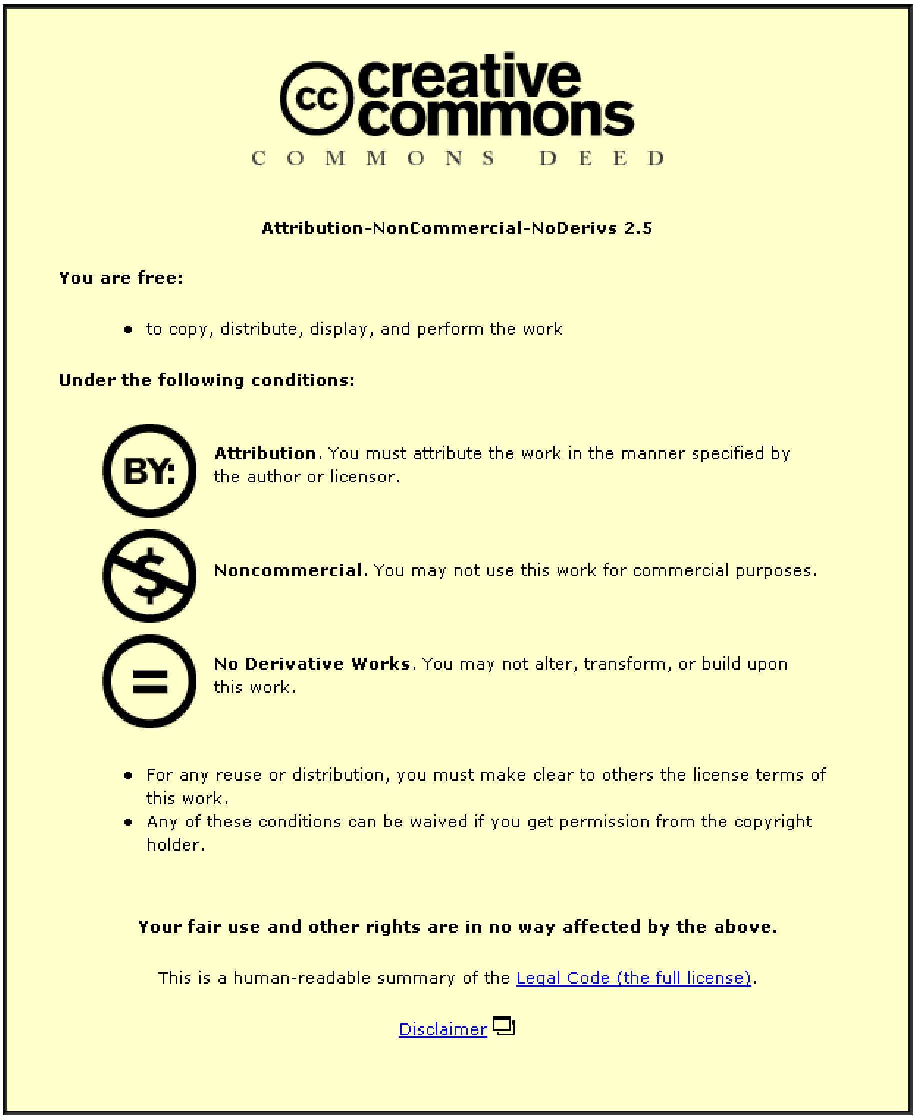


Hazards in Advising Autonomy

Developing Requirements for a Hazard Modelling Methodology Incorporating System Dynamics

Clive G. Downes
R&T, Future Capability
BAE SYSTEMS Military Air & Information
Warton Aerodrome, PR4 1AX, United Kingdom
clive.downes@baesystems.com

Paul W. H. Chung
Department of Computer Science
Loughborough University
Loughborough, LE11 3TU, United Kingdom
P.W.H.Chung@lboro.ac.uk

Abstract— **This paper describes the continuation of a research project to identify and develop tools for the identification and management of hazards likely to arise with the quality and reliability of automatic advice – such as in an automated system advisory function, especially where supporting a "Sense & Avoid" capability as embodied within an airborne autonomous system. An earlier literature survey has been used to map detail onto a Use Case model representing an outline certifiable system development process; thereby helping to identify an appropriate research direction within the broad range of potential end-user requirements. From this direction, an approach has emerged to evaluate hypothetical deviations from declared intent within a behavioral modeling framework to be styled upon Owen's STAMP-Based Hazard Analysis (STPA) [1]. For this approach an outline exemplar describing an air-proximity hazard arising between two air-vehicles has been developed, and the representation of the control structure and system dynamics describing this model are considered. Arising from this model some consideration is then given towards the expression of a more systematic approach in the construction of such models, leading towards new methods to derive safety requirements for implementation within autonomous air systems.**

Keywords- systems engineering; requirements analysis; hazard analysis; autonomous systems; certification; behavioural modelling

I. Introduction and Background

Airworthiness authorities expect Unmanned Aircraft Systems (UAS) to present no greater risk to life than a manned aircraft from hazards arising from unexpected events or airborne failures, and the removal of the pilot alone is unlikely to eliminate the tendency to seek "pilot error" as a possible cause in accidents. Of particular concern to authorities is the possibility of collisions between uninhabited air vehicles and another inhabited aircraft. An autonomous system will be expected to maintain itself in a safe state following any manageable upset event, which must certainly extend to specific critical UAS capabilities – such as the required Sense and Avoid capability for operation in un-segregated airspace.

