Safety Science 92 (2017) 53–65

Contents lists available at ScienceDirect

Safety Science

journal homepage: www.elsevier.com/locate/ssci

Review

An investigation of proposed techniques for quantifying confidence in assurance arguments

Patrick J. Graydon *, C. Michael Holloway

Mail Stop 130, NASA Langley Research Center, Hampton, VA 23681, United States

ARTICLE INFO

Article history:
Received 10 May 2016
Received in revised form 29 August 2016
Accepted 25 September 2016

Keywords:
Safety case
Assurance argument
Confidence
Uncertainty

ABSTRACT

The use of safety cases in certification raises the question of assurance argument sufficiency and the issue of confidence (or uncertainty) in the argument's claims. Some researchers propose to model confidence quantitatively and to calculate confidence in argument conclusions. We know of little evidence to suggest that any proposed technique would deliver trustworthy results when implemented by system safety practitioners. Proponents do not usually assess the efficacy of their techniques through controlled experiment or historical study. Instead, they present an illustrative example where the calculation delivers a plausible result. In this paper, we review current proposals, claims made about them, and evidence advanced in favor of them. We then show that proposed techniques can deliver implausible results in some cases. We conclude that quantitative confidence techniques require further validation before they should be recommended as part of the basis for deciding whether an assurance argument justifies fielding a critical system.

Contents

1. Introduction . . . 54
2. Background . . . 54
   2.1. Safety cases . . . 54
   2.2. The philosophy literature . . . 55
3. Method . . . 55
   3.1. Selection of proposals . . . 55
   3.2. Assessment of proposals . . . 55
4. Results . . . 56
   4.1. The proposed techniques . . . 56
      4.1.1. Techniques based on Bayesian Belief Networks . . . 56
      4.1.2. Techniques Based on Dempster–Shafer Theory, Jøsang's Opinion Triangle, or Evidential Reasoning . . . 57
      4.1.3. Other techniques . . . 57
   4.2. Hypotheses . . . 57
      4.2.1. The purposes of confidence assessment . . . 57
      4.2.2. The maturity of the proposed techniques . . . 58
      4.2.3. Helping analysts make more accurate assessments . . . 58
      4.2.4. Communicating the safety argument . . . 58
   4.3. Evidence of fitness for use in release-to-service decisions . . . 58
      4.3.1. Properties of the underlying theory . . . 58
      4.3.2. The Cyra and Górski experiment . . . 59
      4.3.3. The Nair et al. survey . . . 59
   4.4. Counterargument . . . 59
      4.4.1. Masking missing evidence or counterevidence . . . 59

* Corresponding author.
E-mail address: patrick.j.graydon@nasa.gov (P.J. Graydon).

http://dx.doi.org/10.1016/j.ssci.2016.09.014
0925-7535/

4.4.2. Sensitivity to the arbitrary scope of hazards . . . 60
4.4.3. Technique-specific counterexamples . . . 61
4.4.4. Other issues . . . 62
5. Discussion . . . 62
5.1. Don't let the perfect be the enemy of the good . . . 63
5.2. There is no alternative . . . 63
5.3. Overlooking the counterexamples . . . 63
5.4. Quantification is systematic . . . 63
5.5. Other work . . . 63
6. Conclusion . . . 63
Acknowledgment . . . 63
Appendix A. Example Dempster–Shafer calculations . . . 63
References . . . 64

