

How Can Neuroscience Enhance System Safety?

Quentin Montardy
InMind France
14000 Caen, France
 0000-0002-3617-6038

Katia Potiron
KNDS France
18023 Bourges, France
 0009-0008-4631-9187

James Inge
National Armaments-Materiel
Bristol, UK
 0009-0001-5481-5091

Abstract— Nowadays systems are dynamic and increasingly complex, taking the human into account into the system is needed and human factors are increasingly integrated into system safety. But there is a difference between human factor and understanding the human inside the system, as part of the system. This is particularly true for stress, fatigue, or heat situations where often the system safety relies at least partly on the appropriate human behaviour. This position paper explores how neuroscience can help to have a better understanding and integration of human as part of the system in system safety.

Keywords— *human factors, neuroscience, system safety, systems engineering*

I. INTRODUCTION

Traditionally humans always had a role to play in safety, often being part of the safety strategy as the last barrier between a problem and a disaster. In some domains some average profile is considered, in others training is particularly intense to condition the human to have the expected reactions to situations. But humans are still often considered outside the system of interest or limited to a “faulty system element”.

Moreover, systems are becoming increasingly complex and volatile, often combining elements from different technologies and application domains and changing the perception the human can have of the environment. What impact does this have on human performance?

In many safety analyses, the view of humans is limited by an implicit hypothesis that humans all behave the same and will always behave the same in all “states”, as if stress or tiredness will not change the risk perception or risk acceptance of the human part of the system. But neither systems nor their contexts are static; the world is basically VUCA (Volatile, Uncertain, Complex and Ambiguous) and so are systems.

The neuroscience and human factors disciplines have matured along complementary paths: neuroscience building deep, mechanistic insight into exploitable brain and physiological signals but still largely confined to the lab and the clinic; human factors delivering robust, field-ready methods that have not yet fully drawn on these recent advances. Framing the operator’s state using dynamic risk management is an opportunity to bring the two disciplines closer and let each renew the other.

II. VIEWPOINTS AND IDEAS

A. System Safety Viewpoint

Safety strategies and safety cases often rely on humans as system elements at some point but do not always take them into account as a part of the system. This turns an operator from actor into a spectator yet leaves them responsible for detecting emergencies and taking over control – engaging a completely different set of cognitive tasks, which are still to be integrated into system safety.

System safety needs to mature its ability to consider humans. It needs to move from considering them as outside the system (and so not fully taken into account) to a more

systems engineering view that integrates knowledge of the human state and takes it into account to build the most appropriate system safety strategy. To do so it is necessary for system safety to be able to assure and ensure that a human system element can fulfil its mission. Hence the human role in the system must be studied, explainable and targetable.

B. Human Factors Viewpoint

For many years, HF practitioners have realised the tension between the need to remove humans from systems to avoid human errors, and to give humans sufficient control of unexpected events to exploit their adaptability. Various authors [1]–[4] all discuss the need to balance the tasks or functions allocated to humans with those allocated to engineered system elements, playing to their respective strengths.

Rasmussen [5] suggests that human error can be seen as a mismatch between the tasks demanded by a system design and the operator’s abilities, performance, or limitations. The system needs to be designed to ensure that humans have the understanding, physical capability and situational awareness to take appropriate actions when called upon.

C. Neuroscience Viewpoint

Modern neuroscience brings out new ways to monitor how a system affects the operator’s cognitive and emotional state, whether read as discrete states or followed continuously. Every change of cognitive state reshapes his brain’s priors (the expectations it brings to what will be perceived from the environment), and with them his ability to respond to a threat effectively and in time. Those behavioural shifts are not noise to be averaged away, they must be taken into account.

Modern neuroscience investigates brain transitions between states, how such transitions can be detected and, more importantly, what function they serve. For example, negative emotional states such as anxiety or fear are often treated as ‘hijacks’ of behavioural control that leave a human system element unable to respond appropriately to other system element failures. Yet the same human states can be understood as adaptive: they tune perception and prioritize the anticipation of threat, a functional view that has been interpreted through the lens of Bayesian inferences [6], [7], and been formalized as approximate algorithms for survival-relevant decision-making under uncertainty [8], [9]. Following this idea, human system element affective or cognitive states are firstly a signal to be read (a substrate for dynamic risk assessment) and secondly a behavioural feature that changes the way human systems elements interact with their surroundings (system, team, etc.), which should be considered during system architecture and design.

III. PUTTING EVERYTHING TOGETHER

In IEC 61387, complex systems are defined as those where the collective behaviour of the system elements entails emergence of properties that can hardly, if at all, be inferred from properties of those system elements [10]. This means that emergent system-level properties like safety cannot be

determined by analysing the system elements independently: many safe parts do not imply a safe whole. In the defence sector, military capabilities are often assembled by combining parts such as sensors, weapon systems, vehicle platforms and communication links.

Brought together with human factors and neuroscience, this opens the way to construct a *system/human safety interaction space*: a multidimensional space that combines the states of the system with the states of the human system element. This could improve input to existing human factors and design tools: allowing the system to be built around the cognitive and emotional states the operator can go through, so that its design fits how they perceive, attend and decide in each state. At design time, this space is discrete and static, but during operation, those discrete maps extracted from the full system/human safety interaction space could be integrated with continuous measurement of the human system element's mental state. This turns it into a *dynamic system/human safety interaction space* (a projection of the interaction space from the perspective of the human state or one system state): one can then navigate the space through the appropriate map in real time, or even to predict and direct it, and let other system elements and the human system elements adapt to one another.