Consequently, regulations are expected to ensure that such systems demonstrate at least a level of safety equivalent to that of manned aircraft – some go further requiring that "pilot equivalence" be demonstrated; which with interpretation might be expected to have the following implications:

- A capability to manage faults and respond to events while the system is under autonomous control;
- The diagnoses of functional loss and provision of prognoses of likely outcomes against options contingent upon unpredictable events;
- The identification and reduction of all identifiable hazards caused by false alarms and unintended interactions arising from the associated sensing, advisory systems, users and other participants.

The system dynamics of a UAS interacting with other airspace users, and Air Traffic Control, is by definition mathematically complex, exhibiting emergent behavior. Derived aircraft system safety requirements are conventionally formulated within a system safety assessment processes which tend towards a static view of anticipated system safety behavior conditioned only by estimated human and component failure rates. Where complex system dynamics are known to be present, with interactions between system entities not fully understood, and emergent properties likely, then dynamics and safety constraint violation modeling might also be considered.

II. Hazard Modelling Research Programme Scope

It is noted by Allenby, et al [2], that in developing an integrated approach to system safety assessment that there are similarities between FHA (Functional Hazard Assessment) [3] and HAZOP (Hazard & Operation) processes. In considering further the combining of HAZOP and FHA Allenby, et al [2], also state that both methods systematically consider hypothetical deviations from declared intent, whereas safety analysis techniques such as FMEA (Failure Modes and Effects Analysis) assess only the effects of known behavior.

Therefore, so as to aid understanding of the certification system requirements within the context of a larger systems engineering problem, a provisional model based design representation of the aircraft certification process has been constructed, and requirements thereof analyzed [4]. The resultant model serves as the reference point in the understanding of the process of assessing any safety related aircraft function – at least in outline. Following a system requirements analysis, the further developed model, as depicted

This work is supported by BAE SYSTEMS in conjunction with EPSRC, the UK Engineering & Physical Sciences Research Council

in Figure 1, below, represents a relationship within the final selection of functional requirements derived from the literature.

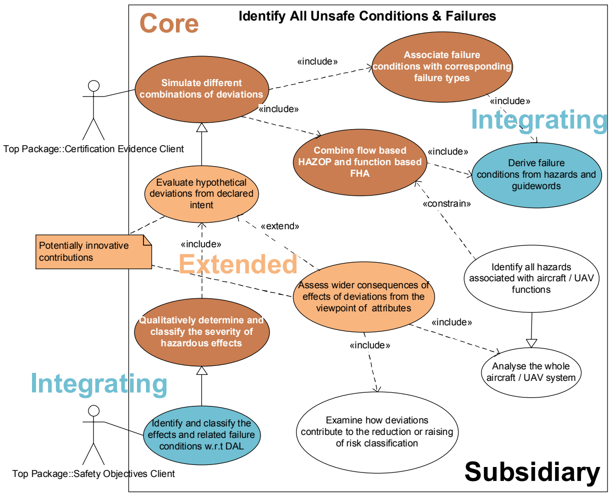

Core
Identify All Unsafe Conditions & Failures
Simulate different combinations of deviations
«include»
Associate failure conditions with corresponding failure types
«include»
«include»
Integrating
Top Package::Certification Evidence Client
Combine flow based HAZOP and function based FHA
«include»
Derive failure conditions from hazards and guidewords
Evaluate hypothetical deviations from declared intent
«constrain»
«extend»
Potentially innovative contributions
«include»
Extended
Identify all hazards associated with aircraft / UAV functions
Assess wider consequences of effects of deviations from the viewpoint of attributes
Qualitatively determine and classify the severity of hazardous effects
«include»
«include»
Analyse the whole aircraft / UAV system
Integrating
Identify and classify the effects and related failure conditions w.r.t DAL
Examine how deviations contribute to the reduction or raising of risk classification
Top Package::Safety Objectives Client
Subsidiary

Figure 1. Scoping a Hazard Modelling Research Programme

A. *Research questions and challenges*

Treating this requirement for a complex system safety assessment method as a Systems Engineering issue, then a number of questions arise. Does a tractable failure analysis methodology exist to predict the consequences of system failure, including false alarms, and unintended interactions within a high integrity autonomous system? What is the nature of any models representing the processes that enforce defensive barriers and safety constraints within the encompassing socio-technical system? And, how might system requirements, design errors, and implementation errors for any system of interest be represented with such a behavioral model?

