MARCH2017 International Workshop on decision Making in Software ARCHitecture
Collocated at ICSA 2017 Gothenburg, Sweden

# Risk-based Decision-making Fallacies: Why Present Functional Safety Standards Are Not Enough

Andreas Johnsen[1], Gordana Dodig-Crnkovic[12], Kristina Lundqvist[1]
Kaj Hänninen[1], Paul Pettersson[1],

[1]School of Innovation, Design and Engineering, Mälardalen University

[2]Chalmers University of Technology and University of Gothenburg

# FUNCTIONAL SAFETY OF A SYSTEM

Functional safety of a system is the part of its overall safety, understood as freedom from unacceptable/unreasonable risks, that depends on a system operating correctly in response to its inputs.

Functional safety elements are examined at every stage of the the software development life cycle, including requirement specification, design, implementation, verification, validation and deployment.

Acceptability of risks is judged within a framework of analysis with contextual and cultural aspects by individuals who may introduce subjectivity and misconceptions in the assessment.

# FUNCTIONAL SAFETY STANDARDS

If functional safety standards control functional safety by the absence of risk judged to be unacceptable, we argue it is critical to also require the absence of unreasonable judgments.

We study common fallacies in risk perception, through a moral-psychological analysis of functional safety standards.

We propose plausible improvements of the involved risk-related decision making processes, with analysis of *the notion of* an acceptable residual risk.

# HARDWARE-SOFTWARE SYSTEMS

Issues of safety are best addressed in a combined hardware-software system, where malfunction of software can lead to hardware causing damage.

As a reference model, we use the functional safety standard ISO 26262, addressing potential hazards caused by malfunctions of hardware and software systems within road vehicles, and defines safety measures that are required to achieve an acceptable level of safety.

# FALLACIES IN RISK PERCEPTION

Fallacies of risk perception: The Majority Rule/Groupthink, Fallacies of Individual Judgment (e.g. the above-average effect), Biases due to Memory Mechanisms, “What You See Is All There Is”, Status Quo Bias, Biases by Subconscious Processes (e.g. priming), etc.

Adequate risk assessment especially important in contemporary developed autonomous vehicles with increasing the role of computational applications.

# ISO 26262 SAFETY
## - the absence of unreasonable risk

ISO 26262 is an automotive-specific interpretation of the basic functional safety standard IEC 61508. ISO 26262 provides a safety lifecycle reference model that complies with standardized safety requirements in the development of E/E systems within road vehicles. The reference model both addresses potential hazards caused by malfunctions and specifies safety measures through which safety is achieved.

The standard defines safety as *the absence of unreasonable risk*:

*"risk judged to be unacceptable in a certain context according to valid societal moral concepts."*

# FUNCTIONAL SAFETY OF A SYSTEM

We argue that *functional safety standards should be complemented with the analysis of potential hazards caused by fallacies in risk perception*, their countermeasures, and the requirement that residual risks must be explicated, motivated, and accompanied by a plan for their continuous reduction.

This approach becomes especially important in contemporary developed autonomous vehicles with increasing computational applications.

# RISK

Risk is defined as:

*“the combination of the probability of occurrence of harm and the severity of that harm”*

Harm and its severity depend on the value system and ethics, which remain implicit in the standard.

# HARM

The notion of *harm* is related to emotions.

Harm, physical as well as mental, instinctively causes unpleasant emotions. Together with the ability of reasoning, humans are able to develop models of right and wrong conduct – ethics.

The “do no harm”-principle is fundamental to ethics, derived from the value of human dignity and the respect of the personal integrity. In applied ethics, the fundamental *principle of beneficence* refers to a moral obligation to “act for the others’ benefit, helping them to further their important and legitimate interests, often by preventing or removing possible harms.”

# HARM FROM “RESIDUAL RISK”

The question is whether the reasoning behind an “acceptable residual risk” includes untrue assumptions, inferences, or theorems.

Such errors have the potential to result in causing harm, that is in unethical conduct, in spite of careful application of functional safety standards.

We identify relevant systematic errors of thinking and analyze the impact these fallacies may have on functional safety.

The study of fallacies is based on Kahneman’s book “Thinking, Fast and Slow”.

