Analysis of Human Factors Failures in an Incident of Self-Driving Car Accident

Ashraf Labib[1*], Yoskue Nagase[1], Sara Hadleigh-Dunn[1]

[1] University of Portsmouth, United Kingdom

Abstract There are always submerged risks involved with advanced technology; therefore, it is necessary for policymakers, inventors and technology companies to scrutinise potential risks when they consider implementing new technology. This paper attempts to extract generic lessons from a failure relevant to autonomous transport systems. We use fault tree analysis (FTA), a reliability block diagram (RBD) approach and failure mode and effect analysis (FMEA), for analysing a fatal pedestrian accident caused by a level-3 self-driving car in 2018. The work highlights the importance of prematurity of test driving self-driving cars on public roads and the potential of an insightful analysis method that can capture human factors. In this work we theorise accident reporting systems, and provide a framework for triple loop learning.

Keywords: Learning from Failures, Self-Driving Car, Accident Analysis

1. Introduction: Learning from failures and advanced technology

While it is hoped that technology makes our society more comfortable, more convenient and safer, risks exist in new technologies and tragic failures associated with technologies have occurred. However, in the world of engineering, learning from failures is indispensable, and it is known that various failures have helped societies evolve. Leoncini (2016) argues that learning-by-failing is a necessity as the very nature of uncertainty that is associated with innovative activities, can lead to failures. Subsequently, learning and failure are closely connected due to the fact that trial-and-error procedures are among the main elements of discovery (Chesbrough, 2010). Also Leoncini (2016) suggests that the learning process is more focused in the case of failure than it would be in the case of success. In this context the learning process is defined as a search process intended to rectify an organisational behaviour that led to a failure.

Furthermore, there is a shift from micro to macro level of emphasis that also includes a shift from technical to human to socio. This led to an interest among researchers in constructing hierarchy of causes of failures, as originally proposed in the Accimap model, proposed by Rasmussen (1997), and other variations of this mental models of hierarchies, and networks, of causal factors such as Hierarchy of Failures (Hatamura, 2015), STAMP (Leveson, 2004), FRAM (Hollnagel, 2016). Such approaches in framing causal factors can support policymakers, inventors or corporations when considering the implementation of a new technology. Correspondingly, this paper attempts to examine policies and recommendations related to recent technological failures with respect to implementation of advanced technology in an evolving hybrid modeling approach.

Autonomous transport systems (ATSs), such as self-driving cars and autonomous surface ships, attract attention as a solution for achieving sustainable development goals (Lim & Taeihagh, 2018). Moreover, it is hoped that these technologies will contribute to the maintenance of national wealth in developed countries that will soon face an aging and diminishing population (Department for Transport, the UK Government, 2015; Cabinet Secretariat, Government of Japan, 2012). In the US, some states have allowed social implementation tests of independent-type autonomous vehicles on public roads, and consequently, the number of accidents is increasing (Nakata, 2018). Accordingly, the current paper examines the failures in a fatal accident using fault tree analysis (FTA), a reliability block diagram (RBD) approach and failure mode effect analysis (FMEA).

The paper is organised as follows: Section 2 reviews the theoretical literature on the subject; Section 3 describes the first case study data on self-driving car accident and demonstrates the use of modelling techniques of FTA, RBD, and FMEA; Section 4 includes some concluding remarks.

2. Literature Review and Theorization related to Triple Loop Learning from Major Incidents:

Barriers to Advanced Technology Adoptions. Barriers to advanced technology adoptions have been discussed in terms of its relation to research policy in different countries (Baldwin & Lin, 2002 ; D'Este et al, 2012; Galia and Legros, 2004; Lhuillery and Pfister, 2009). However none of them applied root cause analysis methods to assess casual factors and analyse vulnerabilities in the system as investigated in this paper.

