notions of *automatic* and *automated*. The Oxford English Dictionary defines autonomy as being *"the right or condition of self-government; freedom from external control or influence; independence."* and autonomous as *"having the freedom to govern itself or control its own affairs; having the freedom to act independently."* (Stevenson 2012)

On first inspection these definitions convey a simple concept of autonomy/autonomous, that which can be defined entirely by the degree of dependency or interaction between two entities (e.g., the human Remote Pilot (RP) and an UAS).

The higher the LoA of the UAS, the less the human RP is involved in the operation and the more the UAS subsumes the role of the human RP. It is important to recognise that in order to maintain system performance, as the degree of independence increases, so too does the need for the UAS to exhibit the more complex properties that were previously provided by the human RP (e.g., the ability of the UAS to perceive its environment, make decisions, and to modify its behaviour and change its goals accordingly, etc.). Thus, the higher the degree of independence, the higher the implied complexity of the UAS. Ultimately, if the performance of the Human-Machine System (HMS) is to remain the same, then full autonomy would imply that the UAS is capable of performing all of the functions traditionally provided by a RP.

This leads us to the second and more common concept of autonomy, that which makes the relationship between increasing independence and the increasing complexity of the machine explicit (e.g., see definitions summarised in Table 1 of Appendix A). Evident in these definitions of autonomy are the properties of higher-order systems in addition to the degree of freedom or independence from external influence or control. Such properties can only exist in more complex systems; those systems residing in the higher tiers of Boulding's General Hierarchy of Systems (Boulding 1956). Put differently, these definitions have associated the concept of autonomy with complex-system properties such as intelligence, self-awareness, adaption, and cognition etc., which are properties not present in low-complexity systems (i.e., those classified as belonging to the tiers of frameworks, clockworks, or thermostats in Boulding's Hierarchy).

This provides some explanation for the ongoing difficulties encountered when defining autonomy. There are two camps. The first considers autonomy as being completely described by the property of independence between two entities (e.g., Clough (2002)). This property can be identified in systems residing at any tier in Boulding's General Hierarchy of Systems. The second camp makes explicit the relationship between increasing autonomy (i.e., independence) and the increasing complexity required of the machine in order to maintain a desired level of performance (i.e., the increasing need for the machine to resemble "people"). This second notion of autonomy is not as clearly defined, having been associated with more complex and equally debated system properties (e.g., system intelligence).

Finally, it is worth noting the difference between autonomous and automatic. The Oxford English Dictionary defines automatic as *"(of a device or process) working by itself with little or no direct human control."* (Stevenson 2012) Definitions of automated/automatic identified in the broader literature are summarised in Table 1. There is little distinguishing the notion of automatic from that of the simpler notion of autonomous. A fully autonomous UAS (i.e., one that is entirely independent of human influence or control) could also be identified as automatic, and vice versa. Take, for example, a wristwatch. A wristwatch can be described as automatic. Under the simpler notion of autonomy, the wristwatch can be determined as having a high LoA as it performs its task (measuring and display of time) largely independent of the person wearing it. Whereas, using the more complex notion of autonomy the automatic wristwatch is likely to be assigned a very low LoA (a wristwatch is not intelligent or adaptive, etc.).

The authors agree with Clough (2002) who advocate that the property of autonomy is distinct from other complex properties such as system intelligence. However, the authors recognise that in order to maintain system performance at increasing LoA, the degree of intelligence, adaptability, etc. of the machine must also increase. A pragmatic framework for assessing the LoA of a system may need to take into consideration these additional relationships. This appears to be the (implicit) position adopted by the majority of the assessment frameworks identified in

the literature.

3 Assessing Levels of Autonomy

Numerous frameworks for assessing the LoA of a system have been proposed. A summary of the more common frameworks and those that have been previously applied to UAS is provided in Table 2 of Appendix A. Some of the key points of difference between these frameworks include:

1. *Measurement Scales* - LoA have been expressed using ordinal, interval and ratio measurement scales (refer to Stevens (1946) for definitions of scales). The majority of existing frameworks measure a LoA on an ordinal scale. Measures of the difference between LoA defined on an ordinal scale are not meaningful. All that can be ascertained is whether the LoA of one system is less than, equal to, or greater than the LoA of another system. Interval scales permit measures of the magnitude of the difference between differing LoA, however this difference has no 'absolute' reference point. Measurements made on a ratio scale permit assessments of the multiplicative difference between the LoA of different systems (e.g., a system has twice the autonomy of that of another).
2. *Number of Levels* - The number of discrete LoA proposed for ordinal scales ranges from four to twelve (refer to Table 2). The literature provides limited justification for the number of levels proposed.
3. *Component Properties* - The assessment frameworks describe a LoA using a range of properties of the HMS. For example, the degree of human/operator control or interaction with the UAS, the allocation or performance of certain functions or tasks to the human or the machine, or the complexity of the mission or environment in which the HMS operates. Differences in the properties used to assess autonomy arise due to differences in the underlying concepts of autonomy and differences in the context in which the assessment frameworks are used. For example, Clough (2002) defines a LoA assessment framework for military UAS. The proposed LoA assessment framework includes factors such as battle space cognisance, targeting and multi-vehicle coordination. On the other hand, the LoA framework proposed by Billings (1991) was primarily intended for the assessment of cockpit automation and hence the LoA are described in relation to the degree of pilot control and management of a single aircraft.
4. *Measurement* - Two approaches have been used to determine the measure of the LoA:
   (a) A single measure of the LoA of a system is determined from the mathematical 'combination' or 'mapping' of the independent properties used to characterise autonomy. For example, Kendoul (2011) determines a LoA for an UAS by combining independent measures of the guidance, navigation and control capabilities of the UAS. Methods for aggregating the component measures into a single LoA include addition or weighted linear combination.
   (b) The LoA is assessed in relation to independent functions or contexts without aggregation into a single LoA for the system. For example, Parasuraman et al. (2000) measure a LoA in terms of four "broad classes" of functions. The four measures are not aggregated to provide an assessment of the overall LoA of the system.

The above differences make comparisons between the various autonomy assessment frameworks difficult.

4 Evaluation of Assessment Frameworks

Six generic case study UAS are defined. The case study UAS serve as test points for evaluating the existing LoA assessment frameworks. Four of the case study UAS describe systems in use today, with the final two UAS describing systems with capabilities that only currently exist in science fiction[2]. The six generic configurations are:

> **UAS A:** An UAS where the Unmanned Aircraft (UA) is capable of executing a pre-programmed behaviour (e.g., following a series of waypoints) and where it is not possible for the Remote Pilot (RP) to interact with the UA after launch except possibly where to terminate flight. There is significant human interaction in the mission planning phase, but no interaction is possible after launch. An example of this type of UAS are early variants of the Ryan Lightning Bug, a reconnaissance UAS that once launched followed a pre-programmed route without interaction with a RP.
>
> **UAS B:** An UAS where the operation of the UAS requires continual input from, or interaction with, a RP. For example, an UA where the flight control surfaces are manipulated by the RP via a data link.
>
> **UAS C:** An UAS where the UA, under normal operating conditions, does not require continual input from, or interaction with, a RP to perform its mission. The RP may or may not interact with the UA during the mission but interaction is possible. The RP continuously monitors the status and performance of the UAS. The behaviour of the UAS can be pre-programmed before the flight or updated by the RP during the flight. RP interactions are predominantly high-level and in the form of deviations to pre-programmed behaviour (e.g., new waypoints) due to changes in the mission objectives, failures in the system or changes in the environment. The behaviour of the UAS is deterministic. The majority of current UAS would be described by this case study system.
>
> **UAS D:** An UAS with all the capabilities of case study type 'C' UAS with the additional capability of the UAS being able to change its behaviour in response to changes in its environment or performance. Given predefined goals the UAS will determine

[2]The films identified by the authors are for illustrative assistance in understanding the described configuration and makes no comment in relation to the quality of the film.

how to achieve those goals within the constraints defined by the RP or system designer. Constraints on the behaviour of the UAS are static and hard (i.e., they cannot be changed or breached by the UAS). Given the goals and constraints the behaviour of the UAS can be non-deterministic, although the bounds on the behaviour are always known. Under normal operating conditions, the RP has the ability to override decisions made by the UAS. This configuration is in the R&D stage with possible systems in flight test.

UAS E: An UAS analogous to the fictitious "HAL 9000 computer" portrayed in the novel and screenplay "2001: A Space Odyssey" (Clarke 1968). Recall the words from HAL "I'm sorry, Dave. I'm afraid I can't do that." An UAS of this type has all the capabilities of case study 'D' UAS with the additional capability of the UAS being able to override or deny the inputs of the RP. The UAS maintains the initial goals and remains within predefined constraints on its behaviour (as determined by the system designer or programmed before the mission). The ability for the RP to change the goals and/or constraints of the UAS is lost or impaired.

UAS F: An UAS analogous to the fictitious "Extreme Deep Invader (EDI)" portrayed in the Columbia Pictures film "Stealth". For this type of UAS, the UAS performs all of the functions that a RP would and like type 'E' UAS, can operate without the need for interaction with the RP. Interaction between the UAS and the RP is possible but only if the UAS so chooses. The UAS may change its goals and/or constraints independent of the RP.

The autonomy of an UAS can change for different phases of flight (e.g., taxi, take-off, landing, etc), for different functions within its mission (e.g., weapons release) or during emergency situations (e.g., loss of communications between the ground control station and the UA). Systems with such a capability are said to exhibit "adaptive autonomy", which adds another dimension of complexity to the regulation of UAS. For this paper, the case study UAS are assumed to exhibit only one LoA.

4.1 LoA of the Case Study UAS

The LoA is assessed for each of the case study UAS and the ICAO concept of an *Autonomous Aircraft*. Only those frameworks that provided a single measure of the LoA on an ordinal scale were evaluated. In order to explore the consistency of the assessment across different assessors, each author independently assessed the LoA of the case study UAS. The assessment was performed twice by each author to explore the repeatability/consistency of the assessment on an individual assessor basis. The consolidated results from all authors are presented in Table 3 of Appendix A.

The frameworks use different labelling conventions for their scales. For ease of comparison Roman numerals are used for all LoA assessments. The class of lowest LoA is assigned the numeral 'I', the next class of higher LoA assigned the numeral 'II', and so on. A '-' in Table 3 is used to indicate those instances where the LoA of the case study UAS could not be determined using the framework. Where a case study UAS could be assessed as having more than one LoA, all possible levels are indicated. An 'NC' in Table 3 indicates those cases where consensus between the authors on the assessment of the LoA could not be reached.

It should be noted that despite the use of a common labelling scheme, a direct comparison of the LoA determined using different assessment frameworks is not meaningful. For example, a LoA of 'IX' assessed for one framework is not necessarily equivalent to a LoA of 'IX' assessed using another framework. Nor is it meaningful to measure the difference between two LoA on any scale (refer to the Stevens (1946) for further discussion on the limitations of measurements made on ordinal scales). However, meaningful observations can be made in relation to how the different frameworks rank/order the LoA of the case study UAS.

4.2 Analysis of Results

With reference to Table 3, it can be observed that the greatest variation is in the assessment of the LoA of *UAS A*. Under some frameworks *UAS A* is assigned a very low LoA but in others it is assigned a very high LoA (i.e., comparable to the LoA assigned to *UAS E* and *F*). Variability in the LoA assigned for *UAS A* arises due to the different concept of autonomy adopted by each framework. *UAS A* is assigned a high LoA in those frameworks that base their concept of autonomy on the degree of independence between the UAS and the RP. This is because *UAS A* can perform its mission without interaction with a human RP. Conversely, *UAS A* is assigned a low LoA in those frameworks that adopt the more complex notion of autonomy as the UAS is not capable of higher-order functions such as decision making or the ability to change its own behaviour in response to changing conditions.

There are a number of cases where a consensus between the authors on the LoA could not be reached (i.e. the 'NCs' in Table 3). This most commonly occurred in the assessment of the LoA for *UAS C* and *D*. It was found that the majority of the inconsistencies between repeated assessments were due to the limited guidance available on how to make assessments using the proposed LoA schemes. The greatest inconsistency between assessments occurred for LoA frameworks that used independent metrics but provided no guidance as to how the independent measures were to be combined. For example, using the framework proposed by Kendoul (2011), some UAS were assessed as having a high LoA in some functions (e.g., control) but low in others (e.g., guidance and navigation). Limited guidance was provided on how to combine the three independent measures into a single LoA for the system. Kendoul (2011) uses '+' and '-' notation to "distinguish between a system that has accomplished that AL [autonomy level] or TRL [technology readiness level] by satisfying all its requirements, and a system that satisfies some of the requirements only." A description on how the overall LoA assignment is made, e.g., whether the system is a 'VII+' or 'VIII-', is not provided. Often the case study UAS satisfied some of the conditions for assignment to a particular LoA but not all. From the description provided it was unclear as to whether the independent measurement dimensions were necessary or sufficient conditions for the assignment to a particular LoA or if other means should be used to perform

the aggregation of independent measurement dimensions (e.g., minimum, maximum, majority, etc.).

Another factor contributing to the inconsistency in the assessments was that there was insufficient information on the case study UAS. Some of the assessment frameworks required detailed information about the UAS (e.g., cognisance of surroundings, % of interaction time with the RP, whether the UAS was reliant on the Global Positioning System or could perform collision avoidance, etc.). Where such information was unavailable an assessor could either indicate that the LoA of the case study UAS could not be determined or make assumptions about the case study UAS. The latter situation contributes to the subjectivity of the assessment process and in turn the potential for inconsistency in the results.

There are many instances in Table 3 where it was not possible to assign a single LoA to the case study UAS. There were also a number of cases where the LoA determined for the different case study UAS overlapped. Overlap in the assessments of LoA was most frequently encountered between *UAS C* and *UAS D*, and between *UAS E*, *UAS F* and *Autonomous Aircraft*. From a regulatory standpoint, a clear distinction between the case study UAS on the basis of their LoA would be required. The higher the LoA of the UAS, the less the human RP is involved in the operation and the more the UAS subsumes the role of the RP. It follows that, as the LoA increases so too does the degree of complexity of the UAS and in turn, greater safety assurance is required of the UAS hardware and software components that perform the safety critical functions previously provided by a human RP. This assurance can be provided through more rigorous standards on the design, implementation, testing and operation of the UAS. The training and licensing requirements on the RP will also depend on the LoA of the UAS, thus, clear distinctions are required.

The ICAO definition of *Autonomous Aircraft* was consistently mapped to the same LoA as that determined for *UAS A*, *UAS E* and *UAS F*. There are significant differences in the capability and complexity of these three case study UAS. Interestingly enough, *UAS A* corresponds to the earliest and most primitive of UAS, whilst *UAS F* corresponds to a much more complex system yet to be fielded in reality. This result is illustrative of the diversity of potential interpretations of the ICAO definition of an *Autonomous Aircraft*. A definition of an *Autonomous Aircraft* that is less open to interpretation is required.

4.3 Assessing UAS Modes of Operation

An *Autonomous Operation* is defined by ICAO as *"an operation during which a remotely-piloted aircraft is operating without pilot intervention in the management of the flight"*. (ICAO 2011) Numerous modes of operation have been defined for UAS (refer to Table 4 of Appendix A). Such modes of operation imply an UAS of a minimum LoA.

The implied LoA for the modes of operation proposed by Huang (2008) and for the ICAO concept of an *Autonomous Operation* are assessed using the same LoA frameworks analysed in Section §4.1. The results are summarised in Table 5 in Appendix A. For convenience, the definitions of the different modes of operation proposed by Huang (2008) can be found in Table 4.

As can be observed in Table 5, the operational mode of *Remote Control* is consistently assigned to the lowest LoA across all assessment frameworks. The LoA associated with the ICAO concept of an *Autonomous Operation* corresponds well with that determined for a *Fully Autonomous* UAS, as defined by Huang (2008). Up to nine LoA could be associated with the concept of a *Semi Autonomous* mode of UAS operation. *Semi Autonomous* would encompass the vast majority of all UAS operations. This class of operations would include UAS of significantly different LoA, and in turn, capability and complexity. It is concluded that the classification provided by Huang (2008) is unlikely to provide a suitable partitioning of the different types of autonomous UAS operations for use within regulations.

