On the Evaluation of Risk Acceptance Principles

H. H. Kron[a,1]

[a] Dr. Kron, Safety Assessor for Railway Signalling and
Maglev Operation Control Systems
Am Mandelberg 1, D-76831 Birkweiler, Germany

Abstract: According to the CENELEC standards EN 50126, 50128, and 50129 [3-5], safety assessment and acceptance is based on risk analysis, and risk analysis is based on risk acceptability („risk acceptance“) principles. There are a few proposed risk acceptance principles (e.g. ALARP, GAMAB, MEM). Others may be derived from railway acts and ordinances (e.g. MGS and NMAU). A closer look shows that they all have drawbacks and limitations. So, the first step must be the definition of properties and evaluation criteria for risk acceptance principles. The second step would be to evaluate the known principles or combinations thereof. One result is obvious: „Risk“ may be defined as the product of hazard probability and hazard severity, but „risk acceptance“ is definitely more complex. Simple safety targets like TIRF (tolerable individual risk of fatality) and THR (tolerable hazard rate) are not sufficient for risk acceptability.

Key Words: Risk Analysis, Risk Acceptance, Risk Acceptability, Safety.

[1] E-mail: dr.kron@t-online.de

1 Introduction

1.1 The Problem

The CENELEC standards EN 50126 and prEN 50129 [3,5] define „safety“ as the freedom from unacceptable levels of risk. So the old question „how safe is safe enough?“ is reduced to the question „what levels of risk are acceptable?“. For an objective, comprehensible and standardized treatment of this question, risk acceptance principles (RAPs) have to be defined.

The CENELEC standards, however, are not very helpful in choosing risk acceptance principles. EN 50126 recommends that a generally accepted principle should be used (4.6.3.3) and gives three examples: ALARP, GAMAB, and MEM (see Subsection 1.2).

In the context of risk analysis, RAPs are applied to the identified hazards,

1. to decide whether a particular safety device, safety function, or safety rule is necessary or not (if it is not already mandatory by other regulations),

2. to determine the safety integrity level (SIL) for the specified safety devices and functions, and

3. to determine other safety-relevant parameters like overlap length, height of railings, braking values, safe reaction times, etc.

Thus, RAPs have great influence on the design process and the quality of the final signalling system. RAPs might boost or inhibit technical progress. RAPs might channel limited resources to the wrong problems, for instance spending more effort on controlling random hardware faults than on controlling systematic errors in specifications and software, or on controlling operator’s errors during degraded levels of system operation.

1.2 Sample Principles

EN 50126 [3] introduces the risk acceptance principles ALARP, GAMAB, and MEM:

- ALARP („as low as reasonably practicable“): between the region of unacceptable risk and the region of broadly accepted risk, there is a tolerability region where risk is undertaken only if a benefit is desired and where each risk must be made as low as reasonably practicable.

- GAMAB („globalement au moins aussi bon“): all new guided transport systems must offer a level of risk globally at least as good as the one offered by any equivalent existing system.

- MEM („minimum endogenous mortality“): the individual risk due to a particular technical system must not exceed 1/20th of the minimum endogenous mortality.

The German ordinances BOStrab § 2 [1], EBO § 2 [2], and MbBO § 3 [12] allow to use the principle

- MGS („mindestens gleiche Sicherheit“ / „at least the same level of safety“).

MGS allows deviations from the generally accepted rules of technology if at least the same level of safety can be demonstrated by an explicit safety case. This is similar to GAMAB, but not the same. Taken literally, MGS only applies to *deviations* from rules and standards and is not meant to serve as a general risk acceptance principle independent of rules and standards.

BOStrab § 3 [1] introduces a general risk acceptance principle:

- NMAU („nicht mehr als unvermeidbar“ / „not more than unavoidable“).

The NMAU principle states that nobody‘s risk, as imposed by the normal operation of installations and vehicles, should be greater than unavoidable. In this sense, safety is the freedom of *avoidable* levels of risk. German jurisdiction accepts the *unavoidable* residue of the hazards typically associated with the operation of railways because of the associated social benefit [15]. A hazard is not *unavoidable* if it can be avoided by *reasonable*[1] safety measures [15]. This leads to considering the public benefit of the railway operation and the reasonable costs[2] of safety measures.

Now it is clear that MEM does not logically include NMAU. So in the realm of BOStrab, MEM is not the first choice. This example indicates that the risk acceptance principle must be chosen very carefully and that a judicious choice depends on more than mere technical aspects.