Accident Analysis Methods. In terms of accident analysis methods and accident causation, there is a systematic review (Hulme, et al, 2019) that examined literature related to accident analysis methods related to sociotechnical systems contexts, mainly; AcciMap, the Human Factors Analysis and Classification System (HFACS), the Systems Theoretic Accident Model and Processes (STAMP) method, and the Functional Resonance Analysis Method (FRAM). The authors have identified the need to upgrade incident reporting systems, or recommendation reports, and have encouraged exploring further opportunities around the development of novel accident analysis approaches, which is also in line with the recommendations by Goode et al (2018).

Learning Loops. The grounding of safety and security in the learning loops is under researched in the literature. Hence we are in agreement with the literature review in the field of safety carried out by Drupsteen and Guldenmund (2014) where they concluded that how learning occurs had been rarely studied, and suggested that safety research 'would benefit from input from organizational learning theories, such as Argyris and Schön's (1978) models of single and double-loop learning' (Drupsteen & Guldenmund, 2014, pp94). We extend their argument into the benefits of triple loop learning in safety and security arguing that safety and security are similar in their dealing with prevention and management of hazardous incidents or threats, where the main difference is the intent. Hence, we are also in agreement with the observation of Aven (2007) related to

the growing interest in applying risk analysis and risk management not only to safety but also to security problems.

In order to conceptualize triple loop learning, we need first to understand the literature related to double-loop learning. The concept of organisational learning can be described as a dichotomy (Cope 2003; Tosey et al 2012;). Single loop learning occurs `whenever an error is detected and corrected without questioning or altering the underlying values of the system', and double loop learning occurs `when mismatches are corrected by first examining and altering the governing variables and then the actions', as defined by Argyris (1999). Hence, single loop concerns preserving and improving status quo, whereas, second loop learning implies changing the status quo itself (Labib, 2016). Accordingly, by extending this logic, triple loop learning can then be described as a `deeper', or `higher', level than, primary and secondary forms of learning, which implicitly means that this level has greater impact. Yet, as noted by Tosey et al, (2012) in spite of its perceived importance, conceptualisations of this level of learning do not always make clear how it differs from, or relates to, primary or secondary forms.

One of the most comprehensive conceptualizations of the organizational triple loop learning can be found in Tosey et al. (2012) in their paper titled 'The origins and conceptualisations of 'triple-loop' learning: a critical review'. They distinguish between three conceptualizations of 'triple-loop learning'. They offer three conceptualisations of triple-loop learning, viz; a) a level superior to single and double-loop learning, a form of shift from operations to strategy; b) a level that involves reflexivity on learning how to learn about the previous two levels i.e. learning about the process of learning; c) a level that involves a change of epistemology; a change in the wisdom in the form of knowing and learning. This third conceptualisation is about a complete, or fundamental, change of belief and opinion. The triple loop learning is about a shift towards 'richness'; as Weick puts it succinctly: 'it takes richness to grasp richness' (Weick, 2007).

In this paper we provide tools and case studies as enablers for realizing second and third loops of learning from failures. In doing so, it is hoped to extend such triple loop conceptualization to both the safety and security fields.

3. Case Study: A self-driving-car accident

On 18 March 2018 in Arizona, for the first time, a self-driving car killed a pedestrian (Wakabayashi, 2018). Recently, some authorities in the US have allowed self-driving tests on public roads, and accidents caused by these vehicles have increased, however, the fatalities in these accidents had been restricted to the drivers. Although there was only one fatality, it is possible that fatal accidents during tests will discourage people and policy makers from accepting this technology. Hence, learning from an accident is a meaningful process for progressing science and technology policy.

3.1 Overview of the accident

The fatal pedestrian accident happened at around 10 p.m. on a road in Tempe, Arizona (Griggs & Wakabayashi, 2018; NTSB, 2018). The female victim was walking her bicycle across the road when the vehicle in autonomous mode struck her at about 45

miles/hr. The vehicle was operated by Uber Technology Inc. and equipped with their original light detection and ranging (or LIDAR) system that enabled the vehicle to drive at night. However, according to the released dashboard-camera recording, the system did not work very well and the driver inside the vehicle looked away without taking evasive action. Moreover, the US National Transport Safety Board (NTSB, 2018) reported that although the system detected an object six seconds before the accident, the system needed emergency braking manoeuvre 1.3 seconds before the accident when the system determined it was a bicycle, and the car was not designed to reduce speed and alert the driver while under computer control.