5 Discussion on a Regulatory Definition of Autonomy

In this section we briefly explore some of the issues in the development of a LoA framework specifically for use in the safety regulation of UAS. Clothier & Williams (2012) present a set of criteria for the evaluation of classification frameworks used in safety regulations. The same set of criteria can be used to evaluate the LoA assessment frameworks and their suitability for use in the regulation of UAS.

5.1 Context

As stated by Clothier & Williams (2012), all classification frameworks have a purpose and this purpose influences their design. The existing LoA assessment frameworks have been developed for a wide variety of purposes, primarily for analysis of mission or operational capability. Subsequently, the properties used to distinguish one LoA from that of the next reflect differences within the specific context. As an example, the ability to declare enemy ground targets and establish their intent has little relevance in determining the LoA for application in civil aviation safety regulations. Within the context of aviation safety regulation the LoA need to be distinguished on the basis of their impact on the safety of UAS operations. Establishing a relationship between safety and the LoA of a system will require assumptions in relation to the missions and environments in which the UAS are operated (see discussion Section §5.5).

5.2 Scale

An ordinal scale is recommended for a regulatory specification of the LoA of UAS. This would allow regulators to "rank" UAS on the basis of their LoA and would serve to divide the continuum of UAS LoA into a finite number of mutually exclusive classes. Regulations could then be developed and promulgated for each class of autonomy in line with the safety and complexity of the system. As discussed in Clothier & Williams (2012), a regulatory classification should be comprehensive, i.e., capable of providing a complete, contiguous and mutually exclusive partitioning of UAS across all foreseeable types. Therefore, the range of the LoA scale should cover all potential cases, from no autonomy through to systems such as that defined by case study *UAS F*. It is evident that highest LoA of some of the reviewed scales do not consider the possibility of autonomous systems which are *completely* independent of a human operator.

5.3 Number of Levels

The more levels defined on the LoA scale, the higher the resolution and the greater ability to distinguish

between differences in the LoA of UAS. Increasing the resolution has the potential to increase the level of flexibility in the regulation of autonomous systems, however, it can increase the regulatory development effort and the complexity of the regulatory classification scheme.

The number of levels defined along the measurement scale can have a significant impact on the effectiveness and practicality of the assessment framework. If too few LoA are defined, then the assessment framework may not be able to distinguish between systems that have a noticeable/significant difference in their respective LoA. In the context of aviation safety regulations, insufficient resolution can result in a failure to distinguish differences in the safety performance of UAS. Conversely, a measurement scale with too many levels would make the framework impracticable to implement. The ideal number of levels is the minimum number needed to distinguish between LoA where there is a significant difference in the safety-performance of the systems. Establishing the relationship between a LoA and the safety performance of the UAS is beyond the scope of this paper. It is important to note that although many of the scales are labelled linearly (e.g., assigned the labels 0, 1, 2...) this does not mean that the relationship between LoA is also linear.

5.4 Component Properties

The set of properties used to determine a LoA should be the simplest and most concise set necessary to distinguish differences in the safety-performance of the systems. The terms concise and simple are included to reflect pragmatic requirements on the regulatory framework for UAS and to simplify the assessment process. For example, the assessment framework proposed by Barber & Martin (1999) requires measurements of how often (% of decisions) and how much (% of time) the RP intervenes in the decision making of the UAS. Such assessments would be sensitive to changes in the mission, task or environment, can be time consuming to evaluate, and are not measures that can be easily verified by a regulatory authority.

5.5 Level of System Complexity

The component properties used to describe autonomy have included properties belonging solely to the machine, properties of the human and the machine, emergent properties of the HMS, or emergent properties of the HMS operation. Thus, autonomy has been assessed at numerous different levels of an hierarchical systems model describing an UAS. This leads to the question as to what level of complexity in the representation of the UAS should the LoA be assessed?

UAS autonomy could be assessed at an operational level, and in so doing take into consideration the complexity of the mission and environment performed by the UAS. An example of this level of representation is the concept of Contextual Autonomous Capability (Huang 2008). A high level requirement for UAS is that they operate seamlessly within the existing airspace system. This high level requirement could be used to establish categories or classes of mission and environment (e.g., UAS operations under visual flight rules and under visual meteorological conditions versus UAS operations in accordance with instrument flight rules), for which "operational" LoA for UAS could be assessed.

UAS LoA could also be assessed in relation to the functions performed by the physical UAS or its component sub-systems. In civil aviation safety regulations flight critical or safety critical functions are determined within the "aircraft system". Safety critical functions can be determined below the operational system level by assuming that the relationships between measures of functional performance and the operational risks are largely independent of a particular mission or environment. For example, a loss of control poses a risk to those people on-board an aircraft irrespective of the mission or environment in which the aircraft is being operated. In such a case, the control function could be classified as having a catastrophic failure condition. From a regulatory perspective, the LoA could be defined in relation to such safety critical functions. Existing function classifications such as those defined in"Part 1309" system safety regulations (FAA 1988, EASA 2012) could be used.

5.6 The Assessment Process

The vast majority of the literature reviewed provided limited or no guidance as to how a LoA scheme could be applied in practice (i.e., the actual measurements required and how they could be aggregated). As described by Clothier & Williams (2012), the LoA measurement scale/scheme and the assessment process are equally important components in any classification framework. If the LoA assessment process is onerous, ambiguous, or overly complex the likely outcome is a miss classification. In the context of aviation safety regulations a miss classification is itself an hazard. An assessment of the LoA of a system must be easily verified by a regulatory authority.

5.7 Adaptive Autonomy

The LoA of an UAS may vary with operating conditions (normal / abnormal / emergency conditions), for particular functions, and/or phases of flight. A methodology will need to be in place to ensure that the safety regulations provide adequate coverage for such systems. This may include the system conditions or process for changing the LoA of the UAS. For example, should the UAS have the ability to change its own LoA *autonomously*? Under some circumstances this may be desirable (e.g., in those situations where the communications link between the UA and the ground control station are lost). If so, what should the safety assurance requirements be for such a capability?

6 Conclusions

Civil aviation safety regulations and guidance material classify Unmanned Aircraft Systems (UAS) as either Remotely-Piloted Aircraft Systems (RPAS) or Autonomous Aircraft Systems (AAS). This distinction is based on the premise that the effective safety risk management of UAS is dependent on the degree of autonomy of the system being operated. However, it is found that there is no consensus on the concept of autonomy, on how it can be measured, or on the nature of the relationship between Levels of Autonomy (LoA) and the safety-performance of UAS operations.

Existing LoA assessment frameworks do not provide a consistent classification of the six generic case study UAS nor of the ICAO concept of an *Autonomous Aircraft*. It was found that the existing LoA assessment frameworks, when applied to the case study UAS, do not provide a consistent basis for distinguishing between the regulatory classes of RPAS and AAS. It was also found that the existing regulatory definition of an *Autonomous Aircraft* is too

broad, covering UAS of significantly different levels of capability and system complexity. Within the context of aviation safety regulations, a new LoA assessment framework for UAS is required.

This is not a surprising outcome as none of the existing LoA frameworks were developed for use in a safety or regulatory context. A new LoA assessment framework will require a clear relationship between autonomous system capability/functionality and safety performance to be established. This needs to be first established at the operational level (i.e., taking into consideration the mission and environment).

The higher the LoA of the UAS, the more it subsumes the functions of the human remote pilot. Hence, the more complex the UAS becomes. A wide range of methods for implementing higher LoA in the hardware and software components of an UAS have been successfully demonstrated. Complex high-level decision making is already being achieved through the use of neural networks, agent-based architectures and probabilistic reasoning. This poses a significant challenge to regulators who must certify such implementations as being "safe". The default performance benchmark is that of the human pilot that has been replaced. Not only does this performance benchmark need to be quantified but new tools for use in the certification process need to be developed. These tools must be capable of exercising the behaviour of the system across a wide range of missions and operating conditions (normal/abnormal/emergency).

7 Acknowledgment

The research presented in this paper forms part of Project ResQu led by the Australian Research Centre for Aerospace Automation (ARCAA). The authors gratefully acknowledge the support of ARCAA and the project partners, Queensland University of Technology (QUT); Commonwealth Scientific and Industrial Research Organisation (CSIRO); Queensland State Government Department of Science, Information Technology, Innovation and the Arts; Boeing and Insitu Pacific. This research was in part supported under the Robust Autonomous Systems collaboration between the University of Newcastle and Boeing.

References

ASTM (2007), Standard terminology for unmanned aircraft systems, Standard F2395-07, ASTM International, West Conshohocken, USA.

Barber, K. S. & Martin, C. E. (1999), Agent autonomy: Specification, measurement, and dynamic adjustment, *in* 'Autonomy Control Software Workshop at Autonomous Agents (Agents '99)', Seattle, USA.

Billings, C. E. (1991), Human-centered automation: A concept and guidelines, Technical Memorandum 103885, National Aeronautics and Space Administration, Moffet Field, California.

Boulding, K. E. (1956), 'General systems theory: The skeleton of science', *Management Science* **2**(3), 197–208.

CAA-UK (2012), Unmanned aircraft system operations in UK airspace – guidance, Civil Aviation Publication CAP 722, Civil Aviation Authority, United Kingdom (CAA-UK), Norwich, United Kingdom.

Clarke, A. C. (1968), *2001: A Space Odyssey*, Hutchinson & Co, United Kingdom.

Clothier, R. A. & Williams, B. P. (2012), A review of regulatory classification systems for unmanned aircraft, Unpublished research report, School of Aerospace, Mechanical and Manufacturing Engineering, RMIT University, Melbourne, Australia.

Clough, B. T. (2002), Metrics, schmetrics! how the heck do you determine a UAV's autonomy anyway?, *in* 'Performance Metrics for Intelligent Systems Workshop (PerMIS -02)', Gaithersburg, MD.

EASA (2012), Certification specifications and acceptable means of compliance for large aeroplanes, CS-25.1309, Technical report, European Aviation Safety Agency (EASA), Cologne, Germany.

Endsley, M. R. & Kaber, D. B. (1999), 'Level of automation effects on performance, situation awareness and workload in a dynamic control task', *Ergonomics* **42**(3), 462–492.

FAA (1988), System design and analysis, advisory circular 25.1309-1, Technical report, Federal Aviation Administration, U.S. Department of Transportation, Washington, DC.

FAA (2011), Unmanned aircraft operations in the National Airspace System (NAS) unmanned aircraft operations in the national airspace system (NAS) unmanned aircraft operations in the National Airspace System (UAS), Notice N JO 7210.766, Federal Aviation Administration.

Galster, S., Barnes, M., Cosenzo, K., Hollnagel, E., Miller, C., Parasuraman, R., Reising, J., Taylor, R. & van Breda, L. (2007), Human automation integration, RTO Technical Report TR-HFM-078.

Hasslacher, B. & Tilden, M. W. (1994), Living machines, Technical report, Los Alamos National Laboratory, Los Alamos, NM.

Hill, A. F., Cayzer, F. & Wilkinson, P. R. (2007), Effective operator engagement with variable autonomy, *in* '2nd SEAS DTC Technical Conference', Systems Engineering for Autonomous Systems Defence Technology Centre (SEAS DTC), Edinburgh.

Huang, H.-M. (2008), Autonomy levels for unmanned systems (ALFUS) framework. volume 1: Terminology, Special Publication 1011-I-2.0, National Institute of Standards and Technology (NIST), Gaithersburg, Maryland.

Huang, H.-M., Messina, E. & Albus, J. (2007), Autonomy levels for unmanned systems (ALFUS) framework, Special Publication 1011-II-1.0, National Institute of Standards and Technology (NIST), Gaithersburg, Maryland.

Huang, H.-M., Pavek, K., Ragon, M., Jones, J., Messina, E. & Albus, J. (2007), Characterizing unmanned system autonomy: Contextual autonomous capability and level of autonomy analyses, *in* 'Proceedings of SPIE Defense and Security Symposium'.

ICAO (2011), Unmanned aircraft systems operations (UAS), Circular 328, International Civil Aviation Organization (ICAO), Montreal, Quebec, Canada.

Insaurralde, C. C. & Lane, D. M. (2012), Autonomy-assessment criteria for underwater vehicles, *in* '2012 IEEE/OES Autonomous Underwater Vehicles (AUV) Conference', Curran Associates, Red Hook, USA, pp. 1–8.

Kendoul, F. (2011), Towards a unified framework for UAS autonomy and technology readiness assessment (ATRA), *in* 'International Conference of Intelligent Unmanned Systems (ICIUS)', Chiba, Japan.

Olson, W. A. & Wuennenberg, M. G. (2001), Autonomy based human-vehicle interface standards for remotely operated aircraft, *in* '20th Digital Avionics Systems Conference, 2001', Vol. 2, IEEE, Wuennenberg, M.G., pp. 7D3/1– 7D3/9.

Parasuraman, R., Sheridan, T. B. & Wickens, C. D. (2000), 'A model for types and levels of human interaction with automation', *IEEE Transactions On Systems, Man, And Cybernetics—Part A: Systems And Humans* **30**(3), 286–296.

Proud, R. W., Hart, J. J. & Mrozinski, R. B. (2003), Methods for determining the level of autonomy to design into a human spaceflight vehicle: A function specific approach, *in* 'Performance Metrics for Intelligent Systems Workshop (PerMIS-03)', National Institute of Standards and Technology (NIST), Gaithersburg, Maryland.

Riley, V. (1989), A general model of mixed-initiative human-machine systems, *in* 'Human Factors and Ergonomics Society Annual Meeting', Vol. 33, Human Factors Society, Santa Monica, California, pp. 124–128.

Sheridan, T. B. & Verplank, W. L. (1978), Human and computer control of undersea teleoperators, Technical Report 15, Massachusetts Institute of Technology, Cambridge, MA, USA.

Sholes, E. (2007), Evolution of a UAV autonomy classification taxonomy, *in* 'IEEE Aerospace Conference', IEEE, Big Sky, Montanna.

Stevens, S. S. (1946), 'On the theory of scales of measurement', *Science* **103**(2684), 677–680.

Stevenson, A., ed. (2012), *Oxford Dictionary of English*, 3rd edn, Oxford University Press.

Taylor, R. M., Brown, L. & Dickson, B. (2002), From safety net to augmented cognition: Using flexible autonomy levels for on-line cognitive assistance and automation, *in* 'RTO HFM Symposium on Spatial Disorientationin Military Vehicles: Causes, Consequences and Cures', La Coruna, Spain.

TC (2008), The review and processing of an application for a special flight operations certificate for the operation of an unmanned air vehicle (UAV) system, Staff Instruction SI 623-001, Transport Canada.

UKMoD (2011), The UK approach to unmanned aircraft systems, Joint Doctrine Notice 2/11, Ministry of Defence, United Kingdom, Swindon, United Kingdom.

USDoD (2011), Unmanned systems integrated roadmap FY2011-2036, Technical report, Department of Defense (DoD), USA.