Assuming that a behavioral modeling approach does prove tractable, then for such a model, how might partitioning across phases of flight, design constraints, annunciation of failure conditions, advisories, recommended actions, and machine autonomous decisions be represented? Also, how might one validate a system hazard and system dynamics model, providing the necessary supporting evidence where these are to be used to obtain a specification of system performance and advisory requirements related to autonomous system behavior? Consequently, what is required from various known safety assessment techniques such as FMEA, HAZOP analysis, and Leveson's STAMP (System Theoretic Accident Model and Processes [5]), to help identify appropriate strategies?

III. HAZARD ANALYSIS

A. *Options for combining methods*

Consequently, ranked pairs of options combining safety assessment methods have been considered, encompassing the various features accommodated by each of the six possible pairings of HAZOP, STAMP, FMEA (itself taken together with FTA – Fault Tree Analysis), and FFA (Function Failure Analysis). Note that FMEA, FTA and FFA together constitute the larger part of FHA, as described in ARP 4761 [3].

B. *Hypothesis and proposal*

Hypothesis – Risk events within complex adaptive systems are caused by deviations from design or operating intentions and unanticipated non-linear casual interactions among system elements that violate safety constraints, therefore, a narrowly focused technique is unlikely to identify all the major problems. Such that:

- The combination of the HAZOP and STAMP related methods will best realize the means to accurately assess these system risk events;
- A meta-model capturing system risk event and system failure behavior can be constructed, unifying the HAZOP and STAMP methodologies;
- The complete coverage of the system risk event space will not be achieved without consideration also of the role of mission phase and probabilistic causes in a unified accident causation model.

In order to confirm, validate, or falsify, this hypothesis it is proposed that a comparison be made ultimately within an exemplar analysis between the "unified" method, proposed above, and a conventional aerospace Functional Hazard Assessment, see ARP 4761 [3]; forming together in effect a complementary combination of safety analysis methods.

IV. HAZARD MODELLING

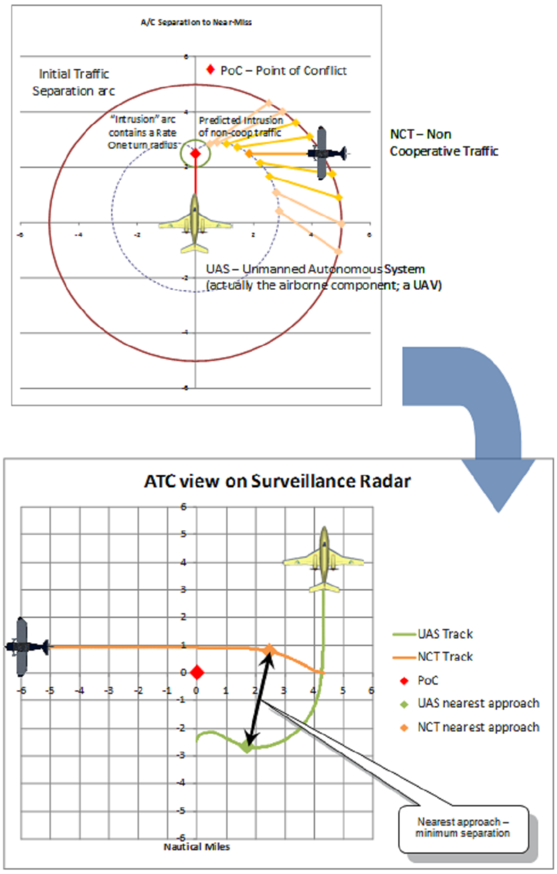


Figure 2. Behavioural Evolution of a Scenario

Pursuant to the objective of developing a “unified” hazard analysis method, eventually also to draw upon the techniques of HAZOP, STAMP [5] and STPA [1, 6], a collision avoidance scenario has been constructed complying with the Rules of the Air [7]. In this interaction certain behaviors have been assigned to each entity. The non-cooperative traffic entity – assumed to be a relatively dumb traffic entity – is assigned only two goals; to maintain its given heading but reconcile this with the goal of doing so within the requirements of the rules of the air. The UAS has been assigned three levels of behavior within the model, and has attributed to it two likely sources of misbehavior. These basic attributes are considered sufficient to investigate the first-cut representational forms for design intent and failure of design intent, along with a representation of hazardous constraint violation – principally a minimum separation constraint, short of actual collision.