3.2 The accident factors and an analysis

First, to confirm the accident factors, the current paper used FTA and an RBD approach (Figures 4 and 5). Using an overview of the accident, the causes were divided into two types: the self-driving car's faults and the pedestrian's faults. The self-driving car's faults were further divided into machine errors and human errors. Finally, the machine errors were classified as being derived from inherent defects (e.g. computer programme, code or mechanical structure) or acquired defects (breakdown), and the human errors were classified as being derived from external factors (invisibility because of light or weather conditions) and internal factors (carelessness or intention). However, a 'basic event 5' in Figure 4 could be removed as it is difficult to determine the next action in a sudden event, as in this case, and it is also unreasonable in this situation that a sensing system could be expected to work appropriately. Moreover, the pedestrian's faults could include intentional and non-intentional behaviours (Figure 4). Correspondingly, the RBD, based on the FTA, illustrates a 3-line parallel structure (Figure 5).

Second, the current paper conducted FMEA to consider each risk-priority number (RPN) from the results of the FTA and RBD (Table 2). Initially, for occurrence, the likelihood of bad conditions for driving and pedestrian road crossing are highly correlated. Next, for severity, the degrees for invisibility and intentional crossing were lowered because people can take evasive actions. Lastly, for detection, almost all the internal factors, except for carelessness, were relatively undetectable. Consequently, Table 2 highlights the criticality of the machine errors.