A Concepts for UAS Autonomy

Concept	Definition
Automatic / Automated	The execution of a pre-defined process or event that requires UAV pilot initiation and/or intervention e.g. automated take-off/landings, way-point navigation, autopilots, pre-programmed manoeuvres etc. (TC 2008) the automatic performance of scripted actions. (ASTM 2007) In the unmanned aircraft context, an automated or automatic system is one that, in response to inputs from one or more sensors, is programmed to logically follow a pre-defined set of rules in order to provide an outcome. Knowing the set of rules under which it is operating means that its output is predictable. (UKMoD 2011) fully preprogrammed and act repeatedly and independently of external influence or control. An automatic system can be described as self-steering or self- regulating and is able to follow an externally given path while compensating for small deviations caused by external disturbances. However, the automatic system is not able to define the path according to some given goal or to choose the goal dictating its path. (USDoD 2011)
Autonomy	the ability of the machine to interpret its environment and make decisions that result in unscripted actions. (ASTM 2007) A UMS's [Unmanned System] own ability of integrated sensing, perceiving, analyzing, communicating, planning, decision-making, and acting/executing, to achieve its goals as assigned by its human operator(s) through designed Human-Robot Interface (HRI) or by another system that the UMS communicates with. UMS's Autonomy is characterized into levels from the perspective of Human Independence (HI), the inverse of HRI. (Huang 2008) The ability to execute processes or missions using on-board decision making capabilities. No intervention by UAV [Unmanned Aerial Vehicle] crew members is required. An autonomous UAV would be capable of dynamic mission management that is not scripted. It would depend on intelligent reasoning and deliberate behaviour for the ability to cope with uncertainty i.e. self-governance. (TC 2008) the quality of being autonomous; self-determination. (FAA 2011) the condition or quality of being self governing. When applied to UAS, autonomy can be defined as UASs own[1] ability of integrated sensing, perceiving, analyzing, communicating, planning, decision-making, and acting/executing, to achieve its goals as assigned by its human operator(s) through designed Human-Robot Interface (HRI) or by another system that the UAS communicates with. (Kendoul 2011) the capability of the system to make decisions based upon an evaluation of the current situation (often referred to as situation awareness). (CAA-UK 2012)
Autonomous	An autonomous system is capable of understanding higher level intent and direction. From this understanding and its perception of its environment, such a system is able to take appropriate action to bring about a desired state. It is capable of deciding a course of action, from a number of alternatives, without depending on human oversight and control, although these may still be present. Although the overall activity of an autonomous unmanned aircraft will be predictable, individual actions may not be. (UKMoD 2011) not controlled by others or by outside forces; independent judgment. (FAA 2011) A UAS is defined to be autonomous relative to a given mission (relational notion) wherein it accomplishes its assigned mission successfully, within a defined scope, with or without further interaction with human or other external systems. A UAS is fully autonomous if it accomplishes its assigned mission successfully without any intervention from human or any other external system while adapting to operational and environmental conditions. (Kendoul 2011) self-directed toward a goal in that they do not require outside control, but rather are governed by laws and strategies that direct their behavior. ...An autonomous system is self- directed by choosing the behavior it follows to reach a human-directed goal. ...autonomous systems may even optimize behavior in a goal-directed manner in unforeseen situations (i.e., in a given situation, the autonomous system finds the optimal solution). (USDoD 2011) that perceives its environment and determines if this affects its goals, and it takes action to ensure as far as practicable (and safe) that its goals will be achieved. It reasons about its course of action from a number of alternatives, to achieve these goals without recourse to human oversight and control. (CAA-UK 2012)

Table 1: Autonomy and related concepts defined in the context of UAS

[1] *own* implies independence from human or any other external system. (Kendoul 2011)

Reference	Number of Levels	Scale	Basis for Scale
Sheridan & Verplank (1978)	10	Ordinal	Autonomy scale based on human-machine role in decision making.
Riley (1989)	12	Ordinal	Autonomy scale is based on the degree of control/authority the machine has over the world (inclusive of the operator).
Billings (1991)	7	Ordinal	Autonomy scale is based on the pilot role in the control and management of system. The control and management continuum is described in terms of "the degree of direct or immediate involvement of the pilot." Olson & Wuennenberg (2001) use the scale to specify requirements on the pilot-aircraft interface for UAS.
Hasslacher & Tilden (1994)	N/A	Interval	A level of autonomy is measured in terms of a Survival Signature Space which is the weighted linear combination of measurements made with respect to the three independent capability vectors of mobility, acquisition, and protection.
Endsley & Kaber (1999)	10	Ordinal	LoA 'taxonomy' is based on the assignment of tasks/functions to a human operator and/or computer. The tasks/functions were defined as monitoring, generating, selecting, and implementing.
Barber & Martin (1999)	N/A	Ratio	Classification based on how often (% of decisions) and how much (% of time) an agent intervenes in the decision making of the robot.
Parasuraman et al. (2000)	N/A	Interval	Level of automation based on the scheme proposed by Sheridan & Verplank (1978) but expressed in relation to the four independent functions of information acquisition, information analysis, decision selection, and action implementation.
Clough (2002)	11	Ordinal	Level of autonomy defined in relation to the combination of three independent functions of perception/situational awareness, analysis/decision making, and communication/cooperation. The framework is used by Sholes (2007) to assess the autonomy of UAS.
Taylor et al. (2002)	6	Ordinal	Level of autonomy defined based on the combined dimensions of pilot authority and contractual authority. A modified version of this scheme is presented by Hill et al. (2007) for describing UAS human-system interaction.
Proud et al. (2003)	8	Ordinal	Level of autonomy defined in relation to four independent dimensions reflecting the decision making processes of Observe, Orient, Decide and Act (OODA). Scheme is based on that developed by Parasuraman et al. (2000), Clough (2002)
Huang, Messina & Albus (2007)	11	Ordinal	LoA relates to the Human Interaction (HI) axis of the ALFUS Contextual Autonomous Capability framework. This axis is described as "the ability for the UMS to identify and communicate and/or negotiate with humans and/or other entities." A consistent measurement framework for the HI axis is difficult to identify from within the various published works. Five levels of HI are defined in (Huang, Pavek, Ragon, Jones, Messina & Albus 2007). Table 1 of (Huang, Messina & Albus 2007) describes the measurement of HI in terms of the % of interaction time and defines eleven "reference levels" of HI.
Galster et al. (2007)	8	Ordinal	Classification is based on the degree of operator control over the UAS.
Kendoul (2011)	11	Ordinal	Similar to the concept of Contextual Autonomy presented in the ALFUS framework, where the LoA of UAS is "characterised by the missions that the UAS is capable of performing (Mission Complexity or MC), the environments within which the missions are performed (Environment Complexity or EC), and independence from any external system including human element (External System Independence or ESI)". The LoA is determined from the ability of the UAS to perform the independent functions of guidance, navigation and control for a specified mission and environment.
USDoD (2011)	4	Ordinal	Autonomy scale is defined in relation to the degree of interaction between human control and the machine motions.
Insaurralde & Lane (2012)	N/A	Interval	Autonomy scale is defined in relation to the five independent contexts of: itself, system, user, environment, and norm. The five contexts are measured in terms of the ability of the system to perform the functions of Observe-Orient-Decide-Act-Check (OODAC). The level of autonomy for each context is determined by the weighted average of the measures for each of the OODAC functions.

Table 2: Autonomy scales identified in the Literature

Autonomy Scale	Case Study UAS						Autonomous Aircraft[2]
	UAS A	UAS B	UAS C	UAS D	UAS E	UAS F	
Sheridan & Verplank (1978)	I	I	I	VIII	IX	X	VII, VIII, IX, X
Riley (1989)	XII	I	VII, VIII	IX	XI	XII	XII
Billings (1991)	VII	I, II	III, IV, V	V, VI	VII	VII	VII
Endsley & Kaber (1999)	III	I, II	II, III	NC	IX, X	X	X
Clough (2002)	II	I	III	NC	NC	X, XI	X, XI
Taylor et al. (2002)	-	I	NC	IV, V, VI	-	-	-
Huang, Messina & Albus (2007)[3]	NC	I	NC	X	XI	-	-
Galster et al. (2007)	VIII	I	V	VI, VII	VIII	VIII	VIII
Kendoul (2011)	II	I	NC	NC	NC	XI	XI
USDoD (2011)	-	I	III	III, IV	-	-	-

Table 3: Assessment of the LoA of the case-study UAS

[2] Defined by ICAO (2011)

[3] Human Interaction (HI) as defined in Table 1 of Huang, Messina & Albus (2007)

Reference	Term	Definition
Clough (2002)	Remotely piloted	The UAV [unmanned aerial vehicle] is simply a remotely piloted aircraft with the human operator making all decisions.
	Remotely operated	The human allows the UAV to do the piloting, but outer loop decisions are made by the human (like where to go and what to do once there). The UAV is a "mother-may-I" system, asking the human permission to do tasks.
	Remotely supervised	The human allows the UAV to execute its own tasks, only taking command if the UAV fails to properly execute them.
	Fully autonomous	The UAV receives goals from the humans and translates that into tasks which it does without human intervention. The UAV has authority to make all decisions.
ASTM (2007)	Semi autonomous	... mode of control[4] of a UAS where the pilot executes changes and conducts the mission through a flight management system interface. Without this input, the UAS will perform pre-programmed automatic operations. This can, but might not, include some fully autonomous functions (like takeoff, landing, and collision avoidance).
	Fully autonomous	... mode of control of a UAS where the UAS is expected to execute its mission, within the pre- programmed scope, with only monitoring from the pilot-in- command. As a descriptor for mode of control, this term includes: (1) fully automatic operation, (2) autonomous functions (like takeoff, landing, or collision avoidance), and (3) "intelligent" fully autonomous operation.
Huang (2008)	Remote control	A mode of UMS [Unmanned System] operation[5] wherein the human operator controls the UMS on a continuous basis, from a location off the UMS via only her/his direct observation. In this mode, the UMS takes no initiative and relies on continuous or nearly continuous input from the human operator.
	Teleoperation	A mode of UMS operation wherein the human operator, using sensory feedback, either directly controls the actuators or assigns incremental goals on a continuous basis, from a location off the UMS.
	Semi-autonomous	A mode of UMS operation wherein the human operator and/or the UMS plan(s) and conduct(s) a mission and requires various levels of HRI. The UMS is capable of autonomous operation in between the human interactions.
	Fully autonomous	A mode of UMS operation wherein the UMS accomplishes its assigned mission, within a defined scope, without human intervention while adapting to operational and environmental conditions.
ICAO (2011)	Autonomous operation	An operation during which a remotely-piloted aircraft is operating without pilot intervention in the management of the flight.

Table 4: UAS modes of operation

[4]*mode of control* is defined as the means the pilot uses to direct the activity of the UAS. Modes include remote control, semi autonomous and fully autonomous. The remote control mode of operations is not explicitly defined.(ASTM 2007)

[5]*mode of operation* defined as the human operator's ability to interact with a UMS to perform the operator assigned missions.(Huang 2008)

Autonomy Scale	UAS Modes of Operation[6]				Autonomous Operation[7]
	Remote Cont	Teleop	Semi Auto	Fully Auto	
Sheridan & Verplank (1978)	I	I, II, III	II, III, IV, V, VI	VII, VIII, IX, X	VII, VIII, IX, X
Riley (1989)	I	II, III, IV, V, VI, VII	VIII, IX, X, XI	XII	XII
Billings (1991)	I	II, III, IV	IV, V, VI, VII	VII	VII
Endsley & Kaber (1999)	I	II	III, IV, V, VI, VII, VIII, IX	X	X
Clough (2002)	I	II, III	II, III, IV, V, VI, VII, VIII, IX, X	XI	X, XI
Taylor et al. (2002)	I	I	III, IV, V	NC	-
Huang, Messina & Albus (2007)[8]	I	NC	II, III, IV, V, VI, VII, VIII, IX, X	XI	XI
Galster et al. (2007)	I , II	II, III	III, IV, V, VI, VII	VII, VIII	VIII
Kendoul (2011)	I	I, II	II, III, IV, V, VI, VII, VIII, IX, X	XI	XI
USDoD (2011)	I	I	II, III, IV	-	-

Table 5: Assessment of the LoA of UAS for different modes of operation

[6] Defined by Huang (2008)

[7] Defined by ICAO (2011)

[8] Human Interaction (HI) from Table 1 Huang, Messina & Albus (2007)

Safety Risk Matrices - Identifying What is Appropriate for Your Business or Undertaking

Tracy A White
System Safety Manager, Public Transport Services – Special Projects
Department of Project Planning and Infrastructure
77 Grenfell Street, Adelaide 5000, South Australia

Tracy.White2@sa.gov.au

Abstract

The use of risk matrices as a risk measurement tool is a common feature in System Safety as it is naturally reflective of the fact that risk has dimensions of both probability and consequence. Whilst the means of safety risk measurement is common, the different risk matrices in use are myriad both in terms of their probabilistic and consequence definitions, but also the granularity of those axes. Often risk matrices are encountered as a tool of organization or company, are generally accepted at face value, and are rarely questioned- they are merely the way we measure risk here. But what are their origins, are they a fundamental element of the safety argument, and are they constant or do they require periodic recalibration (e.g. with the growth of automation)? This paper examines the thinking behind determining an appropriate safety risk matrix for an organisation, one that can be argued as being reasonable and practicable for the undertaking. As a case study, the paper draws on a real life rail problem encountered when happening across an inappropriate risk matrix and establishing and justifying an alternative, which focused on the needs of a modern railway whilst taking into account the existing risk environment and local safety risk influences.

Keywords: System Safety, Risk, Matrix, Safety Criteria

1 BACKGROUND

Public Transport Services – PTS (a rail infrastructure manager and operator responsible for train services in the Adelaide Metropolitan Passenger Rail Network - AMPRN), as part of meeting their obligations under the Work Health and Safety (WHS) law [1][2], have to demonstrate the elimination of safety risks, or where not reasonable and practical to do so, the minimisation of safety risk so far as is reasonably practicable (SFAIRP); an essential contextual element of that argument of reasonableness and practicability is the metrics used to measure that safety risk[1]. In the case of PTS, there is a requirement that the risk metrics is appropriate for the business of conducting rail operation in the first instance and, in particular, conducting those operations in their area of responsibility (AMPRN). PTS are an element of the wider Department of Planning Transport and Infrastructure (DPTI) which has additional responsibilities extending to taxis, buses, road and other infrastructure and which operates a corporate-wide Enterprise Risk Management (ERM) process.

1.1 Applying the Organisational Risk Management Matrix

An initial attempt had been made by PTS Special Projects (PTSP) to compel their rail electrification subcontractors to utilise the umbrella DPTI Corporate Risk Matrix [3] for the Adelaide Rail Revitalisation (ARR[2]) project but when concerns were raised against its practicability and suitability for use in a safety assurance program for rail operations and rail infrastructure development, it gave cause to re-examine that matrix and consider the question of whether there was a case to have a rail-centric safety risk matrix. This process of reflection, which formed the motivation for this paper, attempted to answer the question of whether it is possibly to have a single corporate-wide 'risk matrix' or indeed whether there was an overriding technical requirement to be able justify the appropriateness of the risk measure as a key element of any safety case, which could not be satisfied by the application of the corporate-wide measure.

Whilst the measure of safety risk considers both the probability and consequence aspects, and whilst probability aspects are discussed, for the purposes of this paper, the primary focus is centred on the appropriateness of the consequence measure. The existing consequence definitions (Table 1) was an area that caused the contractors issues, due to the consideration that it was more appropriate for Occupational Health and Safety (OHS[3]), rather than 'process' safety.

C5 – Critical	– Multiple fatalities, – permanent or partial disabilities
C4 – Major	– Single fatality, – permanent or partial disabilities, injuries requiring hospitalisation
C3 – Moderate	– Injuries requiring medical treatment

[1] For the purposes of this paper 'risk' when used in isolation, generally refers to 'safety risk' rather than any other organisational risk consideration

[2] The ARR project delivered additional rail infrastructure, including new track and station, an electrified network and new rollingstock

[3] By current legislation, this would be WHS, or Work Health and Safety.