V. Causal Loop Modelling

A. *Bottom up interaction modelling*

The systems dynamics model associated with the scenario, as illustrated in Figure 2, is constructed using the causal loop notation developed in the work of Jay W. Forrester, and others, describing dynamic couplings and influences between various system elements and parameters in a form suitable for exploration of a system’s dynamical behavior; a small part is shown in Figure 3. This first prototypical model has been constructed to explore the representational and modeling issues from the bottom-up, although the decomposition within safety assessment methods are themselves top-down. The reason for this is that the model attempts to identify a range of suitable behavioral abstractions applicable to the design intent within the physical reality. Higher level abstractions regarding system intent, operations, and management perceptions and behaviors are to be considered in a later stage of the work, as these then must also fit with models of design intent and physical reality. A further reason to employ bottom-up development is that this is more likely to realize a reusable modeling paradigm.

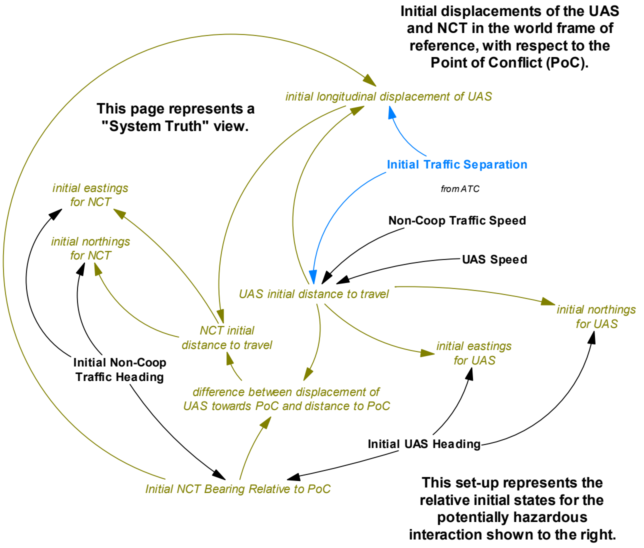

Initial displacements of the UAS and NCT in the world frame of reference, with respect to the Point of Conflict (PoC).
initial longitudinal displacement of UAS
This page represents a "System Truth" view.
Initial Traffic Separation
from ATC
initial eastings for NCT
initial northings for NCT
Non-Coop Traffic Speed
UAS Speed
UAS initial distance to travel
initial northings for UAS
NCT initial distance to travel
initial eastings for UAS
Initial Non-Coop Traffic Heading
difference between displacement of UAS towards PoC and distance to PoC
Initial UAS Heading
Initial NCT Bearing Relative to PoC
This set-up represents the relative initial states for the potentially hazardous interaction shown to the right.

Figure 3. Avoidance Scenario Initial State

At the lowest level of the avoidance scenario the initial set-up for this hazard assumes an imminent breach of horizontal separation, and in the case of head-on or for crossing traffic then the presupposition of a level bust also. It becomes apparent that it would be useful to identify the different categories of interaction:

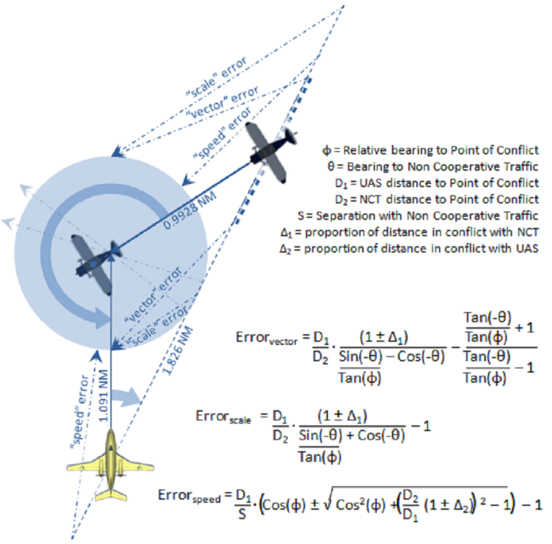
- Fundamental physical and spatial proximity relationships, low-level irreducible behaviors;
- Interactions with other systems (ATC, air vehicles);
- Interactions and derived parameters within the system of interest;
- Parameters derived from fixed external conditions and preconditions;
- Scenario stop events; e.g. Collision.

The intent is to develop a hazard modeling framework as a Model Based Systems Engineering methodology.

B. *Systematic Error Modelling*

The approach taken in the representation of a plausible misbehavior has been to turn the representational problem on its head somewhat – in that the representation is one of how a specific function might act against design intent, rather than a representation of design intent. For example, one obvious act against design intent is to misestimate the separation distance between the two vehicles, as illustrated below in Figure 4. However, such an error will be bounded with the limits of a worst case error in calculating the predicted conflicting track. Rather than build a model of the actual track algorithm, a model of the geometrical plausibility of the track conflict detection is constructed instead, as shown below. Only if an estimate falls within plausible bounds for all of the related systematic errors (ϵ_i) does the model represent certainty of action – in this case emergency avoidance.

$$\forall\, i\ (\epsilon_{\text{lower},\,i} < \text{Estimated Value} < \epsilon_{\text{upper},\,i}) \rightarrow \text{Intended Behaviour}$$



