A Safety-Case Approach to Ethical Considerations for Autonomous Vehicles

C. Menon*, R. Alexander †

*University of Hertfordshire, UK, catherine.menon@city.ac.uk
† University of York, UK, rob.alexander@cs.york.ac.uk

Keywords: safety, autonomous systems, ethics, argument

Abstract

Ethical considerations for autonomous vehicles (AVs) go beyond the "trolley problem" to include such aspects as risk / benefit trade-offs, informed consent, risk responsibility and risk mitigation within a system of systems. In this paper we present a methodology for arguing that the behaviour of a given AV meets desired ethical characteristics. We identify some of the ethical imperatives surrounding the introduction of AVs and consider how decisions made during development can impact the ethics of the AV's behaviour.

1 Introduction

Autonomous systems (a category which includes AVs) have been proposed for use in multiple domains, with examples including nuclear containment, defence systems, health and transport. The ethical requirements across each of these domains will inevitably differ, and in many cases there is no consensus as to which system behaviours would be deemed ethically appropriate.

Ethics is not restricted only to safety, and the discussion of ethical introduction and behaviour of AVs may include considerations of environmental impact, economics, manufacturing processes and adequate financial investment [1]. However, in this paper we will focus on the safety and ethical aspects of the proposed use of AVs for transport.

We present a method for arguing that the behaviour of an AV meets specified ethical characteristics, and that these align with safety. Section 2 provides some ethical background, while Section 3 introduces the safety and ethical landscape around AV introduction. In Section 4 we introduce the concept of risk trade-offs, and in Section 5 discuss safety, ethics and the development of systems. Section 6 presents a methodology for constructing ethical arguments, aligned with safety case arguments and drawing on risk profiles, and Section 7 contains conclusions.

2 Ethical background

The "trolley problem" refers to a well-known ethical thought experiment, in which a train / trolley is on a set of tracks which will cause it to collide with a number of people. The observer is asked whether s/he would choose to switch the train to a second set of tracks which will cause it to collide with a single person only. Amendments and extensions to the trolley problem have couched the problem in terms of an active vs passive choice as well as experimented with the relative "worth" of each person affected.

The trolley problem has a clear analogue in the case of AV behaviour, in that a situation may be encountered in which a collision with at least one group of people is inevitable. In this case, the developers responsible for the behaviour of the AV must address a trolley problem: which group(s) should the AV choose to impact. This is explored further in [2].

2.1 Systems of ethics

The trolley problem can be used to illustrate a number of different ethical systems, providing examples of how these might differ in their application to AV behaviour.

Consequentialism [3] is often considered to provide a reasonable foundation for discussion of AV ethics and behaviour. Consequentialism is an ethical theory which prioritises the outcomes: consequentialist ethics deems acts to be morally acceptable if they lead to a good outcome. This is sometimes summarised as "the end justifies the means". A consequentialist approach to AV safety would be to seek to reduce overall harm by minimising the number of people harmed; a consequentialist solution to the trolley problem would be to switch the trolley onto the section of the track with a single person. Consequentialism as an ethical theory is aligned with more general safety criteria [4] in terms of minimising harm, but does not take into account questions of risk responsibility, informed consent for acceptance of risk and calculations relating to acceptable exposure due to work.

By contrast, deontological theories of ethics prioritise acting in accordance with explicitly stated duties and rules [5]. Deontology therefore does not require the AV to consider the outcomes, but merely to act in accordance with pre-programmed rules (which may include, for example, a rule that the AV must not injure – or cause to be injured – any person). While encoding such rules is conceptually simpler than requiring the AV to perform calculations minimising harm, deontological ethics does require the identification of rules for every situation the AV may find itself in. A deontological approach to the trolley problem would be to consider whether rules exist which govern the acceptability of switching the trolley to a different track, regardless of the risk exposure to any individuals.

A third ethical imperative relevant to AVs is the concept of virtue ethics, typically presented in terms of self-sacrifice [6]. This discusses the extent to which an AV should choose to sacrifice itself and its passenger when placed in a situation in which this would reduce harm to a third party.

2.2 Additional ethical dilemmas

More generally, from a safety perspective we are concerned about the risk posed by the AV to different groups, and the ethical justification for prioritising the safety of one group over another. This extends the trolley problem to other situations in which the risk is the deciding factor. In the following examples where we refer to the decisions or choices made by the AV, this is to be understood to be the decisions and choices made by the AV system developers which result in the defined behaviour.