C2 – Minor	– Injuries requiring first aid treatment
C1 – Insignificant	– Incident with or without minor injury

Table 1: Existing Consequence Subdivisions

It is important to recognise, when referring to 'safety risk' that a distinction is made between OHS and System Safety (and, for that matter, Risk Management). OHS is a term which is generally more related to human-related slips, trips and fall-type events and the application of regulation and codes of practice, whilst system safety is very much an engineering discipline which is an inherent element of systems engineering; the effectiveness of the system safety activity is the primary component for any claim to have assured the safe operation of a railway network (system). This OHS distinction is consistent with, and reflective of, the reporting of the Waterfall Rail Accident Inquiry [4]. It is important to recognise that the focus and needs of these two areas in 'managing safety' will likely not be met by a single 'risk measure'; this paper concentrates on determining what was considered an appropriate rail system safety risk measure for the AMPRN, part of which is and informed attempt to appropriately modify that solution to remain compatible with the wider organisation risk needs. As a reinforcement of the OHS delineation it is worth noting the Baker Report [5] (Texas City) where the principle finding was:

BP management had not distinguished between "occupational safety" (i.e., slips-trips-and-falls, driving safety, etc.) versus "process safety" (i.e., design for safety, hazard analysis, material verification, equipment maintenance, process upset reporting, etc.). The metrics, incentives, and management systems at BP focused on measuring and managing occupational safety, while ignoring process safety. BP confused improving trends in occupational safety statistics for a general improvement in all types of safety.

The oil and gas industry use the term 'process safety', but equally other industries have alternative terms such as 'technical safety' and 'functional safety' to emphasise that OHS/system safety difference.

2 Considerations in Defining Safety Risk

2.1 What is Risk?

To widen the discussion to all risk-interests of an organisation, it is worth qualifying what is meant when talking about 'risk'. There is the perception that all 'risks' (safety, financial, public perception etc.) are 'equal' within an organisation and that all risks should be subject to a single measurement. This approach is flawed not least because there is a legal obligation associated with safety risk [1] [2], which does not attach to other risks such as those related to budget, schedule or reputation. It is important that a desire to impose a single organisational risk measure does not compromise the ability to effectively assess safety and defend that legal duty of care as the arguments of reasonability and practicability invoke markedly different obligations.

The measure of safety risk, being a function of the probability of an event occurring and the consequence (or severity) should that event be realised, is a common construct and one which is ideally suited to be presented in a 2-dimensial graduated axis format (a risk matrix); this aspect has been consistently reflected in system safety standards and publications, both civilian [7] and defence[8] [9], and has been carried forward to the wider rail industry [10]. However, none of these publications define the values, or definitions, for the risk graduations; although some provide 'examples', they explicitly identify the need to develop a matrix appropriate to the domain or application.

It is necessary to distinguish the system safety treatment of 'risk' from the wider 'risk management', particularly when considering employing a single matrix measure. Whilst the definition and measurement of risk in the previously cited system safety publications is consistent throughout, that consistency is not reflected in equivalent risk management publications, be they national [11], or now international [12]. Despite the misconceptions, risk management does not advocate employing a risk matrix, does not propose graduating probabilistic and consequence levels, and does not talk in terms of differing risk levels e.g. extreme, high, medium and low. In terms of potential divergence, the most notable aspect (and would explain the absence of prescribing the use of a matrix) is the meaning and application of risk which for risk management allows consideration of the likelihood, or consequence, or both; for system safety, without exception, it is *always* both. The risk management language may talk in terms of the *risk* of something happening (probability) or the *risk* associated with a train collision (consequence), but for system safety *both* have to feature in any assessment and arguments as to the reasonableness and practicability of any proposed control solution. This difference is discussed here as a means of explaining potential differences which may arise between the safety perspective on the use and definition for a risk matrix and the wider organisational perceived risk management needs/requirements.

2.2 Purpose of the Safety Risk Matrix

Other than the obvious purpose, which is the measurement risk, it is important to reflect on the reason for *measuring risk* in the first instance. A critical aspect for any organisation wishing to be able to claim to be managing risks, and therefore to be meeting their duty under the law [1], is the ability to be able to discriminate between lesser or more important safety risks as it drives the level (and reasonableness) of organisational effort and focus. There are finite funds and resourcing so it is important to deploy resources where they provide the most benefit. In order to do that efficiently, it is important to be able gain a level of relative risk appreciation. If risk scoring is driven by emotional and/or political considerations then there is a real danger that all safety risks end up high and you lose any distinction – then there is no point in scoring them in the first place; the risk matrix is a management tool to make those safety risk

distinctions, not a PR mechanism to show how much you care.

2.3 Safety Risk Datum

As the first consideration, it is important that the risk matrix is reflective of the domain or industry. It is highly questionable to consider cutting-and-pasting a risk matrix from the Oil and Gas industry for use in rail, or aerospace for use in mining, or maritime for roads etc. Simplistically, without going into statistical details, as a relative measure, over the a 10-year period , on average 1400 persons are killed in road traffic accidents pa in Australia (inc. Pedestrians) [14], in comparison just under four people pa were killed on railways over a similar period (this excludes deliberate acts such as suicides) [13], whilst at the bottom of the scale just under that four persons pa died in aviation accidents (but this drops to less than two when excluding recreational flying) [15]. The point to make is that the datum for these different industries is different and for railway operator (or your domain), this needs to be reflected in the adopted safety risk matrix.

If the issue were to be represented graphically (Figure 1), and applying the concept of a 2-dimensional matrix, the datum would refer to the centre, or coverage, of that 2-dimensional area. In the figure below, datum's A to D may be representative of the different domains or industries previously discussed.

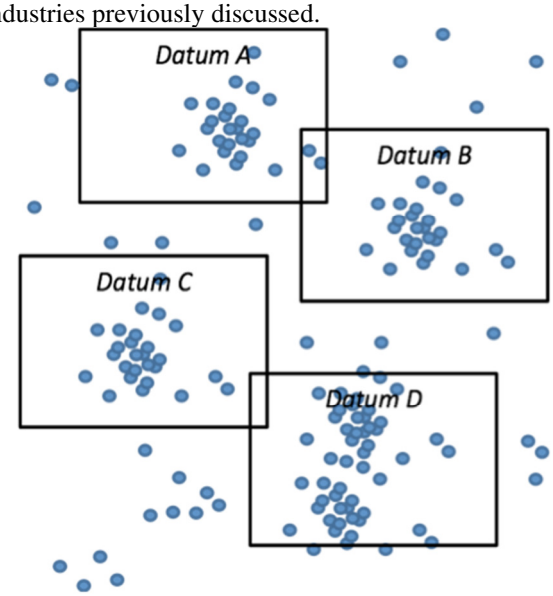

Figure 1 –Distribution of Total Safety Risks

The important thing to recognise is that when defining a safety risk matrix it should reflect the datum, or focal point, which is representative of appropriate local, national or industry norms; there will be obvious variations around the datum as discussed in the following sections.

2.3.1 Passenger Services

Whilst accepting that there is an industry-related datum, even within a given industry, there are operational variations that should be taken into account when measuring risks. Considering the recent break-up of the NSW railways which produced the 3-tier transport system [16] resulting in 3 different rail operators; operationally each will be influenced by different factors such as: frequency and volumes of passengers, logistical aspects such as moving passengers (tunnels and passageways vs. above ground structures), the degree of automation (ATO) etc., so whilst they are all 'rail' and arguably emanate from the same datum, there will be variations required around that datum to reflect the differing local safety risk influences.

As a rail operator PTS need to recognise that there are certain events that will occur on the network, namely suicides (1-fatality), Level Crossing (LX) collisions (2-3 fatalities)[4] and pinnacle safety risk items such as train collision/derailments (multiple >3 fatalities)[5]; these events need to be distinguishable in the risk criteria by their relative (and different) consequence (or severity) levels as a differentiating risk factor. The original DPTI organisational matrix did not, and could not, support that safety risk distinction. One significant determining factor that should be distinguishable in the safety risk matrix (at least for the AMPRN) is the level crossing (LX) contribution; when compared to other states, with 952 crossing sites, South Australia has a significant volume of public access railway level crossings, with only 17 percent having active protection [17]. Given that the human influences such as driver inattention, driver distraction, risk taking and disobeying warning signage which are common contributors to vehicle-train level crossing crashes, this measure of overall safety will (and should) feature heavily in the PTS safety assurance effort and management.

2.3.2 Freight Services

As a counter to the previous case, it needs to be appreciated (reflected in the safety risk matrix) that the dominant influencing factor for rail operations may well not relate to death or injuries. With a freight operator there is a significant risk exposure reduction brought about by the absence of passengers in the activities being conducted, but a derailment will bring with it a significant financial cost which will prompt the business to undertake risk reduction measures commensurate with that cost [18]. The safety risk related to such a derailment may be significantly reduced due to the reduced personnel exposure but the driving measure of 'reasonable and practicable' for the operator would likely swamp the requirement to meet the legal obligation [1]; it is important that the matrix supports this distinction as it is the safety that needs to be argued to a regulator, or court of law, and it should not be obscured by considerations of the financial drivers.

2.4 Dimensioning

The complementary component to the datum'ing issue is that the aspect of the dimensions, namely determining the probability and consequence axes. Dimensioning refers to size and subdivisioning of the risk area. In the example

[4] Representative of an average figure based on a provisional review of a selection of 20+ Level crossing incidents in Australia (1943-2012)

[5] Representative of a worse case rail event involving a major train collision or derailment

proposed by MIL-STD882 [19] that would result in four consequence subdivisions of: catastrophic, critical, marginal; likewise for probability the five subdivisions would be: frequent, probable, occasional, remote and improbable; collectively referred to as a 5x4 matrix.

When considering the previous example (§2.3.2) it can be seen that it would be possible to define two dimensions which could accommodate both the passenger and freight domains (outer area in Figure 2). But the result of that approach would be that, for freight-use, the outer dimensions would be under utilising effectively meaning that the safety risk of primary concern for a freight operator would be unduly compressed into the lower left margin, compromising the effective ability to risk discrimination discussed previously.

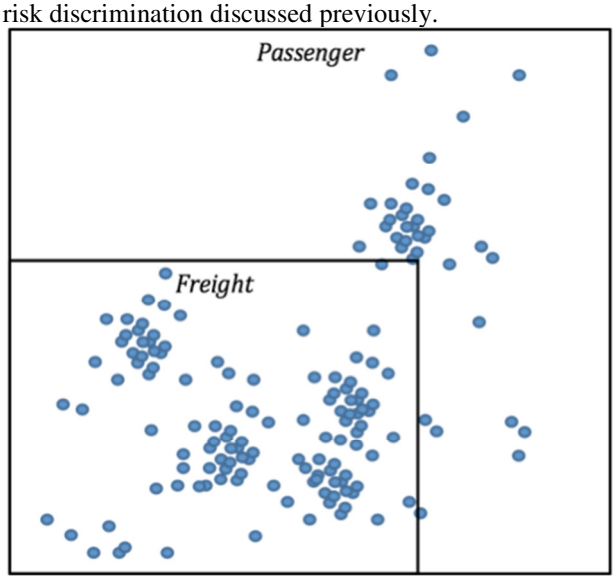

Figure 2 – Passenger and Freight Relative Safety Risk Considerations

Whilst the example is stylised, it illustrates the point that the chosen risk matrix should allow for the spread of risk to enable the organisation to appreciate the relative risk distribution and respond and manage according to the relative importance associated therewith. Inappropriate matrices will not support that differentiated visibility and, within the compressed risk areas, it then becomes a subjective opinion as to what degree of risk mitigation treatment to apply to safety risks (within the same band).

Whilst the datum question sets the primary focus and emanates from a domain consideration, dimensioning should provide the finer tuning focused on the local factors which most influence the specific safety risk problem application (factors such as those previously discussed in relation to the 3-tier transport system, §2.3.1).

3 Considerations in Defining Safety Risk – a Practical Application

3.1 Proposed Safety Matrix

The primary aspect of the safety risk matrix, that the proposed solution attempted to amend (drawing from previous datum and dimension discussions) was the consequence definitions (or dimensioning); it was also a consideration that the proposal needed to reflect current national/international rail best practice, and thereby prove a measure of legal reasonableness in explicitly distinguishing volumes of casualties (inc. single, low and high volumes).

The proposed consequence subdivisions proposed were:

C5 - Critical	– Greater than 4 fatalities[5] – Greater than 40 serious injuries
C4 - Severe	– 2-4 fatalities[4] – Greater than 21-40 serious injuries
C3 – Major	– Single fatality – 11-20 serious injuries – Greater than 200 minor injuries – Significant long term damaged to health or wellbeing (foreshortening of life expectancy) – Large reduction in safety margin or functional capability – Crew physical distress/excessive workload such that operators cannot be relied upon to perform required tasks accurately or completely
C2 - Moderate	– 21-200 minor injuries – Less than 10 serious injuries (1-28 days hospitalisation) – Degree of longer term damaged to health or wellbeing (quality of life) – Pronounced reduction in safety margin or functional capability – Significant increase in operator workload
C1 – Minor (insignificant)	– Less than 20 minor injuries (local first aid or less than 1-day hospitalisation) – Slight reduction in safety margin or functional capabilities – Slight increase in workload such as degraded operations

Table 2: Proposed Consequence Subdivisions

From a legal standpoint, as a responsible rail operator, it was considered that it would not be *reasonably practicable* to be seen as treating the safety risk associated with a LX fatalities[6] the same as a train collision/derailment resulting in notably more fatalities, the latter should *reasonably* attract greater attention, that increased attention being commensurate with the greater safety risk (but the extant matrix would score them the same so would not make that distinction and thereby compel the need to do any 'more').

The proposed matrix provided a safety risk distinction between the Waterfall-type events [4], with low frequency high fatality volumes, which would always be assessed as high (and should always be on the executive

[6] With 952 public access railway level crossings in South Australia [17], this was considered to be significant local influencing safety risk factor

radar), LX events that, with lower fatality volume but a higher frequency, (also likely to be in the higher region and and therefore require appropriate oversight), and suicides which happen quite frequently in comparison and would (on aggregate figures) also appear at the upper end of the scale; with the latter two scenarios, there is a possibility of real risk reduction (by affecting the occurrence rate), in doing so it is important to talk about the 'risk' (likelihood/consequence) and not just the 'consequence' component thereof.

If there was an example required of the need to be able to appropriately differentiate safety risks from a casualty volume perspective, it is apparent from the additional requirements associated with trackside infrastructure. The issue pertains to the escalation event of a derailed train colliding with trackside infrastructure and that structure then collapsing onto the derailed train (see Granville Rail Disaster [6]). A structure (such as a bridge) collapsing onto a train would be considered to significantly escalate the consequence of a derailment; in safety terms the risk therefore should be measurably higher thereby prompting further risk reduction measures from the safety management process. However if the 'worst case' consequence safety risk matrix is 'multiple deaths' (critical) (Table 1), a train derailment hitting a structure (multiple fatalities) and the subsequent collapse of that structure on the train (significantly more fatalities) would not, based on the DPTI risk measure, produce an increased consequence (and thereby risk) rating. The unchanged risk of the escalation event should therefore not attract any different measures or responses (no difference in risk); in reality there are additional safety requirements associated with trackside infrastructure which produces a disjoint of response vs. measured safety risk.

3.2 Constraining Factors

Following on from the base requirement to appropriately measure the safety risk, there was a need to tailor that solution to satisfy the wider organisational risk measurement needs; this section discusses the compromises considered in the final proposed matrix.

3.2.1 Likelihood Dimensioning

One of the constraining factors applied to the proposed matrix was to provide a degree of compatibility with the existing risk matrix. To this end, the same 5x5 dimensioned matrix was retained as were the relative risk divisions (extreme, high, medium and low), which would allow a degree of comparison with other non-safety risks within the organisation; so all 'high' risks should attract similar management oversight albeit for different organisational priorities (safety being just one).

In attempting to maintain some correlation with the extant risk matrix, whilst still providing a more appropriately operationally focused rail safety risk assessment in being able to differentiate between the casualty volumes, the proposal provided in a wider spread for the higher order consequences to achieve that; the lower order likelihood frequency cut-off was retained at 10+years. The frequency cut-off of 10 years is in contrast to other comparable matrices (including TfNSW, NWRL and an ARUP proposed matrix), which included additional bands for 10-100 years and 100years+. The result means that the proposed matrix has stretched (spread) the right hand side of the matrix (consequence) for the higher risk items and but maintained the compression at the lower end of the matrix (likelihood); the likelihood delineations have been driven in part by a previous ARTC example so were still considered to be appropriate for rail operations. The PTS rail operator may consider changing these dimensions should it be felt that there is a need to record and track low frequency/high consequence events (10-100years, 100+years) associated with extreme weather events or environmental factors, but it would be advisable to extend the matrix beyond the current 5x5 to avoid compressing the existing risks still further and loosing the ability to differentiate between risks in the vertical direction.