1.3 Overview

Using the categories proposed by Schnieder [14], we will draw up a set of aspects for the evaluation of risk acceptance principles. The set includes social and legal aspects (Section 2), formal and technical aspects (Section 3), and dynamic aspects (Section 4). For instance, a particular RAP might be formal, quantitative and explicit, but not constitutional. Another principle might be legal, but its results might not be reproducible by independent groups of experts. After introducing the aspects for evaluation, we will apply them to a sample principle (Section 5).

[1] dt. zumutbare

[2] dt. zumutbare Kosten

2 Social and Legal Aspects

2.1 Constitutionality

In some countries, risk limits can only be fixed by legislative bodies, not by standardization bodies [9,11]. According to Leißner et al. [10], the risk acceptance criterion cannot be decided by suppliers or railway authorities alone, but is a sovereign[3] task. The RAPs already given in transport acts cannot be circumvented by risk limits allocated by other bodies than national legislation. Setting risk limits in standards or requirement specifications might be illegal. In Germany, there are no legally fixed risk limits or tolerable hazard rates [15].

2.2 Legality

Legal requirements differ from country to country, and even within a single country there might be different acts for different guided transport modes or systems, e.g. state railways, urban or regional transport, industrial railways, maglev lines.

A RAP may not be legal, even if it is recommended by a standard. It must conform with acts, ordinances (e.g. BOStrab [1]) and regulations. In case of conflict, the RAPs implied by law and legal liability prevails over any RAP recommended by a standard.

Another legal aspect is that any RAP should also be *technically* applicable under the given ordinances and regulations, which may limit the solution space and thereby prevent an optimal risk allocation. Under EBO, permissible speeds of more than 50 km/h on secondary lines require the interlocking of facing points, without exception and regardless of any risk analysis. The RAP should then be technically applicable to the remaining hazards under the precondition that facing point interlocking is compulsory.

2.3 Reasonableness[4]

Simple RAPs only consider risk parameters like hazards, consequences, and their probabilities. However, in most laws, regulations and court decisions, risk acceptance is not a function of risk parameters only, but takes into consideration:

- the public benefit of the potentially hazardous operation,
- the reasonableness of lifecycle costs of safety measures against hazards,
- the operational reasonableness of safety measures, and

[3] dt. hoheitliche
[4] dt. Zumutbarkeit

- their impact on system availability and maintainability.

The obvious disadvantage of „reasonableness“ as a risk acceptance principle is that it is not independent of the safety measures against the given hazards. The solution space of safety measures must be known to be able to estimate the lifecycle costs.

Examples for the application of reasonableness are

- the NMAU principle as required by BOStrab § 3 and

- Requirement 2 of risk analysis as defined by EN 50126, 6.3.3.2 [3]:
  „Requirement 2 of this phase shall be to determine and classify the acceptability of the risk associated with each identified hazard, having considered the risk in terms of any conflicts with availability and lifecycle cost requirements of the system.“

According to German jurisdiction, resonableness shall be judged „ex ante“, i.e. without reference to the particular case of damage and its location, not „ex post“, i.e. with full knowledge of the circumstances and the location of the particular damage [15]. If, for instance, a particular level crossing accident could have been avoided by certain safety measures (say, by additional half-barriers), the reasonable costs of these measures must be judged by considering *all comparable* level crossings, not just the one where the accident actually happened.

The legal and social importance of reasonableness is demonstrated by the following hypothetical case study. Assume that an identified hazard H_i could be controlled

a) only by means of an additional, costly hardware subsystem.
b) just by a few lines of application software in the existing system.
c) only by operational restrictions severely reducing the performance of the system.

Assume that the associated risk was estimated and found to be acceptable. Thus no safety measure was implemented against H_i . What would be the legal assessment of an accident due to H_i? In case b, not implementing a few lines of code would of course be considered an act of gross negligence[5], since the risk would be considered avoidable.

2.4 Legal Certainty[6]

Legal certainty is an essential requirement for any risk analysis. First, the safety assessment and the final approval by the safety authority are based on the validity of the risk analysis.

[5] dt. grobe Fahrlässigkeit
[6] dt. Rechtssicherheit

Second, only a valid risk analysis protects the railway authority and the supplier against liability claims and criminal prosecution in case of accident.