1. Introduction

The *safety case approach* has been used in diverse industries and regulatory domains—e.g., nuclear, manufacturing, oil and gas, rail, aviation, automotive, medical devices, and defense—in some cases for many years (United States Chemical Safety and Hazard Investigation Board, 2014; Office for Nuclear Regulation, 2013; Cullen, 2001; CAP 670, 2010; ISO 26262, 2011; Center for Devices and Radioogical Health (CDRH), 2014; Defence Standard 00-56 and Issue 5, 2014). An organization using the approach takes ownership of the risks to be controlled by adopting an appropriate safety management system, performing a hazard assessment, selecting appropriate controls, and implementing these. The main difference between the safety case approach and other systems safety approaches is the use of a *safety case* to document hazards, controls, and the controls' adequacy (A-P-T Research, Inc., 2014). A safety case combines safety evidence such as fault tree analysis results and test reports with an *assurance argument*, typically defined as "a reasoned and compelling argument . . .that a system, service or organisation will operate as intended for a defined application in a defined environment" (Attwood, 2011). A safety case might serve many purposes. For example, a safety case might communicate the system safety rationale to engineers who will later modify the system. Alternatively, a safety case might explain the safety rationale and evidence to an assessor who must decide whether the hazard controls are adequate. Such use raises a question: how should the assessor determine whether the argument and its evidence are sufficient?

The question of assurance argument sufficiency leads to the concepts of *confidence* and *uncertainty* in the argument's claims. While questions of confidence might be asked of all kinds of assurance cases, research on confidence in assurance argumentation often focuses on the software-intensive, safety-critical systems that are our primary interest. Researchers have defined methods for reviewing assurance arguments (Graydon et al., 2010; Kelly, 2007) and means of associating reasoning about confidence with the parts of the assurance argument they relate to Goodenough et al. (2013) and Hawkins et al. (2011). Other researchers propose adopting quantitative models of argument from disciplines such as philosophy to the problem of assessing confidence in assurance arguments. Some vendors sell tools to perform the necessary calculations (Argevide). But despite the importance of knowing how far confidence estimates should be trusted, little is known about whether proposed techniques for quantifying confidence produce trustworthy results (Graydon, 2013). And as others have observed, the seductive appearance of computational rigor might cause decision-makers to mistakenly place trust in "superficially plausible nonsense" (Littlewood, 2005). A frank appraisal of the evidence for and against the efficacy of proposed quantitative confidence techniques will be of value to safety engineers, assessors, and regulators who must decide how to assess safety arguments and interpret assessments. In this paper, we survey and assess proposed techniques for quantifying confidence in assurance arguments. We identify the proposers' claims and the support given for these, provide specific counterarguments, identify common flaws in the proposals, and assess the evidential basis for quantifying confidence.

2. Background

There is a substantial literature on safety cases, a much larger philosophical literature on argument, and a growing body of work on applying ideas from the latter to the former.

2.1. Safety cases

In the 1970s, the United Kingdom (UK) Committee on Safety and Health at Work observed that prescribing specific risk reduction measures had not ensured safety in diverse workplaces for two reasons (Lord Robens et al., 1972). First, prescription encouraged compliance without thought, resulting in missed opportunities for risk reduction. Second, making law or regulation takes so much time that prescriptions were often out of date before or shortly after they took effect. (This is still true four decades later (Moran et al., 2012).) Accordingly, the UK introduced the safety case process to compel operators to conduct risk assessments, implement appropriate mitigations, adopt an appropriate safety management plan, commission independent audits to verify effective safety management, and revisit safety as circumstances, operations, and technology change (Cullen, 1990; Lord Robens et al., 1972; United States Chemical Safety and Hazard Investigation Board, 2014). In later decades, the safety case process was expanded to applications such as offshore oil and gas installations (Cullen, 1990) and railway operations (Cullen, 2001). The safety case process is now used in the oil and gas sector in the UK and Australia, and a similar process is used in Norway (United States Chemical Safety and Hazard Investigation Board, 2014). Safety cases are used in the UK defense sector (Defence Standard 00-56 and Issue 5, 2014), in automotive applications (ISO 26262, 2011), and with some medical devices in the United States (Center for Devices and Radioogical Health (CDRH), 2014).

Safety cases and their assurance arguments are thought to serve multiple purposes. For example, safety cases communicate safety design intent to those who will modify existing systems so that safety can be maintained during and after the change (Defence Standard 00-56 and Issue 5, 2014). Safety cases also explain the safety rationale and evidence to an assessor—a customer, regulatory agency, or third party—who uses that information to decide whether a system is adequately safe (Ayoub et al., 2013; Duan