3.2.2 Terminology and Current Safety Practice

The proposed consequence delineation change brought with it a problem when trying to retain the original terminology which was more suited to the broader organisational needs. A single death, using the original language, would be termed 'moderate' which would not be socially palatable in the public domain. Whilst 'critical' may generally be the pinnacle importance for OHS-type considerations, it is not appropriate for rail safety. The current international thinking is to use the term 'multiple fatalities', but this term is unlikely have significant use or meaning in the OHS or broader organisational non-rail risk areas.

Ideally the highest consequence term should be catastrophic, but if that was not agreeable to the wider organisation, then an alternative hierarchy could be: critical, severe, major, moderate, minor as employed in the proposed matrix. The 'insignificant' consequence (Table 1) was replaced as, if something is 'insignificant', then the 'risk' (for safety) should be 'closed', there is no sense in scoring a risk as not being a risk (if it is 'significant' then it is scored and tracked, if it is not then there is no basis for the entry); an event 'without injury' has no meaning or relevance for safety risk – if there's no harm, there's no hazard.

3.3 Additional Considerations

3.3.1 Direct and Indirect Risk

A point for clarification in the application of the safety risk matrix is that the risk scoring does not (and arguably should not) make a distinction between the type, or class of injury/fatality, so a fatality through suicide would score the same as a rail track worker being killed in terms of risk. A track worker is an example of a direct risk whereby the organisations exercises considerable control over their behaviour, actions and practices and as such has a clear duty under the law to ensure their safety. In contrast, an organisation does not have the same degree of control or responsibility in managing safety risks related to a suicide or trespass event. It is not the case that it is acceptable to do nothing to prevent a suicide event, not least because of the psychological effect on operational personnel, but what the organization would be

prepared to do, and what would reasonably be expected of it legally, will be different. So the risk-focus to the organization will be the same for both events i.e. the level of management oversight and scrutiny required, but the proposed mitigation action would be argued on a different basis of *reasonableness* and *practicability*.

3.3.2 Aggregate and Relative Risk

There is no single measure of risk and risk matrices need to support different and appropriate approaches including measures related to: single, aggregate and relative risks [20]. Single event risk measures are associated with a discrete activity whilst continuous exposure such as would be experienced in operations are more reflectively considered as an aggregate risk; the risk matrix should support both of these measures with the assessment process applying whichever is more appropriate and reasonable.

Definitions are included in the proposed safety risk matrix (Table 2) which supports different approaches but guidance is required which helps the assessment process to consider where it is more appropriate to be scoring single risk or aggregate risks, or indeed whether there is degradation in safety (relative).

Relative risk measures are common internationally in the aerospace domain where standards and regulations talk in terms of *increased operator workload* and *reduction in safety margins* [21]. However, this is not inconsistent with elements of the Rail Safety Act [22], which target relative safety risk factors such as drugs, alcohol and fatigue which will bring about a *reduction in safety*. This is a sensible recognition that certain events, whilst not leading to an immediate realisation of harm, have significance in their consequence of having increased the overall risk. An organisation quite rightly wishes to manage the occurrence of these events, so instigates suitable controls (e.g. drug & alcohol testing and fatigue management). From a safety point of view it therefore requires that a relative risk measure be assigned which requires an appropriate severity (or consequence) description.

3.4 RECALIBRATING THE SAFETY RISK MATRIX

The formulation of a safety risk matrix, if made on the on the basis of considerations described in earlier sections, in determining the datum and dimensions, should not be closed to recalibration based on organisational experience and findings. Equally, the development of automated technology, which removes, or significantly supplements, current operations, or event an organisational desire to includes risks associated with ultra low frequency events (post tsunami, Fukushima or other 'Black Swan events') should prompt the need for recalibration.

A particular technological effect that would have a profound influence on the safety risk spread would be the introduction of automated systems. A direct effect of some automation will be to reduce the safety risk exposure for operators i.e. driverless systems. What that might mean in the context of the previous rail example (Figure 2) is the elimination of some of those safety risks (Figure 3), or a significant reduction in likelihood e.g. engineering failure probability vs. human failure probabilities. The matrix recalibration would then attempt to adjust the risk spread (dimension) to occupy the vacated (eliminated) safety risk space to optimise the differentiation of the safety risks.

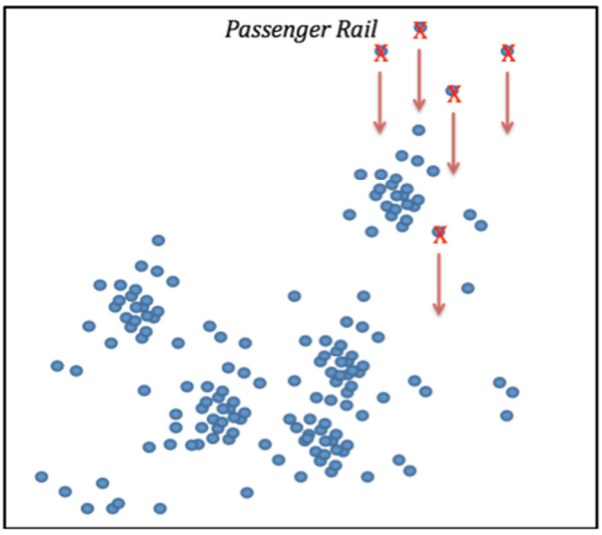

Figure 3: Elimination of Safety Risks due to Increased Automation

4 AFTERTHOUGHT ABOUT RISK PERCEPTIONS & NON-TECHNICAL LANGUAGE

As an anecdote to this paper, in process of presenting and justifying the proposed safety risk measure, it is worth recounting the responses and discussions to the proposal as it illustrates the potential confusion between system safety and risk management thinking. The articulation of the concerns raised brought into stark contrast the misunderstandings arising from inappropriate application of 'risk management' language and concepts to the discipline of system safety, in particular what purpose the risk matrix serves in a program or organisation; it is worth noting that the oft quoted ISO31000 'risk management' standard [12] does not even discuss the use of a risk matrix so it possibly forgivable that 'risk management practitioners' may be confused about its use within a safety program.

The most significant response to the proposed matrix was that, due apparent reduction in consequence for a LX-type event consequence (in the old terms, shifting from a *critical* consequence to *major*) that this indicated an increased 'risk appetite'. The use of the term 'risk appetite' (or tolerance) is an inappropriate application of 'risk management' [23] concepts and thinking to system safety[7]. To talk about the 'risk appetite' of an organisation in regards to safety risk suggests that it has a choice in the level of safety risk it is willing to accept but, unlike organisational risk, safety risk tolerability is not an organisational choice but is a legal measure [1] and is measured by the degree to which is can be argued that the safety risk has been eliminated or minimised SFAIRP.

[7] the application of 'risk' in 'risk management' was previously discussed in §2.1

Rather than indicating some form of *increased* risk *tolerance*, it could be argued that a broader organisation risk matrix which does not discriminate between the significance of a LX event from a major train accident, or 4 suicides a year from 14 (i.e. the consequence would be the same), could be considered to demonstrate a *higher* safety risk appetite due to the equal tolerance for greater casualty volumes.

The reality from the safety perspective (not risk management) is that a risk moving from a high to a medium level would still be subject to a SFAIRP justification; the law [1] attaches no significance to different risk levels so a high to low change would make no difference to the requirement for a justification for the risk mitigation steps taken. This misunderstanding in this department is probably further illustrated by extant matrix (Figure 4) only calling for a SFAIRP justifications for 'high' risk which has no legal or logical basis.

Extreme	Consequences would threaten the survival of not only the activity but also the department possibly causing major problems for clients and the SA Public Sector - requires prompt action by executive management to implement stringent new controls in order to make the risk tolerable
High	Consequences would threaten the survival or continued effective operation of a key business area/service or portion of DPTI - existing controls must be validated and executive management must be able to demonstrate that SFAIRP principles have been applied if risk level is to be tolerated
Medium	Consequences would threaten an activity - existing controls must be effective and possibly additional mitigation action effectively implemented - action may be managed below executive management
Low	Risk is managed by current practices and procedures - consequences are dealt with by routine operations - monitor routine practices and procedures for effectiveness - maintain regime of continuous improvement

Figure 4: SFAIRP requirement against risk levels

5 CONCLUSION & RECOMMENDATIONS

It was found that the use of a broad organisational safety risk matrix was not technically justifiable for supporting system safety activities in delivery of a demonstrably safe Railway. The primary weakness was considered to be its failure to demonstrate consideration of the rail domain and the inability to appropriately differentiate the relative importance of safety risks across the AMPRN (in particular as that relates to casualty volumes). Based on a review of current system safety good practices and the practical needs for conducting an operational rail-focused system safety program, an alternative safety risk matrix was proposed along the lines discussed in this paper.

The proposed safety risk matrix was considered to take into account appropriate domain, technical and local considerations and was proposed as being applicable and justifiable for its application in the management and assurance of safety for the AMPRN.

A recommendation was made that that the basis of safety risk assessment for the organisation be amended to reflect one that was more appropriate for the rail domain and the specific needs of the local rail operations; the technical rationale in terms of datum and dimensioning decisions, as discussed in the paper would form the basis of any operational safety case.

It was further recommended that the safety measure (the matrix), as with any organisation, be periodically reviewed and recalibrated as required, particularly as it may be impacted by significant technological changes such as the advancement of automation.

6 REFERENCES

[1] Work Health and Safety Act 2012 (SA)

[2] Work Health and Safety Regulations 2012 (SA)

[3] Operational Risk Management Procedure, KNet#6681046, version 1, dated October 2012

[4] Special Commission of Inquiry into the Waterfall Rail Accident, Final Report, Volume 2, dated January 2005

[5] James A. Baker, III, The BP U.S. Refineries Independent Safety Review Panel, 2007

[6] Agency No. 6087, Title:, Formal Investigation of an Accident on or about the Up Main Western Railway Line at Granville on 18th January 1977, dated 11 May 1977

[7] IEC 61508 Functional Safety of Electrical/Electronic/Programmable Electronic Safety-related Systems (E/E/PE, or E/E/PES)

[8] Defence Standard 00-56: Safety Management Requirements for Defence Systems

[9] MIL-STD-882D, Department of Defense Standard Practice: System Safety

[10] EN 50126-1, Railway Applications - The specification and demonstration of Reliability, Availability, Maintainability and Safety (RAMS)

[11] AS/NZS 4360:2004 – Risk Management

[12] AS/NZS ISO 31000:2009, Risk management - Principles and guidelines

[13] ATSB Transport Safety Report , Australian Rail Safety Occurrence Data: 1 January 2001 to 31 December 2010, RR-2011-004 Final

[14] Department of Infrastructure and Transport, Road Deaths Australia 2011 Statistical Summary, ISBN: 978-1-921769-74-0.

[15] Australian Transport Safety Bureau, A Preliminary Analysis of Fatal General Aviation Accidents in Australia: 1991 to 2000, Accession Number:00972565

[16] Director General of Transport for NSW, Sydney's Rail Future -Modernising Sydney's Trains, ISBN: 978-1-922030-18-4

[17] South Australian Level Crossing Safety Strategy and Action Plan (2010-2012)

[18] Review of Maintenance Costs and Derailment Costs Associated with Central Queensland Coal Systems, for Network Access Group, Queensland Rail, Himark Consulting Group Pty ltd, dated August 2005

[19] MIL-STD-882D, Department of Defense Standard Practice for System Safety
[20] Safety Risk Aggregation: The Bigger Picture, David Rhys, Safety Assurance Services Ltd, date 18 August 2009
[21] RTCA DO-178B, Software Considerations in Airborne Systems and Equipment Certification
[22] Rail Safety Rail Safety National Law (South Australia) Act 2012
[23] Risk Appetite & Tolerance Guidance Paper, Crowe Horwarth Global Risk Consulting

Risk-management of UAS Robust Autonomy for Integration into Civil Aviation Safety Frameworks

Tristan Perez[1] **Reece A. Clothier**[2] **Brendan Williams**[3]

[1] School of Engineering
The University of Newcastle,
Callaghan, New South Wales 2308, Australia
Email: tristan.perez@newcastle.edu.au

[2] School of Aerospace, Mechanical, and Manufacturing Engineering
RMIT University,
PO Box 71, Bundoora, Victoria 3083, Australia
Email: reece.clothier@rmit.edu.au

[3] Boeing Research & Technology Australia
GPO Box 767, Brisbane, Queensland 4001, Australia
Email: brendan.p.williams@boeing.com

Abstract

This paper discusses a model of the civil aviation regulation framework and shows how the current assessment of reliability and risk for piloted aircraft has a limited applicability for Unmanned Aircraft Systems (UAS) as technology moves towards higher levels of autonomous decision making. Then, a new framework for risk management of robust autonomy is proposed, which arises from combining quantified measures of risk with normative decision making. The term *Robust Autonomy* describes the ability of an autonomous system to either continue or abort its operation whilst not breaching a minimum level of acceptable safety in the presence of anomalous conditions. The term combines reliability, safety, and robustness. The decision making associated with risk management requires quantifying probabilities associated with the measures of risk and also consequences of outcomes related to the behaviour of autonomy. The probabilities are computed from an assessment under both nominal and anomalous scenarios described by faults, which can be associated with the aircraft's actuators, sensors, communication link, changes in dynamics, and the presence of other aircraft in the operational space. The consequences of outcomes are characterised by a loss function quantifies the desirability of the outcomes.

Keywords: Robust Autonomy, Unmanned Aircraft Systems, UAS, Regulation, Certification, Bayesian Reliability

1 Introduction

Unmanned Aircraft Systems (UAS) are a rapidly growing sector of the civil aviation industry. A key challenge facing this emerging sector is the lack of a regulatory framework, either prescriptive, performance or goal based, that can provide assurance in the safety of UAS operations. This framework must account for the varying Levels of Autonomy (LoA) in UAS.

There is an increasing demand for higher LoA in UAS. This demand largely stems from the need to lower the operational cost of UAS by reducing the number of people required to operate the system and to reduce the need for continuous communication links. The role of the human Remote Pilot (RP) in the operation of the UAS depends on the LoA of the system. The higher the LoA, the more the UAS subsumes the role (i.e., functions) of the human RP (Clothier & Walker 2013). Thus, as the LoA increases so does the complexity of the UAS and the higher the degree of safety assurance that must be demonstrated by the components of the UAS. This creates a new and challenging paradigm for National Airworthiness Authorities (NAAs) responsible for managing the safety of the civil UAS industry.

One of the first steps in the safety management process is the establishment of stakeholder goals in relation to the safety performance of the system. A risk assessment is then undertaken to estimate the safety performance of the system. The risk estimates are then compared against the goals to determine which of the identified hazardous scenarios require mitigation and control. In a civil aviation safety management context, the mechanisms put in place to reduce (mitigate and control) the identified risks establish the framework of regulations, standards and procedures relating to

1. design, manufacture, maintenance, and operation of the aircraft (i.e., UAS),
2. training and licensing of personnel, and the
3. responsibilities of the organisation.

Compliance with this framework provides a degree of assurance or confidence in the safety performance of the system. Accidents involving civil aviation aircraft are nowadays extremely rare events. Subsequently, significant operational experience is needed before estimates of the actual safety performance of the regulated system can be determined with a reasonable degree of confidence.

Differences between the Conventionally Piloted Aircraft (CPA) and UAS safety paradigms will influence how the safety risks associated with the operation of the two technologies can be most effec-

tively managed. Researchers have explored a number of different components of the safety risk management process and its application to UAS. A number of factors that can give rise to differences in the specification of safety goals for CPA and UAS are presented in Clothier & Walker (2013). Issues specific to the evaluation risks associated with UAS are explored in Clothier, Williams, Fulton & Lin (2013). An overview of the safety risk management process and some of the issues specific to its application to UAS can be found in Clothier & Walker (2013).