ϕ = Relative bearing to Point of Conflict
θ = Bearing to Non Cooperative Traffic
D_1 = UAS distance to Point of Conflict
D_2 = NCT distance to Point of Conflict
S = Separation with Non Cooperative Traffic
Δ_1 = proportion of distance in conflict with NCT
Δ_2 = proportion of distance in conflict with UAS

$$\text{Error}_{\text{vector}} = \frac{D_1}{D_2}\cdot\frac{(1\pm\Delta_1)}{\frac{\text{Sin}(-\theta)-\text{Cos}(-\theta)}{\text{Tan}(\phi)}} - \frac{\frac{\text{Tan}(-\theta)}{\text{Tan}(\phi)}+1}{\frac{\text{Tan}(-\theta)}{\text{Tan}(\phi)}-1}$$

$$\text{Error}_{\text{scale}} = \frac{D_1}{D_2}\cdot\frac{(1\pm\Delta_1)}{\frac{\text{Sin}(-\theta)+\text{Cos}(-\theta)}{\text{Tan}(\phi)}} - 1$$

$$\text{Error}_{\text{speed}} = \frac{D_1}{S}\cdot\left(\text{Cos}(\phi)\pm\sqrt{\text{Cos}^2(\phi)+\left(\frac{D_2}{D_1}(1\pm\Delta_2)\right)^2-1}\right)-1$$

Figure 4. Misestimating Separation

Within the overall dynamical behavioral model this then describes a plausible and missed conflict detection model, capturing a particular likely failure of design intent. The effects of this can be dynamically explored by varying the degree of misestimating – right out to the most implausible of limits; for example a negative estimate placing the conflicting vehicle on the wrong side of the encounter. Therefore, this model defines the outer limits to be applied to errors within which plausible detection will occur. If the actual given error is outside of the range of any of these error limits then the detection of the threat from the conflicting traffic is deemed as unlikely to occur – with the likely consequence being that the UAS vehicle continues upon its present heading, or otherwise as dictated by the Rules of the Air.

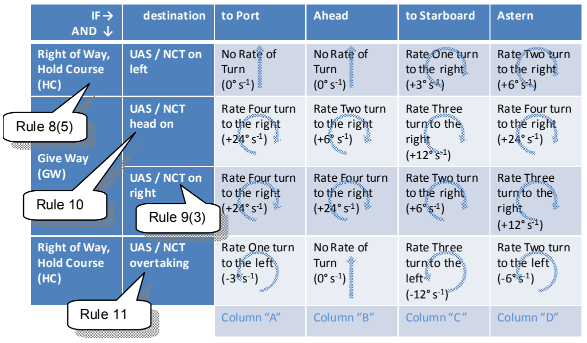
This misestimating of the distance to the threatening traffic, as perceived by the system, is only one plausible source of error. Another likely source of misbehavior lies also in the delay in responding to the threatening traffic. Therefore, a simple information delay model is included in this example.

C. Rule Modelling

Another requirement of a model capturing design intent is to consider the incorporation of operational rules; and at higher levels of abstraction design rules also. In this case these are the Rules of the Air Regulations 2007, General Flight Rules [7]:

- Avoiding aerial collisions – 8(5) … an aircraft which has the right-of-way under this rule shall maintain its course and speed.
- Converging – 9(3) … when two aircraft are converging in the air at approximately the same altitude, the aircraft which has the other on its right shall give way.
- Approaching head-on – 10 When two aircraft are approaching head-on, or approximately so, in the air and there is a danger of collision, each shall alter its course to the right.
- Overtaking – 11(1) … an aircraft which is being overtaken in the air shall have the right-of-way and the overtaking aircraft, whether climbing, descending or in horizontal flight, shall keep out of the way of the other aircraft by altering course to the right.

These rules, Figure 5, have been implemented within a simple two input fuzzy controller using causal loop notation.



IF → AND ↓	destination	to Port	Ahead	to Starboard	Astern
Right of Way, Hold Course (HC)	UAS / NCT on left	No Rate of Turn (0° s^{-1})	No Rate of Turn (0° s^{-1})	Rate One turn to the right (+3° s^{-1})	Rate Two turn to the right (+6° s^{-1})
Give Way (GW)	UAS / NCT head on	Rate Four turn to the right (+24° s^{-1})	Rate Two turn to the right (+6° s^{-1})	Rate Three turn to the right (+12° s^{-1})	Rate Four turn to the right (+24° s^{-1})
	UAS / NCT on right	Rate Four turn to the right (+24° s^{-1})	Rate Four turn to the right (+24° s^{-1})	Rate Two turn to the right (+6° s^{-1})	Rate Three turn to the right (+12° s^{-1})
Right of Way, Hold Course (HC)	UAS / NCT overtaking	Rate One turn to the left (-3° s^{-1})	No Rate of Turn (0° s^{-1})	Rate Three turn to the left (-12° s^{-1})	Rate Two turn to the left (-6° s^{-1})
		Column "A"	Column "B"	Column "C"	Column "D"