In [6] a case is presented whereby an AV may choose to position itself within a lane closer to a smaller car than to a truck. This decision might be justified in two ways: firstly, that this behaviour is typical of a human driver, and secondly that this reduces the risk to the AV (a collision with a small car may reduce harm to the occupants of the AV). From a safety perspective, this decision has prioritised the safety of the AV occupants – and the truck occupants – over that of the smaller car. Such a decision would need to be justified within the safety case and from an ethical perspective.

Another situation arises whereby an AV may take the opposite course; choosing to drive closer to (or in the worst case, impact) a heavier vehicle, or a vehicle with safety systems which are known to be better [6]. In this case the severity of an accident may be reduced, compared to an impact with a vehicle with poor safety systems. However, implementing such a decision into the behaviour of the AV represents a deliberate choice to increase the risk to drivers of certain vehicles known for their safety features. Again, this decision would need to be justified both ethically and in the safety case.

Other situations discussed in the existing literatures include the decision of an AV to sacrifice itself (place itself in the path of another vehicle to save a third party from impact) [6], as well as choosing to impact a motorcyclist wearing a helmet over one not wearing such protective devices [7].

3 Safety and ethical landscape

As we discussed in Section 1, the ethical landscape surrounding the introduction of AVs is not limited only to the trolley problem and to AV behaviour during collisions. While we do not go into detail on the ethical issues which are not directly relevant to safety (e.g. environmental impact, job loss, capability benefits, inequality of access to technology etc.), there are a number of issues which do impact indirectly on the safety considerations for AVs.

The first of these is the question of commercial forces driving early adoption of AVs. There is significant public interest in AVs, particularly around self-driving cars, and engineering companies are alert to the advantage of bringing out the "first of kind" of an AV. However, unlike the military and nuclear domains, the high-profile nature of commercial AVs can encourage the categorisation of safety as a competitive advantage. This means that best practice can be difficult to establish, and known problems may not be shared for reasons of commercial interest.

In addition, there are currently no applicable standards which fully address the safety of AVs, including safety of the intended function [8]. Consequently, while there is a clear economic and reputational imperative for a company to bring out the "first of kind" in autonomous vehicles, it is much less clear that such an AV could be demonstrated to be acceptably safe. There is a risk that the push to produce and market AVs can encourage "quick and dirty" practices during the development lifecycle which can have an effect on the system as released to the public. While standards do exist around ethical design of systems [9], these are relatively new and their general applicability has not been fully determined.

Another question which arises is that of risk transfer and system safety, as previously introduced in discussions of the trolley problem. We expand on this in Section 4, but in brief, a simplistic argument that AVs reduce the overall harm does not go far enough. It may be the case that a segment of the population bears an unfair degree of the risk and therefore, although the overall risk is lower, this segment faces either an absolute or a relative increase in the proportion which they bear. The question of consent is also relevant here, in that other road users may be unwittingly bearing a portion of risk to which they have not consented. This concern also applies to the passenger of an AV; if passengers are unaware of the principles governing AV behaviour, they are not able to consent to the consequent risks.

When we move from human drivers to automated ones, we move the intelligence in the decision from conditions of extreme time stress to a much calmer, slower-paced environment. This may raise the standard of ethical performance the public expects. In the case of a human driver, any decisions made in a collision situation are judged according to that environment (e.g. there is little time to choose between different options, the drivers are under stress, and – except where their actions have been negligent – are generally not considered culpable should they make the "wrong" decision [6]). However, an engineer developing the AV is not under the same pressure, and may therefore be expected to ensure that the AV reacts in a morally acceptable way, regardless of how a human driver might.

More generally, equating the actions of an AV with the actions of a human driver may appear defensible from a risk acceptance perspective, but it is not clear that the general public will necessarily be willing to accept the same risk when it is posed by a machine as opposed to a person.

A more general concern is that of the impact of AVs on the wider road network. This network can be viewed as a system of systems (SoS), with the AVs comprising one component only. The risk posed by an AV may therefore affect any

portion of this network, leading to unforeseen interactions and emergent behaviour. One example of this may be an increase in traffic jams due to all AVs following the same route, as it is in the interest of no individual AV to change route. Another example may be the effect on driver norms where, for example, human drivers may customarily let other vehicles exit from a side street and the road planning is such that it presumes this type of essentially human interaction. These situations will be exacerbated in the case of AVs which make use of machine learning algorithms, where local optimisations made by these algorithms can negatively affect traffic flow, safety or efficiency of the wider network.