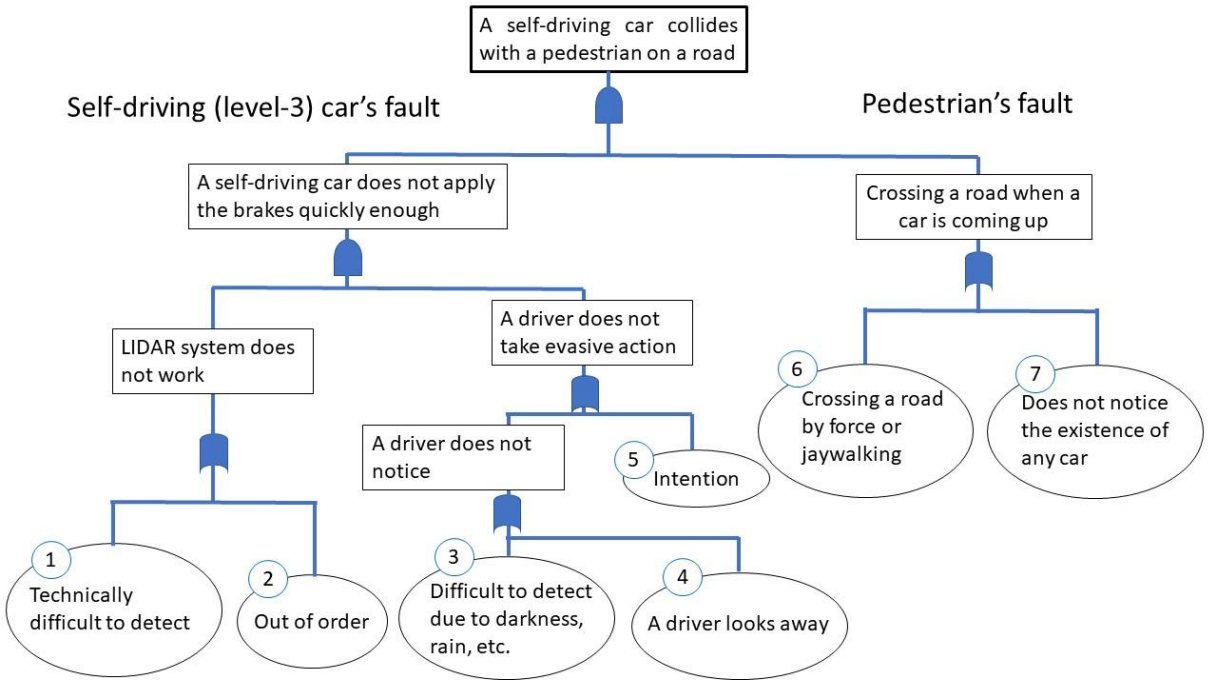

Figure 4. Fault tree analysis of Case 1.

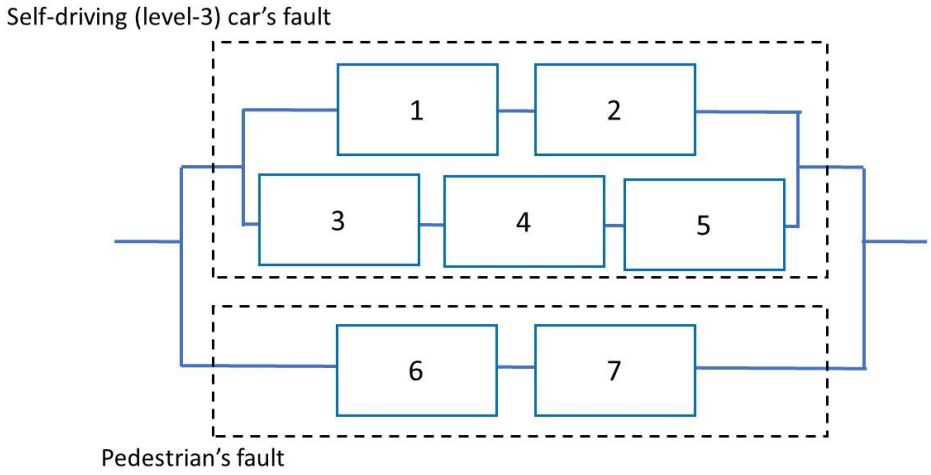

Figure 5. Reliability block diagram for Case 1.

Table 1. Failure mode effect analysis (existing condition)

Potential Failure Mode	Potential Effects of Failure	Potential Failure Causes	Current Controls	O	S	D	RPN
A self-driving car does not apply the brakes quickly enough (self-driving car's fault)	LIDAR system does not work	Technically difficult to detect a pedestrian	Debugging	3	5	5	75
		Out of order	Regular maintenance	3	5	4	60
	A driver does not take evasive action because the driver does not notice the existence of a pedestrian or does not take intentionally	Invisibility (darkness, rain, fog, etc.)	Reducing speed	5	3	1	15
		A driver looks away	Guideline, regulations	4	5	1	20
		Intention	Guideline, regulations	1	5	5	25
(Pedestrian's fault)	A pedestrian crosses a road when a car is coming	A pedestrian crosses a road by force or jaywalks	Education	5	4	1	20
		A pedestrian does not notice the existence of a car coming	Education	5	5	1	25

Note: O = Occurrence, S = Severity, D = Detection and RPN = risk-priority number (all shown as a 5-point scale)

3.3 Discussion: Lessons from the accident and the limitations of failures and advanced technology

First, a noteworthy point when analysing factors is to take into account the current autonomous level. Because a driver is necessary to take evasive action in the case of an emergency, the current technology level of self-driving cars, including the one which caused the accident, is level-3 (Figure 6). This directly means that an AND-gate should be added as a fail-safe to the side of the self-driving car's fault compared with a level-4 or higher autonomous level (basic events 3 and 4, shown in Figure 4) and, accordingly, reliability blocks were added (Figure 5). As a result, the RBD forms a simple 3-line parallel structure; considering each RPN, however, technological difficulties become remarkable (see Table 2). To summarise, currently the status is that a driver of a level-3 self-driving car assumes an important role in avoiding critical dangers.