Figure 5. Premises for Rules 8, 9, 10 & 11

D. Behavioural Modelling

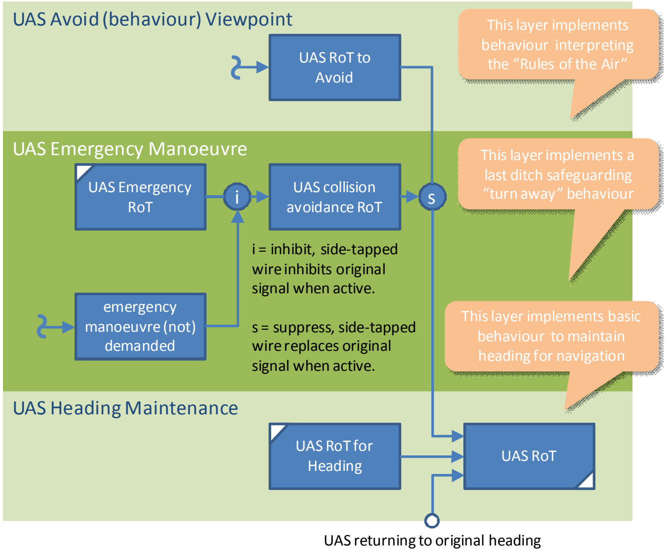

UAS Avoid (behaviour) Viewpoint
UAS RoT to Avoid
This layer implements behaviour interpreting the "Rules of the Air"
UAS Emergency Manoeuvre
UAS Emergency RoT
i
UAS collision avoidance RoT
s
This layer implements a last ditch safeguarding "turn away" behaviour
i = inhibit, side-tapped wire inhibits original signal when active.
emergency manoeuvre (not) demanded
s = suppress, side-tapped wire replaces original signal when active.
This layer implements basic behaviour to maintain heading for navigation
UAS Heading Maintenance
UAS RoT for Heading
UAS RoT
UAS returning to original heading

Figure 6. Avoidance Behavior as Subsumption Architecture

A further consideration is the requirement to capture design intent with respect to behavior. For this it is proposed that a suitable model might be found in the field of behavior-based robotics, for example the Subsumption Architecture, after Brooks [8], and as illustrated in Figure 6; at least for machine based behavioral intent. In this example, collision avoidance has been modeled as behavior-based robotics architecture – illustrating here the suppression of the Rules of the Air in the event of an emergency behavior to avoid an imminent collision with threatening traffic. Whilst this representation may not model the actual software implementation, it might be argued that this does capture design intent, and instead be used to capture a behavioral requirement rather than a design solution. Again, this model might be represented with system dynamics model prototype in causal loop notation, as shown in Figure 7.

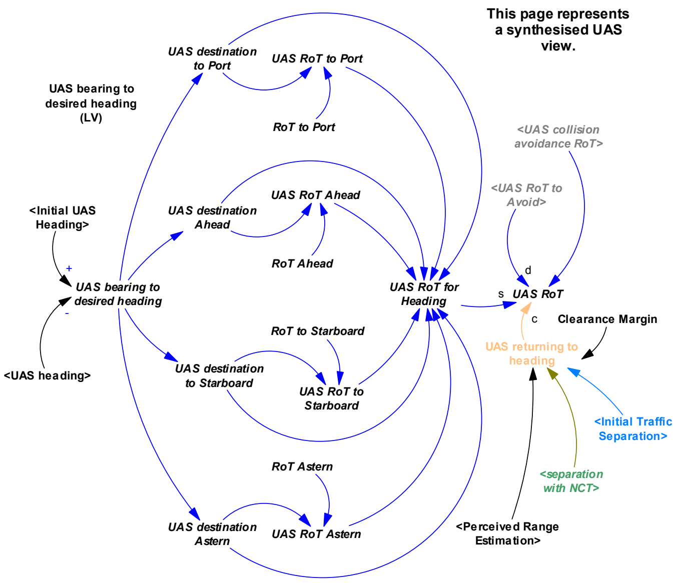


Figure 7. Incorporating Rules and Behaviour

VI. Emergence

In terms of exploring design intent within a complex system, that is one likely to exhibit emergent behavior; it might be argued that is sufficient to attempt to correctly capture characteristic behavior. That is capturing the characteristics of both the desired and undesired behavior, and the behavior arising from the design intent and the emergent properties – ever-present within complex systems interacting within an uncertain non-static environment. This, rather than exact system behavior, might better typify how we then expect a partial system to behave, such that to this we may then later add additional controlling interactions and higher levels of behavior to explore. For example, Figures 8 and 9 both represent resultant dynamic behavior from the model demonstrating the characteristic behavior of this interactive system; constrained to two spatial dimensions. Within the terms of the "Rules of the Air" both of these represent legal characteristic behaviors – notwithstanding the actual degree of separation loss. In both cases each vehicle executes a maneuver in a legal manner.