4 Risk Trades and Risk Profiles

The ethical dilemmas introduced thus far focus on the situations where the AV behaviour prioritises the safety of one group over another. That is, in these situations a choice has been made to reduce one risk posed by the system (e.g. the risk posed to pedestrians) at the potential cost of increasing another risk (e.g. that posed to other vehicles).

In general, there may be multiple ways to reduce the overall risk posed by the system to As Low As Reasonably Practicable (ALARP). Individual risks can be traded-off, or balanced against each other as described above, where an increase in one risk is accepted in return for a decrease in another. Many safety guidance documents [4] provide little information on how to make these choices, requiring only that the overall system risk should be ALARP. It should be noted, however, that where the concept is discussed in standards [10] [11], these emphasise the need to balance individual risks within a system and consider established good practice.

Risk trade-offs and balances can happen at three levels throughout system development. At the micro level a developer might make development choices which reduce certain risks at the cost of potentially increasing others. For example, a choice of C over SPARK ADA may provide increased access to experienced developers, but at the cost of static analysability. At the macro level, as already discussed, one risk posed by the system may be mitigated at the cost of potentially increasing another. Section 3.1 presents this in more detail. Finally, in some situations accepting an increase in risk in one domain or system may lead to a benefit in another. This is discussed further in Section 4.5.

4.1 Risk Profiles

In [12] we presented a number of different risk reduction approaches, or risk profiles, which provide alternative ways of balancing individual risks in order to achieve an ALARP system risk. An ontology of these is briefly given below, and it should be noted that these risk profiles can be combined in a number of ways to produce a "custom" profile.

4.2 Fairness in improvement

The aim of this approach is to achieve a similar absolute risk reduction for all individual risks. A fairness in improvement approach prioritises the reduction of *all* risks A, B… N regardless of the relative cost of these reductions (provided these are reasonably practicable), and regardless of whether making these reductions to one risk A means that for technical reasons further reductions cannot then be made to another risk B. Using a fairness in improvement approach can mean that no individual risk is as low as technically possible considered in isolation. However, this approach ensures that the risk reduction effort confers a certain minimum benefit on all system risks.

A fairness in improvement approach for AV risk reduction may correspond to attempting to mimic the actions and risk reduction behaviour exhibited by a human driver. The risks posed by an AV will therefore bear a similar relationship to each other (e.g. some higher, some lower) as the risks posed by a human driver. It should be noted that an AV developer is still required to minimise the system risk ALARP, so it may be the case that the AV presents a lower overall system risk than the human driver.

4.3 Fairness in outcome

The aim of this approach is to achieve a similar level of risk for all individual risks. Fairness in outcome means that our risk reduction attempts prioritise the reduction of a more severe risk A over the reduction of a less severe risk B. This is the case regardless of the relative cost of reducing risks A and B compared to each other, and regardless of whether making these reductions to A means that for technical reasons further reductions cannot be made to B. Using a fairness in outcome approach can mean that the risk reduction efforts are concentrated on only a few risks, with no benefit for the other risks. However, this approach ensures that the areas of greatest risk are targeted by reduction efforts.

A fairness in outcome approach for AV risk reduction may correspond to a focus on reducing the greatest risks posed by the AV (e.g. reducing the risks posed to motorcyclists without helmets, given the correspondingly greater severity of any collision). In this case a solution to some manifestations of trolley problem is presented by the choice of this risk profile: impact with other vehicles is likely, for example, to be a preferred hazard over impact with pedestrians. However, it should of course be noted that this does not negate the requirement for AV system developers to balance these individual risks such that an increase in one risk is only permitted given an equivalent or greater decrease in another.

4.4 Long-term risk benefit

The question of system risks that change over time can also be relevant when balancing individual risks. Standards such as [10] also consider the possibility of accepting a higher short-term risk if this results in a long-term risk reduction.

For AV risk reduction, taking a long-term risk benefit approach prioritises the introduction of AVs, along with any concomitant short-term increase in risk, should it be possible to demonstrate that this would lead to fewer lives being lost

over the long-term. Long-term risk benefit requires explicit justification within the safety case, as it may not be possible to demonstrate that in the short term the system risk is ALARP. Consequently, long-term risk benefit should be used only to customise and refine other risk profiles.