The default position of NAAs is to seek to apply and adapt the existing CPA framework of regulations to UAS (Clothier et al. 2011). It has been argued that the existing regulations may not provide for an effective management of the risks for UAS due to the differences between the CPA and UAS safety risk paradigms (Hayhurst et al. 2006, Clothier et al. 2011, Clothier & Wu 2012). One of the most significant differences between the two paradigms is that the primary risks associated with UAS operations are to people external to the system (i.e., third parties on the ground or secondary parties onboard other aircraft). Conversely, the primary risks associated with CPA operations are to people onboard the aircraft (Clothier & Walker 2013). Another difference, and the principle subject of this paper, is in the integral contribution of the "human element" to the safe operation of the system.

For UAS, many of the functions that once were performed by a human pilot are now provided by hardware and software systems. The application of the existing CPA framework of regulations and standards may not account for differences in the allocation of functions to hardware, software and "human" components of an UAS. The allocation of functions will also depend on the LoA of the UAS[1]. Hayhurst et al. (2006) states that

> "It is not clear, however, whether existing regulations that are based on a historical pairing of pilot and plane can be adapted to accommodate UASs, or whether UASs constitute a fundamentally different category of aircraft requiring their own set of regulations."

This issue is further explored in this paper. In order to satisfy the safety performance objectives established for UAS, safety assurance is required at higher 'levels' within the aviation safety system. New tools for providing assurance in the safety of autonomous aviation systems are needed. These tools must be capable of assessing the safety of the autonomous system under a wide range of operating conditions, missions and failure scenarios.

The rest of the paper discusses the current safety assurance framework and then proposed a new framework for risk management of UAS with increased levels of autonomy.

2 The Safety Assurance Framework

Figure 1 depicts a hierarchical model of the aviation safety system. This model describes different levels of organisational complexity of the aviation system. The focus of this section is on describing the different mechanisms for providing assurance in the safety at the level of *Operations* and below.

[1] A review of different frameworks for describing levels of autonomy is provided by Clothier, Perez and Williams (Clothier, Perez & B. Williams 2013)

The *Operators* component describes the framework of organisational policies, procedures and resources in which *Scenarios* are conducted. A *Scenario* encompasses the interaction of the components of *Aircraft*, *Mission*, and *Weather*. While the *Airspace System* (which includes the various Classes of Airspace and the Air Traffic Management (ATM) services provided) is determined at a level higher than *Scenarios* the *Airspace Requirements*, which draws on the specifics from the *Airspace System*, is determined by the *Mission*. The organisational environment is a significant factor in the safety of aviation activities. However, further consideration of the *Operators* component is beyond the scope of this paper.

The lower tiers of the model are inspired by the *SHEL Model* (Edwards 1972). Where, *'S'* stands for the software, *'H'* for the hardware, and *'L'* for the liveware (or human) components of an *Aircraft* and *'E'* for the aircraft environment (referred to as the component of *Weather* in Figure 1). The component of *Liveware* represents the interaction of a team of *Individuals* performing a range of *Roles*. The behaviour of a *Machine* emerges from the interaction of both of its components of *Hardware* and *Software*, and the behaviour of the *Aircraft*, through the interaction of the components of *Machine* and *Liveware*, for given *Missions* and *Weather* conditions.

The existing framework of regulations for CPA and its relationship to the components of the model illustrated in Figure 1 is discussed in the following section.

2.1 Safety Assurance for CPA

Historical aviation accident and incident data can be used to provide estimates of the actual safety performance of CPA at the system levels of *Operations* and *Scenario*. The observed "safety performance" is largely due to the safety assurance provided by a framework of regulations, standards and procedures governing different components of CPA *Operations*. The regulatory framework for CPA separates regulations pertaining to the initial and continuing airworthiness of the system (e.g., requirements on the design, manufacture and maintenance of the *Machine*) from those regulations pertaining to the operation (i.e., independent of the *Mission* and *Weather*), and from the training, licensing, and proficiency of the aircrew (i.e., the component of *Liveware*). Each component of the regulation can be thought of as having a 'contribution' to the overall safety performance of CPA *Operations*. The separate contributions are illustrated in Figure 2. It is important to note that in reality the safety performance of CPA *Operations* cannot be reduced to discrete contributions. Safety performance is an emergent property, specifically, an irreducible property and one "which is not determined solely from the properties of the system's parts, but which is additionally determined from the system's structure and behaviour" (Thomé 1993). However, for the purposes of providing a simple illustration for discussion it is represented as an "aggregate" relationship.

For CPA, the primary entities at risk are the crew and passengers onboard the aircraft (Clothier & Walker 2013). Consequently, CPA regulations implicitly aim to limit or eliminate harm to those aboard the aircraft, and secondarily to those over-flown (Hayhurst et al. 2006). This philosophy is reflected in the framework of initial and ongoing airworthiness regulations, which govern the design, manufacture and maintenance of a CPA. As described by Haddon & Whittaker (2002), as far as is practicable, the airworthiness codes of regulatory requirements avoid any

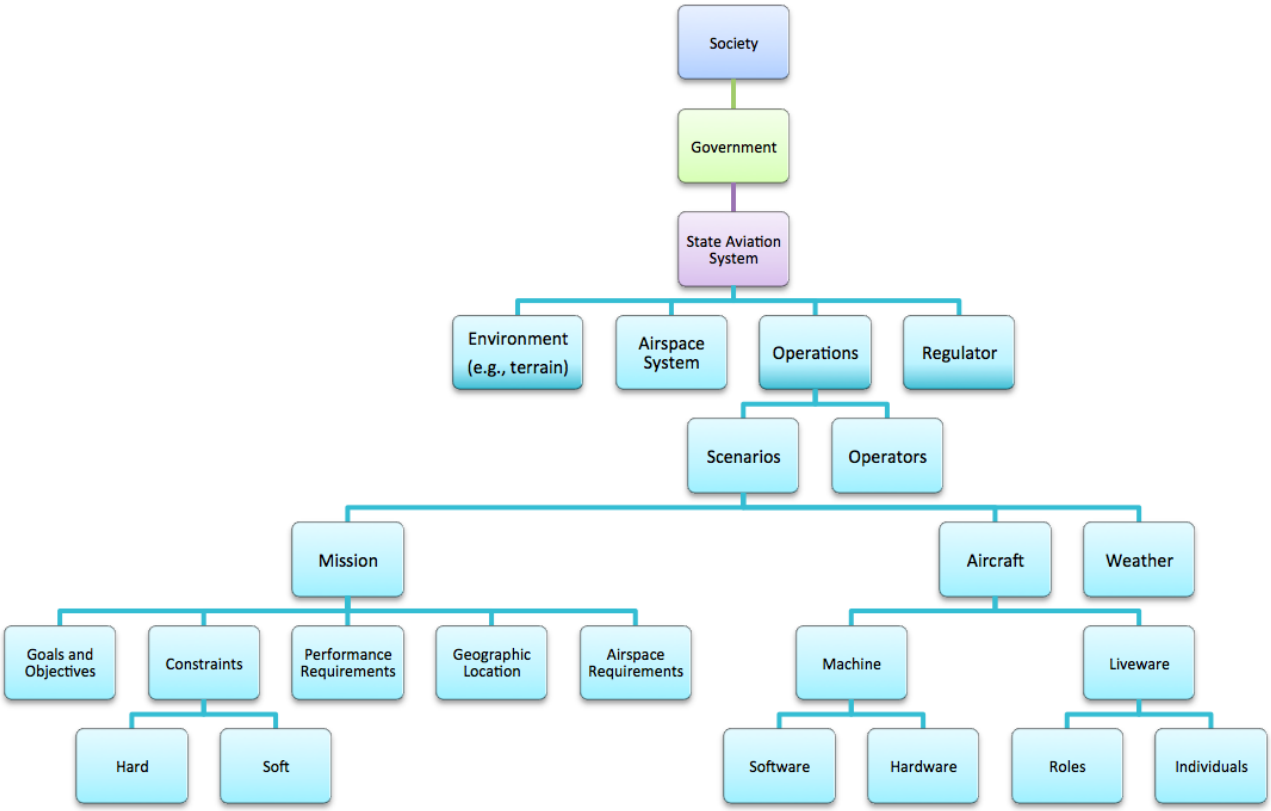

Figure 1: A model of the civil aviation regulation framework.

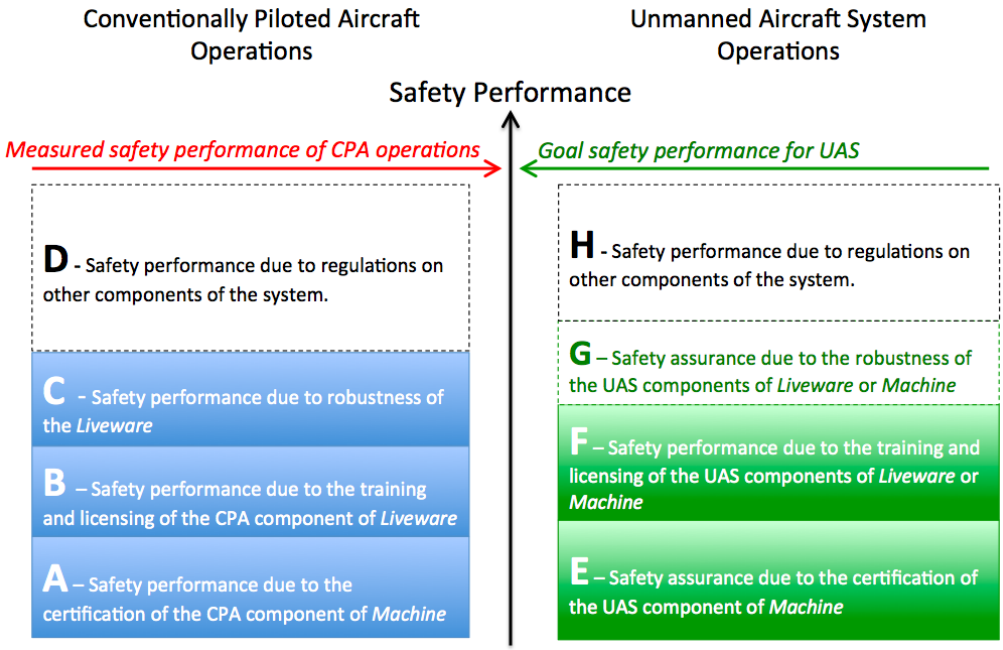

Figure 2: Comparison of Contributions to Safety Performance of CPA and UAS Operations.

presumption of the purposes for which the aircraft will be used in service. Referring to Figure 1, the airworthiness regulations provide safety assurance below the level of the *Aircraft* at the component of the *Machine* (i.e., largely independent of the components of *Mission* and *Weather*). The airworthiness regulations are specific to the component of the *Machine*. Compliance to these regulations and standards is recognised through the issuing of a Type Certificate and Certificates of Airworthiness—see, for example, CASA (2000). The airworthiness regulations and standards contribute to a proportion of the overall safety performance of CPA *Operations*, labelled as region '*A*' in Figure 2. It should be highlighted that while an *Aircraft* design (formalised by the Type Certificate) will limit the types of *Mission* and *Weather* in which operations can take place there is no certification requirement at the level of *Scenarios*. There is a consideration of suitability conducted at the *Operations* level, including continuation training requirements.

CPA airworthiness regulations are defined largely independent of the human component of an aircraft system (the component of *Liveware*). Whilst standards for human-machine interface and handling qualities are defined, the airworthiness standards are implicitly defined based on the assumption of a "nominal pilot". Separate regulations and standards pertaining to the training, licensing and currency of aircrew provide safety assurance in the component of *Liveware*. Aircrew are trained to operate safely a particular type of aircraft under a wide range of normal and abnormal operating conditions, missions and emergency scenarios. The proficiency of the crew in performing these functions is provided through training, examination and through demonstration by means of a flying test(s). Satisfactory performance results in the issuing of a licence. Training and licensing of *Liveware* contributes to a proportion of the overall safety performance of CPA *Operations*. This contribution is represented as region '*B*' in Figure 2. It is important to note that the performance of the *Liveware* can be highly varied and unpredictable. The issuing of a pilot licence provides no guarantee that all pilots will perform as trained. Subsequently, the contribution of a trained pilot to the overall safety performance of the system can vary significantly (i.e., the region '*B*' is highly dependent on the individual). Continuation training, performance and currency monitoring is undertaken at the *Operations* level.

There is an additional contribution to the overall safety performance of CPA *Operations*, which arises from the ability of the pilot (*Liveware*) to handle undesired, emergent and unforeseen scenarios that arise due to failures in, or interactions between, the subcomponents of the *Aircraft*, or due to the interaction of the *Aircraft*, *Mission* and *Weather*. This contribution can be viewed as a "Guardian Angel" function. Examples of where the *Liveware* has successfully adapted to accommodate such undesired scenarios would include Air Canada Flight 142, known as the "Gimli Glider" (Williams 2003). As the Gimli Glider example highlights, there is an additional contribution made by the *Liveware* to the overall safety performance of CPA operations. (illustrated as region '*C*' in Figure 2). It is important to note that the component of *Liveware* is not infallible nor can it handle all foreseeable scenarios. As CPA accident data would attest, the component of *Liveware* can be both a mitigator and significant contributor towards CPA accidents. Thus, the region '*C*' represents the net contribution of all pilots to the safety performance of CPA *Operations*. Hypothetically, if safety performance *could* be measured down to the level of individual pilots, then for some pilots the net contribution could be negative.

A significant contributor to the safety performance of CPA is provided by the *Liveware*. This contribution is not due to the safety assurance provided by the existing regulations and standards. This is because the regulatory framework for CPA only provides safety assurance with respect to the individual components of the *Aircraft* (e.g., the separate components of *Machine* and *Liveware*) and not with respect to the system as a "whole". The question of interest in this paper, is whether the same framework of regulations for CPA would result in the same degree of safety performance if applied to UAS. Of particular interest are those UAS which exhibit a high LoA (i.e., where the role of the human in the operation of the UAS is significantly reduced).

2.2 Application to UAS

The high-level safety objective for UAS is that they demonstrate, as a minimum, a level of safety performance equivalent to that currently demonstrated by CPA. This objective is commonly referred to as Equivalent Level of Safety (ELoS). A summary of qualitative and quantitative statements of the ELoS objective is discussed by Clothier & Walker (2013). We can establish the "goal" safety performance for UAS with reference to the "measured or observed" safety performance for CPA, as illustrated in Figure 2. The primary risks associated with the operation of CPA are to those onboard the aircraft (first parties). Whereas for UAS, the primary risks are to those people onboard other aircraft (secondary parties) or to those on the ground (third parties) (Clothier & Walker 2013). The ELoS objective for UAS must therefore be determined in relation to the risks CPA pose to second and third parties.

With reference to the hierarchical model presented in Figure 1, the ELoS objective can be defined at the component of the UAS *Operation*. Regulations on other components of the aviation safety system (e.g., *Operators*) will contribute to the overall safety performance of the system (illustrated by Regions '*D*' and '*H*' in Figure 2). Discussion in this paper is focusses on the safety-contributions provided by the regulations on the components of the *Aircraft*. Thus, the comparison being made in this section assumes the components of *Mission* and *Weather* are constant.[2]

The application of the existing framework of regulations and standards governing the airworthiness of UAS will have a hypothetical contribution to the overall safety performance exhibited by UAS *Operations* (illustrated as region'*E*' in Figure 2). Similarly, training and licensing of the human RP will also provide a degree of assurance in the safety performance of UAS operations (illustrated as region '*F*'). Similar to the pilot of a CPA, the presence of a human RP can also provide a degree of "robustness" to the UAS operation, which contributes to the overall safety performance of the UAS (illustrated as region '*G*'). These contributions ('*E*', '*F*' and '*G*') will depend on the

[2]It is important to note that applying CPA regulations to UAS (at the level of the *Machine*) will not account for potential differences in the nature of the typical *Missions* performed by UAS or differences in the typical environments (*Weather*) in which UAS are operated. For example, High Altitude Long Endurance UAS or UAS operating at low levels in urban environments. (Clothier & Wu 2012) illustrates how the application of existing CPA system reliability requirements to UAS can, in some cases, lead to an unacceptable level of risk at the level of an UAS *Operation*.