However, due to the continuously changing relative positions and relative bearings of the two vehicles, and the further interaction of the internal competing behavioral models, a variety of emergent behaviors have been observed; some of which might not be so obvious or as originally conceived.

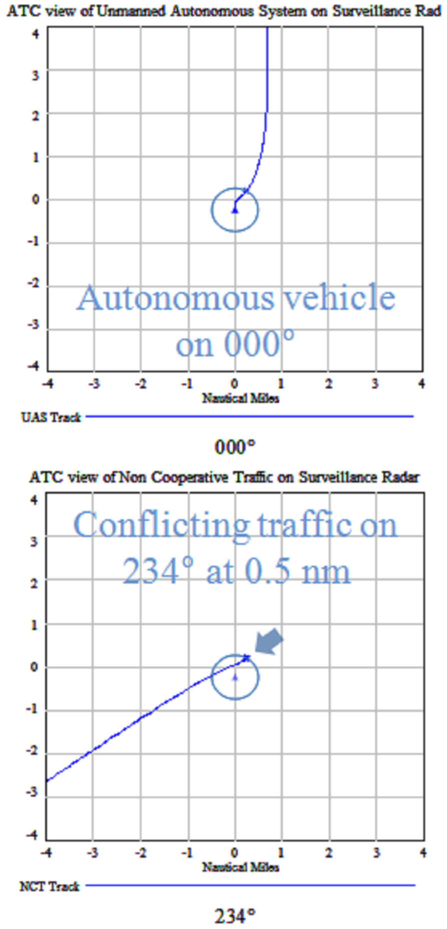


Figure 8. Interactions for Initial Separation at 234° & 0.5 nm

For example an expected emergent behavior is observed at a "break out" point where the UAV elects to turn left to pass behind the conflicting traffic coming from the right – near to 238.5° in the 0.5 nm case. This ensures that a collision does not occur and fulfills at least a basic requirement of the rules of the air to pass behind crossing traffic from the right. Less expected is an entrainment of the vehicles that can occur when arriving at parallel tracks and travelling at nearly the same speed.

For this encounter and scenario, the minimum separation falls to within 449.32 ft, which could be considered an unsafe air proximity event. Figure 8 represents this close but marginally unsafe interaction, whilst directly applying the rules of the air. Figure 9 demonstrates a phase transition due to a further internal interaction occurring with the UAS emergency maneuver behavior, giving a closest approach of 1787.84 ft.

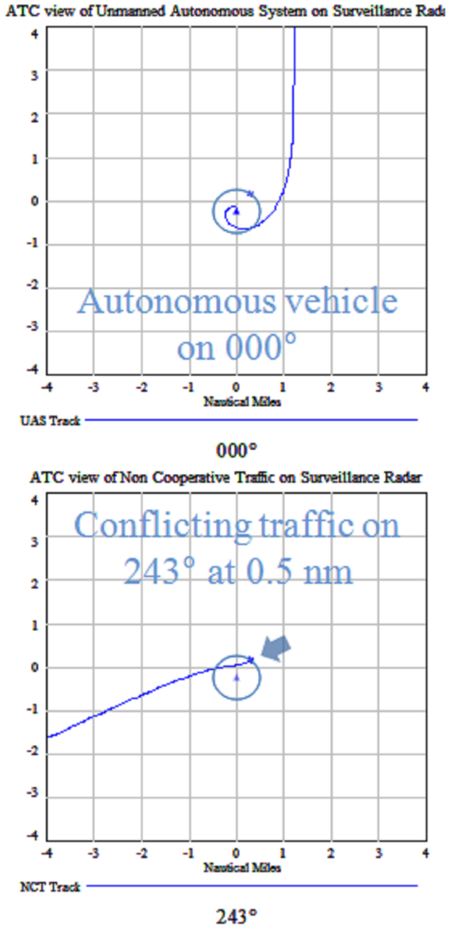


Figure 9. Interactions for Initial Separation at 243° & 0.5 nm

VII. Extending the Hazard Modelling

The modeling of the overall effect upon hazards, such as the violation of separation constraint, is a complex problem and a means of obtaining insight is to model the system dynamics. In this example, the 0.5 nm scenario at 94.25 kts exhibits five distinct dynamic behaviors dependent upon the relative heading: (A) entrainment (near the same speed); (B) direct compliant avoidance; (C) compliant avoidance with reduced separation; (D) loss of safe separation (insufficient space for right turn); and (E) suppressed Rules of the Air; in Figure 10.