4.5 External risk transfer

Risk transfer refers to the situation where there are multiple components or interacting subsystems, such as in the presence of a SoS. In this case, an ALARP claim for each subsystem considered in isolation does not necessarily lead to the lowest overall system risk. In these situations an increase in a local risk associated with one system may be accepted in return for a decrease in the risk associated with the wider system. This is presented in further detail in [12].

More generally, in some cases an increase in a safety risk may result in a benefit in an external domain. For example, the presence of certain security features such as Intrusion Detection Systems (IDS) provides a security advantage while making it harder to demonstrate the safety of the system (amongst other concerns, IDS need to be regularly updated, which is difficult given the rigorous testing and validation required by safety-critical systems [13]). It should be noted, however, that this external risk transfer cannot be deemed acceptable from an ALARP perspective, as the ALARP principle does not consider benefits outside the safety domain.

4.6 Risk profiles and ethical behaviour

A risk profile represents a means of balancing risks against each other, and can be used to describe a set of ethical drivers or priorities. This can be seen most easily by applications of the trolley problem: a risk profile prescribes a balance of risks which prioritises some over others. This corresponds to prioritising the safety of those groups who are impacted by the risks deemed by the risk profile to be higher priority. We can therefore use risk profiles to describe ethically desired AV behaviour by framing it in terms of risk reduction.

5 Safety, ethics and development

Risk profiles allow us to bring safety and ethics together for AV behaviour by explicitly presenting the risk balancing and trade-offs inherent in any implemented solution to the trolley problem. The safety case must then justify these trade-offs and balances.

Although ethical questions are not limited to safety (see Section 3), those that do concern safety deal with the most severe harms. A safety case which does not consider the underlying ethics of decisions around risk and harm can be considered deficient. In order for all stakeholders to adequately understand the implications of the decisions made around risk management, the ethical foundation for these needs to be made explicit within a supplementary "ethics case". In this section we propose the use of such ethics cases and demonstrate how they can be used in conjunction with a safety case to adequately support arguments around the behaviour of AVs.

5.1 Engineering and implemented ethics

When referring to the development and operation of AVs there are two interrelated but distinct applications of ethics and ethical systems. The first of these we will term *engineering ethics* and the second *implemented ethics*.

Engineering ethics refers to the ethical principles adhered to by engineers during system and software development. These may be in the form of principles or codes of conduct formalised by a professional organisation [14]. They typically include criteria such as honesty, integrity, respect for law and the public interest, accuracy, rigour, fairness, objectivity and leadership. In addition, they encourage further thought and assessment to determine if any given engineering action is ethically defensible. It is important to note that adherence to a code of engineering ethics does not, in itself, mean that the behaviour of any resultant system will necessarily be considered ethical by all stakeholders (this can be seen particularly in the defence domain). However, adherence to a code of engineering ethics helps to support arguments about the behaviour and properties of the system by providing confidence in the integrity of any lifecycle artefacts. Should developers not adhere to any professional code of ethics, any argument about the safety of the system or its behaviour can only be weakly supported.

Implemented ethics, by contrast, refer to the ethics which govern the behaviour of the AV itself in the field. These can include the extent to which the safety of the driver is balanced against that of third parties, and more generally the choices the AV makes when confronted with various forms of the trolley problem. Other aspects of behaviour governed by implemented ethics include the extent to which the AV shares data, the dynamic measures performed during driving to reduce environmental impact and the extent to which social aspects of courteous driving are implemented. Unlike engineering ethics, there may not be consensus on the "right" implemented ethics which will govern the behaviour of the AV. Acceptable ethical behaviour will vary across different societies (including different countries) as well as different domains of use. Section 2 discusses this in more detail.

6 Ethics case and argumentation

Just as a claim relating to the safety of the system is supported by a compelling argument, we propose that a claim relating to the ethics of the AV should be supported likewise.

As with safety arguments, there is no single "one-size-fits-all" method of creating an argument to support claims relating to the ethics of an AV. However, any adequate argument would need to present a number of foundational principles that demonstrate the ethics of the AV is adequate, and argue that these have been shown to be met. In this section, we present a methodology for doing this which is in line with the

principles discussed in [9] as well as relevant safety and legal criteria [4].