The starting point is to characterize and anticipate those interactions between cognitive states and other system elements states, i.e. how a human system element is likely to behave within a given system and context. This can be done in two complementary ways. The first, during system design/conception, is to describe the human system element state by state: what perception, attention and decision look like in each condition. We can chart this as a first, static space of a multidimensional system/human safety interaction space whose axes combine cognitive and system states. The second tracks the human system element state continuously through sensors (physio, brain, behaviour), as the work unfolds, treating cognitive state as an intra-individual variable that fluctuates within a mission.

The neuroscientific proposition is to couple the two, to use real-time measurement of the human system element state changes as a new sensing modality that reports his position in the system/human safety interaction space (precursor ideas in [11]): Coupling the static space built by design with the live signal turns it into a *dynamic safety space*, one that can be navigated (or predicted) in real time, so that system and operator adapt to one another as the mission unfolds.

This reconciliation of offline design and online measurement works on two levels. The first is representational and essentially statistical: the initial state-by-state characterization acts as a generative model, a decoder that gives the continuous signal its meaning, while the signal places the operator at a vector on the map, with a direction and a force/speed. Without the map the raw signal is ambiguous; without the signal the map stays static. The second level is functional: a negative cognitive or affective state is read in the light of the map, to update the joint system's estimated risk, and because the same physiological response can mean different things in different contexts, it is the offline map that resolves the ambiguity. This coupling of design-time analysis and run-time feedback is expected to improve safety.

IV. WAY FORWARD

Building such a system first needs to get information about cognition on the human side of the loop. The states of a system

may be known by design, while those of the operator are not, and we need to know which cognitive functions an operator uses within the safety space. To do so, one needs to gather cognitive multitasking information in a representative environment. Multitasking paradigms like MACE (Multi-Assets Cognitive Evaluation [12]), in the lineage of the NASA Multi-Attribute Task Battery [13] allow multiple cognitive tasks to run simultaneously, in a fully controllable simulated environment and synchronous physiological recording platform. This fills the human axes of the safety space beforehand, in controlled conditions, providing an offline map that later gives the live signal its meaning.

Further work will need to demonstrate that such evaluation is practical for real-world systems and show how it can guide designers to simplify human system element tasks and ensure they are manageable. This will become ever more important – and challenging – as systems incorporate AI elements that humans are asked to supervise.

V. CONCLUSION

To address the above challenges, a systems engineering approach can be used to manage the complexity of such systems, but co-engineering and trade-offs should include neuroscience results. Just as there is a need to integrate safety engineering more fully into systems engineering, there is a need to do the same for human factors engineering [14]. But there is also a need to integrate neuroscience.

REFERENCES

- [1] L. Bainbridge, "Ironies of automation," in *Automatica*, 19(6):775–779 Elsevier, November 1983. doi: 10.1016/0005-1098(83)90046-8.
- [2] J.T. Reason, *Human error*, Cambridge University Press, 1990.
- [3] N.G. Leveson, *Safeware: system safety and computers*, Addison-Wesley Professional, 1995.
- [4] N. Storey, *Safety critical computer systems*. Pearson Education, 1996.
- [5] J. Rasmussen, "Approaches to the control of the effects of human error on chemical plant safety," in *International Symposium on Preventing Major Chemical Accidents*. American Inst. of Chemical Engineers, 1987.
- [6] D. Grupe and J. Nitschke, "Uncertainty and anticipation in anxiety: an integrated neurobiological and psychological perspective," in *Nat Rev Neurosci* 14, 488–501, 2013. doi: 10.1038/nrn3524.
- [7] M. El Zein, V. Wyart, and J. Grèzes, "Anxiety dissociates the adaptive functions of sensory and motor response enhancements to social threats," *eLife* 4:e10274, 2015.
- [8] D. Mobbs, C.C. Hagan, T. Dalgleish, B. Silston, and C. Prévost, "The ecology of human fear: survival optimization and the nervous system," *Front. Neurosci.* 9:55, 2015. doi: 10.3389/fnins.2015.00055.
- [9] D. Bach and P. Dayan, "Algorithms for survival: a comparative perspective on emotions," in *Nat Rev Neurosci* 18, 311–319, 2017. doi: 10.1038/nrn.2017.35.
- [10] IEC, "IEC 61317-1: Systems engineering – System safety – Complex systems in defence programmes Part 1 – Concepts, terminology and requirements," Committee Draft for Vote 65A/1187/CDV, Geneva: International Electrotechnical Commission, September 2025.
- [11] F. Dehais, A. Lafont, R. Roy, and S. Fairclough, "A neuroergonomics approach to mental workload, engagement and human performance," in *Front. Neurosci.* 14:268, 2020. doi: 10.3389/fnins.2020.00268.
- [12] MACE: proprietary cognitive multitasking battery developed by InMind; methods paper in preparation.
- [13] J.R. Comstock, Jr. and R.J. Arnegard, "The multi-attribute task battery for human operator workload and strategic behavior research," No. NAS 1.15: 104174, 1992.
- [14] J. Inge, K. Potiron, M. Carter, and I. Hartwell, "Integration of Human Factors into complex systems safety," in *Safety-Critical Systems Symposium 2026*, York, UK: Safety-Critical Systems Club, February 2026. doi: 10.65391/r3347.