LoA of the UAS. The higher the LoA of the UAS, the less the human RP is involved in the operation of the UAS and the more the UAS subsumes the role (i.e., functions) of the RP (Clothier & Walker 2013). It follows that as the LoA increases, the more functions are required to be performed by the *Machine*.

Consider an UAS with a LoA where the RP performs the exact same functions as a human pilot of a CPA, with the only difference being that the RP is remotely located from the Unmanned Aircraft (UA). Despite the 'functions' performed by the RP and conventional pilot being the same, their relative contribution to the safety performance of their respective operations will not be the same. This difference in safety performance is a result of the limited experience in the operation of UAS in civil airspace and as a consequence the training and licensing of RPs may not provide the same degree of safety assurance as that typically provided by the training and licensing of pilots of CPA. There are also human performance considerations unique to UAS (Fothergill et al. 2013) that will impact on the net contribution of the *Liveware* to the overall safety performance of an UAS *Operation*. For example, psychological differences (e.g., trust in autonomy, the lack of "pilot shared fate" with the UA), and differences in RP situational awareness, etc. In sum, even if UAS exhibited the same degree of reliability in the *Machine* and the functions performed by the *Liveware* were the same, the relative contributions and the overall safety performance exhibited is likely to be different for UAS compared to CPA.

Now consider the other extreme, where an UAS has a LoA where the human RP no longer has a role in the flight of the UAS. All of the functions of the *Liveware* are now provided by the component of *Machine*. The safety-contributions previously provided by the *Liveware* (i.e., 'F' and 'G') must now be provided by the *Machine*. Existing software and hardware standard can provide a degree of safety assurance in the *Machine* for those functions for which RP are trained (i.e., 'F'). However, the existing safety contribution due to the adaptability of the RP (i.e., region 'G') would not be assured. Specifically, can the functions previously provided by *Liveware* and now provided by the *Machine* still be adequately performed in the presence of failures in the *Machine* or under unforeseen conditions?

To answer this question, safety assurance must be provided at the component-level of the UAS *Scenarios* in addition to the existing safety assurance mechanisms for the components of *Software* and *Hardware*. New certification tools are needed that can assess the UAS *Scenarios* as a "whole" under varying missions, environmental conditions and failures. The next section (Section 3) describes one possible tool that can be used to provide such a certification assessment.

2.3 Implications for CPA

The discussion of Safety Assurance for CPA reflects current practice. It is important to note that the discussion based on UAS are not limited in applicability to UAS, and it is highly likely that as safety assurance is studied from a UAS context that it will influence and update thinking and practices in CPA.

3 A Proposed Framework for UAS Robust Autonomy Certification

The management of risk of UAS requires quantifying both uncertainty and consequences of outcomes, where the latter refers to behaviour of the system as a whole. While CPA operations may be conducted under similar uncertainty, the previously discussed abilities of the *Liveware* has been the basis for acceptable safety and performance. In UAS, these decisions under uncertainty have shifted from *Liveware* to *Machine*, or some combination of both. The uncertainty is related to three characteristics:

- the actual environmental conditions that the UAS will encounter during the missions (weather and complexity of the operational airspace),
- the reliability of the different components and subsystems of the UAS (airworthiness), and
- the ability of the autonomy to make the rational decisions regarding guidance, navigation and motion control in both normal and anomalous conditions (robust decision making).

All these characteristics are encapsulated in the term *Robust Autonomy*, which describes the ability of an autonomous system to either continue or abort its operation whilst not breaching a minimum level of acceptable safety in the presence of anomalous conditions (Perez et al. 2011*b*, Perez & Williams 2012, Perez et al. 2011*b*). Robust autonomy encompasses both safety and reliability. In addition, the qualifier "*robust*" highlights the feature that the autonomous operation is being considered under both normal and anomalous conditions.

In the absence of a regulatory framework for the certification of highly autonomous UAS, we propose a process based on six steps:

1. Adopt a set of missions for which a UAS is being certified.
2. Adopt the relevant measures of safety and performance and their associated sets of acceptable level.
3. Adopt the envelope of operational conditions within which the missions must be performed.
4. Conduct an evaluation of autonomy to compute probabilities of maintaining acceptable safety and performance.
5. Present a certification case to the NAA.
6. NAA makes a certification decision according to the probabilities and levels of risk deemed acceptable.

The outcomes of Steps 1 to 3 should be a clear set of requirements for particular missions and classes of UAS deemed appropriate for these missions. For example, UAS operations for bush fire monitoring, sea search and rescue, and traffic monitoring over populated areas may all have different requirements and envelopes of weather and environments in which they are allowed to operate. In addition the characteristics of the mission may also have a bearing on the required sensor and actuator physical redundancy. Step 4 requires a probabilistic assessment of safety and performance, which takes into account the reliability of the *Machine* (e.g., the airframe, sensors, and actuators), the likelihood of the operational conditions (e.g., weather, faults, failures, and complexity of the space in which the mission is conducted), and the capabilities of the autonomy to make rational decisions in regard to mission execution, guidance, navigation, communication, and motion control (the equivalent of the *Liveware* in the CPA). Step 4 is expected to

be conducted by third-party testing organisation using methods and tools approved by the NAA. Step 5 involves the operator seeking certification to gather information to put a case forward for the NAA. This will include a certificate from third-party testing organisation. Step 6 involves the decision making by the NAA about certification, which requires the quantification of the consequences of the potential outcomes as perceived by the NAA.

The development of steps 1 to 5 above are to be addressed in collaboration with the NAA. In the following, we discuss specific aspects of each of the steps of the proposed framework above.

3.1 Step 1 - Classification of UAS Missions

UAS are designed for specific operations and environments under which the operations need to be conducted. The objective of Step 1 is to agree with the NAA on a classification of UAS type and its associates class of missions. The class may be related to operations such as search and rescue at sea, bush fire monitoring, border patrol, mineral exploration, *etc.* The classes can be defined by attributes such as mass, sensor redundancy, actuator redundancy, limits on the envelope of environmental conditions for which operations can be conducted, whether operations are to be conducted over populous or non-populous areas, and whether the operations are to be conducted in operational spaces of limited complexity (segregated airspace). Once the classes are defined, the measures of safety and performance associated with each class needs to be determined in Step 2.

3.2 Step 2 - Safety and Performance Indicators

UAS within a particular class should be able to conduct specific operations or missions in prescribed environments. The measures of safety and performance can be evaluated in terms of performance indices r_i $(i = 1, 2, ..., l)$ related to safety requirements of the operation and perhaps also performance attributes that can impact on safety. For example, indices related to safety include separation between aircraft, location with respect to no-fly zones, ability to declare and communicate emergencies, ability to re-route after declaring emergencies, ability to detect and accommodate certain faults, detect and avoid, kinetic energy on emergency landing, *etc.* Performance attributes with bearing on safety are related to the particular aircraft flight envelope. Hence, climb rate, bank angle, loading factor, airspeed, and angle of attack, may be considered in relation to a flight envelope. The adage "Aviate, Navigate, Communicate" can assist in identifying measures of safety and performance, as well as providing a prioritised ordering.

For each quantifiable performance index r_i $(i = 1, 2, ..., l)$, we can associate a set $\mathcal{R}_i$, such that satisfactory performance is attained whenever the value of the index is in the set for the complete mission, namely $r_i \in \mathcal{R}_i$. Then, we can define the *Event of Satisfactory Performance* as such in which a performance index remains inside its region of satisfactory performance for the complete mission:

$$S_i = \{r_i \in \mathcal{R}_i\}, \quad i = 1, 2, ..., l. \tag{1}$$

These events are a logical statement or hypotheses which once a mission is tested can either be true or false.

3.3 Step 3 - Envelope of Operational Conditions

The missions for which the aircraft is being certified are to be conducted under an envelope of operational conditions which encompass weather, aircraft health, and complexity of the airspace. The uncertainty as to which weather condition W_j $(j = 1, 2, ..., m)$ can occur during the mission can be described by the probability distribution $P(W_j|I)$, where I represents background information. These probabilities can be estimated from meteorological data for a particular geographical location and time of the year. Note that the weather conditions to be considered for the operation of the UAS may depend on the type of mission. For example, an UAS used for bush fire monitoring is expected to operate in potentially turbulent high windconditions, whereas a UAS used for aerial spraying of crops may be certified only for operation in light wind conditions.

The UAS may also be subjected to faults, $F_k (k = 0, 1, ..., n)$, which can be associated with the aircraft's actuators, sensors, communication link, changes in dynamics, and the presence of other aircraft in the operational airspace. The condition F_0 denotes the faultless or nominal case (healthy platform operating in anticipated airspace complexity). The uncertainty as to which fault may occur during the mission is described by the probability distribution $P(F_k|I)$. If a fault is associated with a component or a subsystem, for example a servo for a control surface, then $P(F_k|I)$ can be computed from the failure rate function of the component or system (Singpurwalla 2006, Hamada et al. 2010). Faults associated with the complexity of the operational space, $P(F_k|I)$ can be computed from air traffic data or other background information.

3.4 Step 4 - Probabilistic Assessment of Safety and Performance

The work by Perez et al. (2011*b*,*a*) provides a probabilistic assessment of robust autonomy in terms of behaviours. This is motivated by the fact that when pilots are evaluated, it is their behaviour in specific situations which is being assessed rather than their neurophysiological process that leads to the behaviour. A similar procedure can be followed for assessing the rationality of decision making of the autonomy, in which the evaluation is done without specific knowledge of the implementation of autonomous decision making. The potential to evaluate the performance of the autonomy of the *Machine* under failure is critical to providing assurance in the robustness of the UAS (i.e., the additional contribution 'G' in Figure 2.)

The performance during the mission can then be assessed in terms of the predicted probabilities of the events of satisfactory performance S_i as defined in Step 2 (Perez et al. 2011*a*). The hardware that implements the autonomous decisions can be connected to a hardware-in-the-loop (HIL) simulator and tested under the selected set of weather and operational conditions (Step 3). The data, D, collected during a HIL test consists of aircraft motion, location, and also information exchanged over available communication channels. These data can be used to compute the predicted distribution $P(S_i|D, I)$. These probabilities are called *Measures of Robust Autonomy* (Perez et al. 2011*b*) and can be computed via marginalisa-

tion:

$$P(S_i|D,I) = \sum_j \sum_k P(S_i|W_j,F_k,D)\, P(W_j|I)P(F_k|I). \quad (2)$$

Each of these measures involves different aspects of the system which contributes to its total reliability and in turn the overall safety performance of the UAS operation. The distributions $P(W_j|I)$ and $P(F_k|I)$ capture uncertainty about the environment in which the system is to operate. The distribution $P(S_i|W_j,F_k,D),(i=1:l)$ evaluates the quality of autonomous decision making of the UAS under a particular scenario given by the combination W_j,F_k. The latter encompasses aspects of robustness and performance of the vehicle control system, fault detection and diagnosis system, and on-line decisions about reconfiguration of the control system and mission rerouting and trajectory planning.

The distribution $P(S_i|W_j,F_k,D)$ in (2) is related to the concept of *coverage* discussed by Wu (2004), that is, the probability of maintaining a desired level of performance and safety given that a particular scenario has occurred. In the context of this paper, coverage encompasses not only the low-level motion controller but, depending on the degree of autonomy of the platform, also the guidance, communications and navigation systems.

Note that the distribution $P(S_i|W_j,F_k,D)$ is related to the level *Mission* in Figure 1 since the risk measures or performance indices are defined for a set of operations or missions as discussed in Step 2. Also, the weather $P(W_j|I)$ and fault $P(F_k|I)$ distributions are related to the levels of *Aircraft* and *Weather* in Figure 1. Therefore, the measures of robust autonomy $P(S_i|D,I)$ in (2) relate to the level *Scenarios* in Figure 1.

Since the events S_i, defined in (1), are proper hypotheses about the system's behaviour, which can either be true or false, the probability of obtaining a number of successes in a certain number of missions can be modelled using the Bernoulli distribution where π_i is a parameter that that gives the probability of success in one mission. If we collect data D from N replications of evaluation of performance using a hardware-in-the-loop simulator, we can update our prior $p(\pi|I)$ to the posterior

$$p(\pi_i|D,I_{jk}) = \frac{p(D|\pi_i,I_{jk})p(\pi_i|I)}{\int p(D|\pi_i,I_{jk})p(\pi_i|I)\,d\pi_i}, \quad (3)$$

where, $I_{jk} = \{W_j,F_k,I\}$ represents the information related to the particular condition being tested. The posterior densities $p(\pi_i|D,I_{jk})$ encode our uncertainty on the value of π_i under the condition W_j,F_k.

From the posterior $p(\pi_i|D,I_{jk})$, we can obtain the coverage probabilities $P(S_i|D,W_k,F_k,I)$ as predicted probabilities for the success in a single mission (Perez et al. 2011*a*):

$$P(S_i|D,W_k,F_k,I) = \int_0^1 \pi_i\, p(\pi_i|D,I_{jk})\,d\pi_i. \quad (4)$$

In some cases, it may be convenient to have a single figure of merit for robust autonomy. The natural procedure to obtain this figure would be to evaluate the probability that all the indices are jointly within their regions of acceptable performance, namely,

$$\begin{aligned} P(S_1,\ldots,S_l|D,I) \quad (5)\\ &= \sum_j \sum_k P(S_1,\ldots,S_l,W_j,F_k|D,I)\\ &= \sum_j \sum_k P(S_1,\ldots,S_l|W_j,F_k,D)\, P(W_j|I)P(F_k|I). \end{aligned}$$

Details of these computations including a discussion on the choice of prior in (3) according to the principle of maximum entropy can be found in (Perez et al. 2011*a*).

If some indices are not within their region of acceptable performance then it would be expected that they degrade in a manner that meets the priorities established in the adage "Aviate, Navigate, Communicate".

3.5 Step 5 - Present Certification case

The NAA is not likely to conduct Step 4. The assessment can be conducted by an NAA-approved third-party testing organisation. The testing organisation would issue a report on the outcomes of the assessment, which the UAS *Operator* could present to the NAA as part of a case for certification.

Inspiration for the framework proposed in this paper is taken from the offshore shipping industry, where a third-party industry dedicated to assessment has assisted the Classification Societies[3] to develop regulations for testing of behaviours of on-board power management systems and ship motion control systems using hardware in the loop. The services of the assessment industry can also be used by UAS developers to obtain feedback on aspects of autonomy that should be improved. For example, (Perez et al. 2011*b*) discusses a case where information about such an evaluation identifies the need to either improve on fault detection and handling and mission re-planning or reduce the probability of failure of an actuator. From the point of view of design, the framework can suggest areas that may need improvement (for example fault-tolerance, mission re-planning and guidance, or increase the reliability of particular component or sensor to reduce its failure probability), the actual solution to achieve such improvement is not, however, the objective of proposed framework.

3.6 Step 6 - The Certification Decision

Given a case presented for certification, the NAA must make a decision. This decision must be made under uncertainty since the information provided by Step 4 is a probability that the system will attain the required levels of safety and performance, namely (5), mandated by the class under which the UAS seeks certification. Note that this is similar to the decision made as to issuing a licence to a pilot - a decision under uncertainty.

The NAA is likely to compare the measures of robust autonomy with threshold probabilities. To aid in the setting of these thresholds to make such a decision, we can look at normative decision theory (Peterson 2009, Singpurwalla 2006, Berger 1985, Jaynes 2003). In this theory, a decision problem has three key ingredients:

$$\Pi = \langle \mathcal{A}, \mathcal{X}, \mathcal{O} \rangle,$$

[3] The Classification Societies are entities that develop standards and issue certificates for design and construction practices as well as operations of ships and offshore structures. These certificates are required by insurance companies for insuring vessels.

where

- $\mathcal{A} = \{A_1, A_2, \ldots, A_m\}$ is the set of actions being considered,
- $\mathcal{X} = \{X_1, X_2, \ldots, X_n\}$ is the set of states of nature about which there is uncertainty: $P_k = P(X_k)$,
- $\mathcal{O} = \{O_{ij}\}, i = 1, 2, \ldots, m, j = 1, 2, \ldots, n$ is the set outcomes.