The argument we present consists of three independent and interacting legs, each supporting a different claim. The overall claim is:

G0: The behaviour of the AV is ethically appropriate for its proposed context of use.

This claim is supported by three sub-claims:

A0: Engineering ethics are adequately defined, implemented and adhered to during the development lifecycle.

B0: Implemented ethics are adequately specified and comply with the legal, social and ethical norms of the environment of use.

C0: The risk management and design decisions are such that the AV behaviour adheres sufficiently closely to these implemented ethics.

We address each of these claims in further detail in the following sections.

6.1 Claim A0

A0: Engineering ethics are adequately defined, implemented and adhered to during the development lifecycle.

The purpose of this claim is to demonstrate that the engineering codes of practice and prescribed ethical principles are not compromised or impacted by any decisions relating to the ethical behaviour which it is decided the AV should demonstrate.

The desired engineering ethics may be identified by referencing codes of conduct ([14], [15]), domain good practice and relevant previous decisions and their adequacy should be justified. Evidence to support this claim may be in the form of Continuing Professional Development records, audit records, lifecycle artefacts, documented processes and policies and so forth.

6.2 Claim B0

B0*: Implemented ethics are adequately specified, and comply with the legal, social and ethical norms of the environment of use.*

We recommend that this claim is broken down into sub-claims for clarity of argument. A template example is given below.

B1: The implemented ethics are adequately specified.

This specification may be in the form of references out to legal documents, to standards and policies, to previous system design decisions, records of public consultations and so forth. The specification of implemented ethics must be sufficient to address all issues raised in Section 5.1, as well as to provide a justification that the issues under discussion are sufficient and complete.

B2: The implemented ethics comply with the legal, social and ethical norms of the environment of use.

As stated in [9], the norms of the relevant community (or environment of use) must be considered when assessing the behaviour of the AV. The implemented ethics must be compatible with these norms. It should be noted that this does not mean that an AV should behave in exactly the same way as a human driver (that is, the implemented ethics do not have to be identical to the ethics currently embedded within the environment of use), but the two must be compatible, and any discrepancies identified and a justification provided.

6.3 Claim C0

C0: *The risk management and design decisions are such that the AV behaviour adheres sufficiently closely to these implemented ethics.*

We recommend that this claim is broken down into a number of sub-claims for clarity of argument. A template example is given below.

C1: System design and intended AV behaviour are adequately specified.

This sub-claim should be supported with evidence relating to the system design and implementation. Its intent is to demonstrate that the AV system design is specified sufficiently well enough to reduce the likelihood of unexpected behaviours. Should the intended behaviour or the design of the AV be underspecified, then it becomes much harder to predict whether the resultant operational actions of the AV will be considered ethically acceptable.

C2: Design decisions and risk reduction decisions reflect the specified implemented ethics

This claim should firstly be supported by nomination and definition of a specified risk profile (customised if required, as described in Section 4). It must also be demonstrated that this risk profile reflects the desired implemented ethics. The nomination of a risk profile, with the consequent requirement that this describe a mechanism for reducing the system risk ALARP, is necessary in order to ensure that the specified implemented ethics do not contradict any of the legal requirements around safety [4].

For example, should the implemented ethics require that the AV behaviour mimic the behaviour of a human driver (thereby resulting in no change in relative risk distribution across the road network from the introduction of AVs), then we would expect to see a “fairness in improvement” risk profile selected. In practice, the desired implemented ethics are likely to be sufficiently complex such that a significant amount of customisation is needed to any of the “base” risk profiles.

Secondly, this claim should be supported with evidence that the risk management and risk reduction decisions reflect the selected risk profile. In practice, this may best be done by referring out to individual claims in the safety argument and demonstrating how the risk prioritisation decisions have been reflected in the mitigations.

C3: Any gaps between the behaviour resulting from the design and risk reduction decisions and the implemented ethics are adequately justified.

The final sub-claim addresses the fact that, like safety, ethics is a limit concept [16]. Just as a system cannot be guaranteed to be absolutely safe, it cannot be guaranteed to be absolutely ethical (this is exacerbated by the difficulty in adequately specifying a comprehensive set of ethical principles).