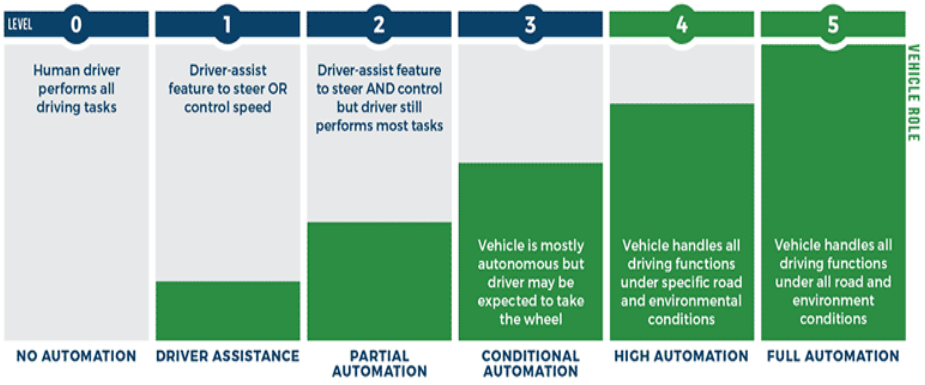

Figure 6. Self-driving car's autonomous levels (Guerra, 2017).

Second, it is obvious that the technical difficulty of detecting people is the greatest problem for the safeguarding of pedestrians (Table 2). To achieve further high autonomous levels, the development of highly robust systems is indispensable because if the autonomous level reaches level-4 or higher, there is no driver command. Even recently, what some US states permitted is to allow technology companies to test self-driving cars on roads with a driver onboard (Wakabayashi, 2018; Nakata, 2018); the safety benchmark have not been clarified (Guerra, 2017; Wakabayashi, 2018). While the NTSB (2018) focused on how the sensing systems worked and indicated the necessity for at least compulsory braking and alert systems, no probable cause for the accident has not been specified. The situation surrounding the detection capability of self-driving cars illuminates the prematurity of the implementation of level-3 or higher autonomous vehicles.

Third, considering this prematurity, it is possible that assumed variables on public roads are too difficult to deal with. This means that the hybrid environment, where people and cars simultaneously exist, makes the challenge too complicated for the state-of-the-art technology. In addition, there is a wide range of variables, including the shapes of roads, road signs, the weather, brightness and other objects. Therefore, it appears necessary for policy makers to reduce such variables as much as possible, not to suddenly allow tests on public roads. However, reducing failures by changing the prerequisites seems beyond the analysis of FTA, which seems a useful analysis method for analysing risk factors and considering palliative countermeasures.

Finally, in this paper recommendations were derived from a hybrid approach in accident causation analysis. In this paper we used FTA an analysis to study causal factors. The FTA is a logical 'language' that consists of two main logical symbols (a language of just two letters). Such constraint can act as both a point pf weakness and a blessing; the former is due to its simplicity, whereas the latter is due to it forcing one to think logically, and this acts as a 1st loop learning. The higher levels, or richness, in learning, occurs when this feeds into RBD (for venerability analysis) FMEA (for occurrence, severity and detection analysis). This subsequently leads to recommendations related to prevention, mitigation and enacting new early warning measures. In order to fully satisfy triple loop learning, we propose future work on benchmarking analysis that is coupled with high reliability organisations (HRO) principles in order to provide a framework that fully satisfies all three loops of learning.

References

1. Aven, T.: A unified framework for risk and vulnerability analysis covering both safety and security. *Reliability engineering & System safety*, *92*(6), 745-754 (2007).
2. Argyris, C. (1999). On organizational learning 2nd Edition. *Malden, MA: Blackwell*.
3. Argyris, C., & Schön, D.:What is an organization that it may learn. *Organizational learning: A theory perspective, Argyris C, Schoen DA (eds). Addison-Wesley: Reading, MA*, 8-29 (1978).
4. Baldwin, J., & Lin, Z.: Impediments to advanced technology adoption for Canadian manufacturers. *Research Policy*, *31*(1), 1–18 (2002).
5. Cabinet Secretariat. (2012, July 31). *日本再生戦略 [Japan revitalization strategy]*. Retrieved from https://www.cas.go.jp/jp/seisaku/npu/pdf/20120731/20120731.pdf
6. Chesbrough, H.: Business model innovation: opportunities and barriers. Long Range Plan. 43, 354–363 (2010).