To solve a decision problem, we require a *Loss function*, which quantifies the consequences of the outcomes. That is the loss function L maps the states of nature and the actions into numbers that measure the consequences of the outcomes:

$$L : A_i, X_j \mapsto L_{ij}(O_{ij}).$$

That is, L_{ij} measures the consequence of taking the action A_i were X_j be the true state of nature.

A decision can then be made by adopting a *Decision Criterion*, which selects the preferred action based on the consequences of the outcomes and the uncertainty of the states of nature:

$$C : \{L_{ij}, P_k\} \mapsto \ A^\star.$$

For the decision about certification, a simplified decision problem involves a set of a binary actions that the NAA can make:

A_1 - Certify, $\qquad$ A_2 - Require Imrpovements.

The states of nature of interest can be defined as the state in which all the performance indices are jointly satisfied and its complement:

$$X_1 = S_1 \cap S_2 \cap \cdots \cap S_l, \qquad X_2 = \bar{X}_1,$$

and then

$$P_1 = P(X_1) = P(S_1, \ldots, S_l | D, I),$$
$$P_2 = P(X_2) = 1 - P_1.$$

The outcomes are

- O_{11} - Certifying a UAS that will satisfy the safety and performance requirements.
- O_{12} - Certifying a UAS that will not satisfy the safety and performance requirements.
- O_{21} - Not certifying a UAS that will satisfy the safety and performance requirements.
- O_{22} - Not certifying a UAS that will not satisfy the safety and performance requirements.

The loss function in terms of the parameters L_{ij} shown in Table 1. A positive value L_{ij} represents

Table 1: Decision Matrix

	X_1- Safe	X_2-Not safe
A_1- Certify	L_{11}	L_{12}
A_2- Do not certify	L_{21}	L_{22}

a loss, and a negative value represents a gain or a reward. For example,

- L_{11} is a negative number, which reflects the reward for certifying a UAS that satisfies the safety and performance requirements.
- L_{12} is a positive number, which reflects a loss for certifying a UAS that is not safe.
- L_{21} is a positive number, which reflects a loss for denying certification to a safe UAS. (L_{21} is likely to be less than L_{12} as it would be a worse to certify an unreliable UAS compared to denying certification to a worthy UAS.)
- L_{22} is a negative number, which reflects the reward for denying certification to an unsafe UAS.

These four numbers and the scale in which they are measured reflect the the attitude of the NAA (risk proneness or risk averseness). The determination of the these numbers is a complex multidisciplinary task that will require involving not only the NAA, but also industry, social scientists and other subject matter specialists.

At the time of making a decision, whether the true state of nature is X_1 or X_2 will not be known with certainty: only their probabilities P_1 and P_2 will be known through the calculations made in Step 4. The decision criterion that satisfies the requirements of consistency and rationality is that of taking the action that minimises the Bayesian risk (Jaynes 2003). The Bayesian risk is the expected loss over the posterior distribution of the uncertain states of nature at the time of making the decision (Singpurwalla 2006, Berger 1985), namely,

$$A^\star = \arg\min_{A \in \mathcal{A}} E^X[L(A, X)],$$

where $E^X[\cdot]$ denotes the expectation operator with respect the distribution of X. That is, the NAA should take the action that gives the minimum of the two risks:

$$\rho(A_1) = L_{11}\, P_1 + L_{12}\, P_2,$$
$$\rho(A_2) = L_{21}\, P_1 + L_{22}\, P_2.$$

As discussed by Peterson (2009), the choice of expected loss as decision criterion is not to be interpreted as making the decision that minimises the loss in average (as if decisions were made several times). This criterion arises from the satisfaction of axioms of rationality and consistency, from which it follows that rational agents making decisions behave *as if* they minimise the expected loss.

4 Conclusion

This paper discusses a model of the civil aviation safety framework and shows how the current assessment of reliability and risk for conventionally piloted aircraft may not provide an appropriate framework for the same degree of assurance in the safety of Unmanned Aircraft Systems (UAS) operations of varying levels of autonomy. This is because existing approach does not certify the system as a whole. A new framework for certifying UAS, based on the principles of risk management is proposed. This framework arises from combining quantified measures of risk with a probabilistic assessment and normative decision making. The decision making requires quantifying probabilities associated with the measures of risk and also consequences of outcomes related to the behaviour of an UAS. These probabilities are measures of uncertainty about the events of satisfactory safety and performance, and they are computed using a Bayesian approach from an assessment under both nominal and anomalous scenarios described by

faults. The framework poses the decision making as a normative decision problem and solves it in terms of the minimisation of the expected loss - a criterion which satisfies the requirements of consistency and rationality of probability theory.

5 Acknowledgment

This research was in part supported under the Robust Autonomous Systems collaboration between the University of Newcastle and Boeing.

References

Berger, J. (1985), *Statistical Decision Theory and Bayesian Analysis.*, Springer.

CASA (2000), *Aircraft Airworthiness Certification Categories and Designations Explained - Advisory Circular, AC 21.1(1)*, Civil Aviation Safety Authority (CASA), Canberra, Australia.

Clothier, R., Palmer, J., Walker, R. & Fulton, N. (2011), 'Definition of an airworthiness certification framework for civil unmanned aircraft systems,', *Safety Science* **49**(6), 871–885.

Clothier, R., Perez, T. & B. Williams, B. (2013), A review of the concept of autonomy in the context of the safety regulation of civil unmanned aircraft systems, *in* 'Australian System Safety Conference (ASSC2013), Adelaide, Australia.'.

Clothier, R. & Walker, R. (2013), *Safety Risk Management of Unmanned Aircraft Systems*, Springer Science + Business Media B.V., Dordrecht, Netherlands., chapter 3.

Clothier, R., Williams, B., Fulton, N. & Lin, X. (2013), ALARP and the risk management of civil unmanned aircraft systems, *in* 'Australian System Safety Conference (ASSC2013), Adelaide, Australia.'.

Clothier, R. & Wu, P. (2012), A review of system safety failure probability objectives for unmanned aircraft systems, *in* 'Eleventh Probabilistic Safety Assessment and Management Conference (PSAM11) and the Annual European Safety and Reliability Conference (ESREL 2012)', Helsinki, Finland.

Edwards, E. (1972), Man and machine: Systems for safety, *in* 'British Airline Pilots Associations Technical Symposium', British Airline Pilots Associations, London.

Fothergill, S., Clothier, R. & Coyne, J. (2013), Human performance considerations for civil unmanned aircraft systems, *in* 'Australian System Safety Conference (ASSC2013)', Adelaide, Australia.

Haddon, D. & Whittaker, C. (2002), *Aircraft Airworthiness Standards for Civil UAVs*, UK Civil Aviation Authority (CAA), London, UK.

Hamada, M., Wilson, A., Reese, C. & Martz, H. (2010), *Bayesian Reliability*, Springer.

Hayhurst, K., Maddalon, J., Miner, P., DeWalt, M. & McCormick, F. (2006), Unmanned aircraft hazards and their implications for regulation, *in* 'IEEE/AIAA 25th Digital Avionics Systems Conference (DASC)', Portland, OR, United States of America. Portland, OR, USA.

Jaynes, E. (2003), *Probability Theory - The Logic of Science*, Cambridge University Press.

Perez, T. & Williams, B. (2012), Assessment of robust autonomy for unmanned systems – progress and challenges., *in* 'AUVSI's Unmanned Systems North America', Las Vegas, NE, USA.

Perez, T., Williams, B. & de Lamberterie, P. (2011*a*), Computational aspects of probabilistic assessment of uas robust autonomy, *in* '28th International Congress of the Aeronautical Sciences ICAS, Brisbane, Australia.'.

Perez, T., Williams, B. & de Lamberterie, P. (2011*b*), Evaluation of robust autonomy and implications on uas certification and design, *in* '28th International Congress of the Aeronautical Sciences ICAS, Brisbane, Australia.'.

Peterson, M. (2009), *Introduction to Decision Theory*, Cambridge introduction to phylosophy, Cambridge Unversity Press.

Singpurwalla, N. (2006), *Reliability and Risk*, Wiley Series in Probability and Statistics.

Thomé, B. (1993), *Systems Engineering: Principles and Practice of Computer-Based Systems Engineering*, John Wiley & Sons Ltd, New York.

Williams, M. (2003), 'The 156-tonne gimli glider', *Flight Safety Australia* **27**, 22–27.

Wu, N. (2004), 'Coverage in fault-tolerant control', *Automatica* **40**(4), 537–548.

Author Index

Recent Volumes in the CRPIT Series

ISSN 1445-1336

Listed below are some of the latest volumes published in the ACS Series *Conferences in Research and Practice in Information Technology*. The full text of most papers (in either PDF or Postscript format) is available at the series website `http://crpit.com`.

Volume 113 - Computer Science 2011
Edited by Mark Reynolds, The University of Western Australia, Australia. January 2011. 978-1-920682-93-4.
Contains the proceedings of the Thirty-Fourth Australasian Computer Science Conference (ACSC 2011), Perth, Australia, 1720 January 2011.

Volume 114 - Computing Education 2011
Edited by John Hamer, University of Auckland, New Zealand and Michael de Raadt, University of Southern Queensland, Australia. January 2011. 978-1-920682-94-1.
Contains the proceedings of the Thirteenth Australasian Computing Education Conference (ACE 2011), Perth, Australia, 17-20 January 2011.

Volume 115 - Database Technologies 2011
Edited by Heng Tao Shen, The University of Queensland, Australia and Yanchun Zhang, Victoria University, Australia. January 2011. 978-1-920682-95-8.
Contains the proceedings of the Twenty-Second Australasian Database Conference (ADC 2011), Perth, Australia, 17-20 January 2011.

Volume 116 - Information Security 2011
Edited by Colin Boyd, Queensland University of Technology, Australia and Josef Pieprzyk, Macquarie University, Australia. January 2011. 978-1-920682-96-5.
Contains the proceedings of the Ninth Australasian Information Security Conference (AISC 2011), Perth, Australia, 17-20 January 2011.

Volume 117 - User Interfaces 2011
Edited by Christof Lutteroth, University of Auckland, New Zealand and Haifeng Shen, Flinders University, Australia. January 2011. 978-1-920682-97-2.
Contains the proceedings of the Twelfth Australasian User Interface Conference (AUIC2011), Perth, Australia, 17-20 January 2011.

Volume 118 - Parallel and Distributed Computing 2011
Edited by Jinjun Chen, Swinburne University of Technology, Australia and Rajiv Ranjan, University of New South Wales, Australia. January 2011. 978-1-920682-98-9.
Contains the proceedings of the Ninth Australasian Symposium on Parallel and Distributed Computing (AusPDC 2011), Perth, Australia, 17-20 January 2011.

Volume 119 - Theory of Computing 2011
Edited by Alex Potanin, Victoria University of Wellington, New Zealand and Taso Viglas, University of Sydney, Australia. January 2011. 978-1-920682-99-6.
Contains the proceedings of the Seventeenth Computing: The Australasian Theory Symposium (CATS 2011), Perth, Australia, 17-20 January 2011.

Volume 120 - Health Informatics and Knowledge Management 2011
Edited by Kerryn Butler-Henderson, Curtin University, Australia and Tony Sahama, Qeensland University of Technology, Australia. January 2011. 978-1-921770-00-5.
Contains the proceedings of the Fifth Australasian Workshop on Health Informatics and Knowledge Management (HIKM 2011), Perth, Australia, 17-20 January 2011.

Volume 121 - Data Mining and Analytics 2011
Edited by Peter Vamplew, University of Ballarat, Australia, Andrew Stranieri, University of Ballarat, Australia, Kok–Leong Ong, Deakin University, Australia, Peter Christen, Australian National University, , Australia and Paul J. Kennedy, University of Technology, Sydney, Australia. December 2011. 978-1-921770-02-9.
Contains the proceedings of the Ninth Australasian Data Mining Conference (AusDM'11), Ballarat, Australia, 1–2 December 2011.

Volume 122 - Computer Science 2012
Edited by Mark Reynolds, The University of Western Australia, Australia and Bruce Thomas, University of South Australia. January 2012. 978-1-921770-03-6.
Contains the proceedings of the Thirty-Fifth Australasian Computer Science Conference (ACSC 2012), Melbourne, Australia, 30 January – 3 February 2012.

Volume 123 - Computing Education 2012
Edited by Michael de Raadt, Moodle Pty Ltd and Angela Carbone, Monash University, Australia. January 2012. 978-1-921770-04-3.
Contains the proceedings of the Fourteenth Australasian Computing Education Conference (ACE 2012), Melbourne, Australia, 30 January – 3 February 2012.

Volume 124 - Database Technologies 2012
Edited by Rui Zhang, The University of Melbourne, Australia and Yanchun Zhang, Victoria University, Australia. January 2012. 978-1-920682-95-8.
Contains the proceedings of the Twenty-Third Australasian Database Conference (ADC 2012), Melbourne, Australia, 30 January – 3 February 2012.

Volume 125 - Information Security 2012
Edited by Josef Pieprzyk, Macquarie University, Australia and Clark Thomborson, The University of Auckland, New Zealand. January 2012. 978-1-921770-06-7.
Contains the proceedings of the Tenth Australasian Information Security Conference (AISC 2012), Melbourne, Australia, 30 January – 3 February 2012.

Volume 126 - User Interfaces 2012
Edited by Haifeng Shen, Flinders University, Australia and Ross T. Smith, University of South Australia, Australia. January 2012. 978-1-921770-07-4.
Contains the proceedings of the Thirteenth Australasian User Interface Conference (AUIC2012), Melbourne, Australia, 30 January – 3 February 2012.

Volume 127 - Parallel and Distributed Computing 2012
Edited by Jinjun Chen, University of Technology, Sydney, Australia and Rajiv Ranjan, CSIRO ICT Centre, Australia. January 2012. 978-1-921770-08-1.
Contains the proceedings of the Tenth Australasian Symposium on Parallel and Distributed Computing (AusPDC 2012), Melbourne, Australia, 30 January – 3 February 2012.

Volume 128 - Theory of Computing 2012
Edited by Julián Mestre, University of Sydney, Australia. January 2012. 978-1-921770-09-8.
Contains the proceedings of the Eighteenth Computing: The Australasian Theory Symposium (CATS 2012), Melbourne, Australia, 30 January – 3 February 2012.

Volume 129 - Health Informatics and Knowledge Management 2012
Edited by Kerryn Butler-Henderson, Curtin University, Australia and Kathleen Gray, University of Melbourne, Australia. January 2012. 978-1-921770-10-4.
Contains the proceedings of the Fifth Australasian Workshop on Health Informatics and Knowledge Management (HIKM 2012), Melbourne, Australia, 30 January – 3 February 2012.

Volume 130 - Conceptual Modelling 2012
Edited by Aditya Ghose, University of Wollongong, Australia and Flavio Ferrarotti, Victoria University of Wellington, New Zealand. January 2012. 978-1-921770-11-1.
Contains the proceedings of the Eighth Asia-Pacific Conference on Conceptual Modelling (APCCM 2012), Melbourne, Australia, 31 January – 3 February 2012.

Volume 131 - Advances in Ontologies 2010
Edited by Thomas Meyer, UKZN/CSIR Meraka Centre for Artificial Intelligence Research, South Africa, Mehmet Orgun, Macquarie University, Australia and Kerry Taylor, CSIRO ICT Centre, Australia. December 2010. 978-1-921770-00-5.
Contains the proceedings of the Sixth Australasian Ontology Workshop 2010 (AOW 2010), Adelaide, Australia, 7th December 2010.