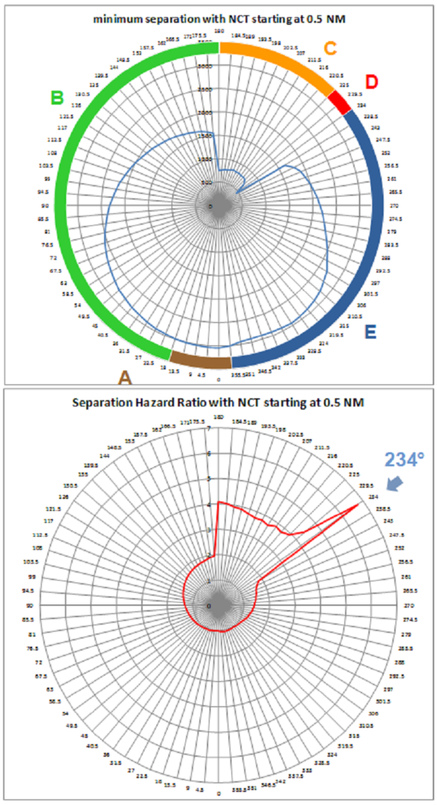


Figure 10. Separation as Constraint Violation

However, the "quality" of the design intent of such a system is as likely to be judged on the possibilities of generating an air proximity arising as it is for speculative probabilistic collision rates. Therefore the constraint to be modeled might more appropriately relate to this possibility, wherein different categories of potential hazard are defined with categorized causal factors such as those given in Table 1.

TABLE I. DOMINANT AIRPROX CAUSAL FACTORS [9]

Air Proximity Causal Factors	
Cause	***Attributed to***
Did not separate / poor judgement	Controller
Climbed / descended through assigned level	Pilot
Did not see conflicting traffic	Pilot
Inadequate avoiding action / flew too close	Pilot
Late sighting of conflicting traffic	Pilot
Misinterpretation of ATC message	Pilot
Not obeying orders / following advice from ATC	Pilot
Penetration of CAS/SRZ/ATZ without clearance	Pilot
Conflict in other type of airspace	Other
FIR conflict	Other
Sighting Report	Other

VIII. Conclusions

Therefore further work is yet to be undertaken to formulate a representation to encompass the more complex view of a system incorporating air-proximity constraints, and later maintenance process models. These are to be layered upon the current physical model layer, and developed into representational models in SysML. The intent is to initiate the further development of tools to investigate cause & hazard effect sensitivity scenarios – and to support trade studies. This would include the facilitation of the free exploration of "What-if" hypotheses – enabling investigations with multiple cause analysis, and anomalous hazardous event sequencing, false alarm, and misdiagnosis coupling; advisory and diagnostic function discrimination performance determination – supporting performance trade-off against Design Assurance Level; and hopefully the possibility of re-usable and extendable hazard model modules. The challenge is to find appropriate, or better yet, correct paradigms for behavioral modeling of system hazards.

Acknowledgment

The authors wish to thank BAE SYSTEMS and EPSRC for their support of this project, operating through the Systems Engineering Doctorate Centre at Loughborough University.

References

[1] B. D. Owens, M. S. Herring, N. Dulac, N. G. Leveson, M. D. Ingham, and K. A. Weiss, "Application of a Safety-Driven Design Methodology to an Outer Planet Exploration Mission," in Aerospace Conference, 2008 IEEE, 2008, pp. 1-24.

[2] K. Allenby and T. Kelly, "Deriving safety requirements using scenarios," in Requirements Engineering, 2001. Proceedings. Fifth IEEE International Symposium on, 2001, pp. 228-235.

[3] ARP 4761 - Guidelines and Methods for Conducting the Safety Assessment Process on Civil Airborne Systems and Equipment. Warrendale, Pennsylvania, USA: Society of Automotive Engineers, SAE International, 1996.

[4] C. G. Downes, P. W. H. Chung, and A. Morris, "Hazards in advising autonomy: A structured approach seeking novelty in developing the requirements for an exemplar," in System of Systems Engineering (SoSE), 2010 5th International Conference on, 2010, pp. 1-7.

[5] N. G. Leveson, "A systems-theoretic approach to safety in software-intensive systems," Dependable and Secure Computing, IEEE Transactions on, vol. 1, pp. 66-86, 2004.

[6] M. V. Stringfellow, N. G. Leveson, and B. D. Owens, "Safety-Driven Design for Software-Intensive Aerospace and Automotive Systems," Proceedings of the IEEE, vol. 98, pp. 515-525, 2010.

[7] CAP 393 - Air Navigation: The Order and the Regulations. London, UK: TSO (The Stationery Office) on behalf of the UK Civil Aviation Authority, Safety Regulation Group, 2010.

[8] R. A. Brooks, Cambrian Intelligence: The Early History of the New AI: Bradford Books, The MIT Press, 1999.

[9] "Twenty Second Report by the UK Airprox Board (January 2009 to June 2009)," Analysis of Airprox in UK Airspace, vol. 22, 2009.