7. Cope, J.: Entrepreneurial learning and critical reflection: Discontinuous events as triggers for 'higher-level'learning. *Management learning*, *34*(4), 429-450 (2003).
8. D'Este, P., Iammarino, S., Savona, M., & von Tunzelmann, N.: What hampers innovation? Revealed barriers versus deterring barriers. *Research policy*, *41*(2), 482-488 (2012).
9. Department for Transport. (2015, February). *The Pathway to Driverless Cars: Summary report and action plan.* Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/401562/pathway-driverless-cars-summary.pdf
10. Drupsteen, L., & Guldenmund, F. W.: What is learning? A review of the safety literature to define learning from incidents, accidents and disasters. *Journal of Contingencies and Crisis Management*, *22*(2), 81-96 (2014).
11. Galia, F., & Legros, D.: Complementarities between obstacles to innovation: evidence from France. *Research policy*, *33*(8), 1185-1199 (2004).
12. Goode, N., Salmon, P. M., Lenne, M., & Finch, C.: *Translating systems thinking into practice: a guide to developing incident reporting systems*. CRC Press (2018).
13. Griggs, T., & Wakabayashi, D.: *How a self-driving Uber killed a pedestrian in Arizona.* Retrieved from The New York Times (2018, March 21).
14. Hatamura Y.: Evacuation and decontamination in response to the Fukushima nuclear power plant accident. The 2011 Fukushima Nuclear Power Plant Accident. Boston: Woodhead Publishing (2015).
15. Hollnagel, E.: *Barriers and accident prevention*. Routledge. (2016).
16. Hulme, A., Stanton, N. A., Walker, G. H., Waterson, P., & Salmon, P. M.: What do applications of systems thinking accident analysis methods tell us about accident causation? A systematic review of applications between 1990 and 2018. *Safety science*, *117*, 164-183 (2019).
17. Leoncini, R.: Learning-by-failing. An empirical exercise on CIS data. *Research Policy*, *45*(2), 376-386 (2016).
18. Leveson, N.: A new accident model for engineering safer systems. *Safety science*, *42*(4), 237-270 (2004).
19. Lhuillery, S., & Pfister, E.: R&D cooperation and failures in innovation projects: Empirical evidence from French CIS data. *Research policy*, *38*(1), 45-57 (2009).
20. Lim, H. S. M., & Taeihagh, A.: Autonomous vehicles for smart and sustainable cities: An in-depth exploration of privacy and cybersecurity implications. *Energies*, *11*(5), 1062 (2018).
21. Nakata, A. (2018, November 7). *2018年に急増した自動運転車の事故一覧、米カリフォルニア州で50件超.* Retrieved from NIKKEI x TECH: https://tech.nikkeibp.co.jp/atcl/nxt/column/18/00155/110500034
22. Rasmussen, J.: Risk management in a dynamic society: a modelling problem. *Safety science*, *27*(2-3), 183-213 (1997).
23. Waterson, P., Jenkins, D. P., Salmon, P. M., & Underwood, P: 'Remixing Rasmussen': The evolution of Accimaps within systemic accident analysis. *Applied ergonomics*, *59*, 483-503 (2017).
24. Tosey, P., Visser, M., & Saunders, M. N.: The origins and conceptualizations of 'triple-loop'learning: A critical review. *Management learning*, *43*(3), 291-307 (2012).
25. Wakabayashi, D.: *Self-Driving Uber Car Kills Pedestrian in Arizona, Where Robots Roam.* Retrieved from The New York Times (2018, March 19).