This sub-claim should therefore be supported by a gap analysis of how well the AV system design and the risk reduction decisions reflect the implemented ethics. In practice, restricting this gap analysis to risk reduction decisions will not be sufficient, and the overall AV design and behaviour should be considered also. This is because not all implemented ethics refer to safety (some may refer to aspects of environmental sustainability, others to elements of courteous driving etc.). The risk profiles, dealing only with safety, will not be able to be used to argue that the "non-safety" requirements of the implemented ethics are met. Where the behaviour or design is underspecified, this should be considered as a gap.

For any identified gaps, the argument must demonstrate that mitigations have been put in place to reduce the effect of these gaps so far as is reasonably practicable. This parallels the ALARP requirement for safety, and similar argument techniques may be used.

7 Conclusions

In this paper we have identified the ethical landscape and imperatives that govern discussion of AV behaviour. We have introduced and formalised the concept of risk trade-offs, which are typically dealt with superficially by applicable safety-critical guidance. We have considered the ethical drivers behind these risk trade-offs, and identified the need for transparency in risk balances and risk trade-offs.

We have presented a methodology for arguing that the behaviour of an AV meets ethical criteria deemed relevant to safety. This methodology draws on aspects of safety argumentation to support a number of claims relating to the definition of ethically acceptable behaviour, the applicability of this in the proposed environment and the design decisions made during AV development. We draw on the concept of risk profiles to transform ethical principles into the language of safety and to provide a foundation for discussing how our ethical principles impact our risk mitigation decisions.

We distinguish between the principles of ethical conduct constraining the professional actions of engineers, and the principles of ethics constraining the behaviour of the systems these engineers design. We recognise that ethics of system behaviour, like safety, is a limit concept and extend the consideration of ALARP into the ethical domain. This allows us to examine whether the behaviour demonstrated by the AV is sufficiently close to the ethically desired behaviour in the environment of use.

There is the potential for significant further work in this area, particularly in the areas of balancing risk trade-offs. It would be of value to further extend the ontology of risk profiles to consider which refinements are of most use across multiple domains. In addition to this, the consideration of ethical drivers outside safety is also a relevant topic. Security and privacy are topical concerns for AVs, while human trust and social integration are issues of note for autonomous systems in general. There is scope for considering the extent to which safety, security, ethics and trust interact, and how the requirements of these can be balanced for a general autonomous system.

Acknowledgements

The authors wish to thank the SCSC Safety of Autonomous Systems Working Group.

References

[1] Harris, C., Pritchard, M., Rabins, M. "Engineering Ethics: Concepts and Cases", *Wadsworth, Cengage Learning*, 2009.

[2] MIT, "MIT Moral Machine", http://moralmachine.mit.edu/, 2017.

[3] Goodall, N. "Machine Ethics and Automated Vehicles", *Road Vehicle Automation*, pp 93 – 102, 2014.

[4] Health and Safety Executive, "Reducing Risks, Protecting People", 2001.

[5] Goodall, N. "Ethical Decision Making During Automated Vehicle Crashes", *Transportation Research Record Journal of the Transportation Research Board*, 2014.

[6] Lin, P. "Why Ethics Matter for Autonomous Cars", *Autonomous Driving*, pp 69 – 85, 2015.

[7] Gerdes, J., Thorntson, S. "Implementable Ethics for Autonomous Vehicles", *Autonomous Driving*, pp 87 – 102, 2016.

[8] International Standard for Organisation, "Road Vehicles – Functional Safety", *ISO 26262*, 2011.

[9] IEEE Global Initiative. "Ethically Aligned Design", *IEEE Standards v1.0*, 2016.

[10] Health and Safety Executive, "Safety Assessment Principles for Nuclear Facilities", 2006.

[11] Office for Nuclear Regulation, "Guidance on the Demonstration of ALARP", 2013.

[12] Menon, C., Bloomfield, R., Clements, T. "Interpreting ALARP", *Proceedings of the 8th IET International System Safety Conference,* 2013.

[13] Johnson, C. "Barriers to the Use of Intrusion Detection Systems in Safety-Critical Applications", *SAFECOMP*, 2014.

[14] Royal Academy of Engineering, "Statement of Ethical Principles", 2017.

[15] IET, "Rules of Conduct 2015", 2015.

[16] Habli, I., Kelly, T., Macnish, K., Megone, C., Nicholson, M., Rae, A. "The Ethics of Acceptable Safety", *Proceedings of the 23rd Safety-critical Systems Symposium,* 2015.