# DECISIONS BASED ON VALUES

Values serve as a guide to decision making and action. They are relevant to all aspects of engineering practice.

Contrary to what many believe, cognitive scientists have found values to be integral parts of STEM (Science, Technology, Engineering, and Mathematics).

TUANA. COMMUNICATIONS OF THE ACM | DECEMBER 2015 | VOL. 58 | NO. 12

# TYPES OF VALUES

Various types of values can be involved in decision making and reasoning:

- *economic* values, etc.
- functional & extra-functional values (dependability, reliability, robustness, safety, security, accuracy, integrity, availability, responsiveness, throughput, etc.)
- ethical values (the good of society, equity, sustainability)
- *aesthetic* values (simplicity, elegance, complexity), or
- *epistemic* values (predictive power, reliability, coherence, scope).

# CONCLUSIONS

Functional safety is assessed in different phases - from initiation and requirements specification to development, testing and verification and maintenance of a system.

Decision-making includes subjective elements in which values and ethical judgments are part of the decision process.

We propose that residual risk assessment related decision-making is grounded in  value-based ethical aspects, which today are implicit, and should be made visible and a subject of scrutiny.

Analysis points out the importance of a robust safety culture with developed countermeasures to the common fallacies in risk perception, which are not addressed by contemporary functional safety standards.

# FUTURE WORK

Study of the process of software architecting of cyber physical systems when functional safety is addressed, including stakeholders.

Which values enter the decision making processes in different phases of research and development and in what way

What would be the best way to assure transparency and document decision-making with respect to risk assessment and specifically residual risk

The role of safety culture – from the standard to technology design and application. BUILDING IN LEARNING IN THE SYSTEM.

# REFERENCES

Avizˇienis, J.-C. Laprie, and B. Randell, Dependability and Its Threats: A Taxonomy. Boston, MA: Springer US, 2004, pp. 91–120.

International Organization for Standardization, "ISO 26262-1:2011 Road vehicles - Functional safety," Geneva, Switzerland.

R. Hugman, E. Pittaway, and L. Bartolomei, "When 'Do No Harm' Is Not Enough: The Ethics of Research with Refugees and Other Vulnerable Groups," British Journal of Social Work, 2011.

T. Beauchamp, "The Principle of Beneficence in Applied Ethics," in Stanford Encyclopedia of Philosophy, 2008.

D. Kahneman, Thinking, Fast and Slow. New York: Farrar, Straus and Giroux, 2011.

J. M. Doris, The Moral Psychology Handbook. Oxford University Press, 2010.

H. Kienle, D. Sundmark, K. Lundqvist, and A. Johnsen, "Liability for Software in Safety-Critical Mechatronic Systems: An Industrial Questionnaire," in Proceedings of the 2nd International Workshop on Software Engineering for Embedded Systems, June 2012.

# REFERENCES

P. Kemp, En teknolgietik. Stockholm, Sweden: Brutus O¨ stlings Bokfo¨rlag Symposion, 1991.

T. Krause, “The Ethics of Safety,” http://ehstoday.com/safety/best-practices/ehs_imp_67392, October 2016.

G. Dodig Crnkovic and B. Cürüklü, “Robots: ethical by design,” Ethics and Information Technology, vol. 14, no. 1, pp. 61–71, 2012.

Sapienza, G., Dodig-Crnkovic, G. and Crnkovic, I. Inclusion of Ethical Aspects in Multi-Criteria Decision Analysis. Proc. WICSA and CompArch conference. Decision Making in Software ARCHitecture (MARCH), 2016 1st International Workshop. Venice April 5-8 2016. DOI: 10.1109/MARCH.2016.5, ISBN: 978-1-5090-2573-2. IEE

Thekkilakattil, A. and Dodig-Crnkovic, G., Ethics Aspects of Embedded and Cyber-Physical Systems In IEEE Proceedings of COMPSAC 2015: The 39th Annual International Computers, Software & Applications Conference, Symposium on Embedded & Cyber-Physical Environments (ECPE). Taichung, Taiwan - July 1-5, pp. 39-44, 2015. DOI: 10.1109/COMPSAC.2015.41