Legal certainty can be created by strict adherence to the law. In most cases, however, the legal requirements are not formal in a mathematical sense, but refer to qualitative (non-measurable) parameters. So, legal certainty depends on a convincing transformation between the qualitative requirements of the law and the RAP used.

Note that probabilistic calculations are based on a formal framework, but not always on convincing statistical data. And even proponents of the MEM principle admit that the risk allocation and reduction factors used in deriving tolerable hazard rates are *arbitrary* [7].

2.5 Reproducibility

The opposite of „arbitrary" is „*reproducible"*: The application of the RAP by two or more independent groups should produce similar results. In this context, Leißner et al. [10] give a report on the attempt to derive a common ETCS risk acceptance criterion for both DB and SNCF. Remarkably, the target is just one simple quantitative value TIRF (tolerable individual risk of fatality); still, the national results differ by two orders of magnitude.

Note that quantitative approaches are no guarantee for reproducibility. Reproducibility depends - among other factors - on exact and reliable input data, on the clear definition of the parameter domains, on not having too many „adjusting screws" (reduction factors), and on correct application of methods and tools.

Reproducibility is related to interoperability (see Section 3) and to stability (see Section 4).

2.6 Validity

Some RAPs do not support the most basic legal distinctions like slight or gross negligence, intent or malicious intent - in the actions of personnel, passengers (including suicides), and third parties (e.g. at level crossings or within communication networks). But we must decide whether or not the associated risks (say, of malicious intent) should be included in the statistics, the calculations and the risk limits. A risk excluded from risk analysis and hazard control is, in effect, an implicitly accepted risk which does not show up in the calculations. There is the danger that statistical data and quantitative risk analysis do not match.

Another aspect is the validity of risk limits. What does a tolerable hazard rate (THR) of 10^{-9} hazardous events per hour really mean? Certainly it is more safe than a THR of 10^{-8} h^{-1}, but neither value guarantees that the hazardous event will not happen tomorrow or during system lifetime.

2.7 Personal and Third-party Responsibility

Many papers on risk acceptance have stressed the difference between the personal responsibility for one's own safety and third-party responsibility. The public accepts very high levels of risk in the pursuit of individual hobbies, high levels of risk for activities with predominately individual responsibility (like road traffic), low levels of risk for predominately third-party responsibility (like public transport), and very low levels for exclusively third-party responsibility (like living near chemical or nuclear power plants). Some risk acceptance proposals are ignorant of this and suggest that the individual risk due to public transport should equal the individual risk due to any other hazardous technology, say 1/20 * minimum endogenous mortality. This neglects the modal differences in responsibility *and* the fact that rail traffic is 75 times safer than road traffic (in Germany, 1981, based on passenger fatalities per travel distance [6]).

2.8 Transparency

Social risk acceptance depends on the perception of risks, which might differ from quantitative and the ease of safety demonstration. For achieving social risk acceptance, Krieg [8] defines three transparency criteria:

- *familiarity*[7], i.e. long-term experience with hazards and safety measures,
- *comprehensibility*[8] of safety measures, and
- *verifiability*[9] of safety measures.

Familiarity favors general (generic) safety functions and solutions. For instance, at a level crossing the train has *in general* the right of way, even if a detailed quantitative risk analysis of a specific application were to recommend that there the train should give way. In the level crossing example, road barriers are more *comprehensible* then just a flashing light, even if a specific risk analysis might point out the risk of trapping a road user in the danger zone or the risk of collisions with the barriers. The risk reduction provided by an over- or underpass is more easily *verifiable* for the public than the dependability of an automatic (or even manually controlled) level crossing.

Note that transparency, like legal reasonableness, takes the safety measures into account, i.e. risk acceptance cannot be achieved without considering the solution space.

---

[7] dt. Vertrautheit
[8] dt. Verständlichkeit
[9] dt. Nachprüfbarkeit

2.9 Local Risk Acceptance

When renewing, modifying, or extending existing transport systems, it might be necessary to consider „local risk acceptance“. For instance, at the stations of the Schwebebahn Wuppertal there are no physical obstacles to prevent passengers from entering the danger zone between the station platforms and falling from there into the river below. This would probably not be acceptable for a new transport system elsewhere, not even for public streets, but it can be locally acceptable for a very long time.

3 Formal and Technical Aspects

Without going into details, we list some formal and technical properties of RAPs here.

Risk concept. An important aspect of any RAP is its *underlying formal risk concept*. Take, for instance, EN 50126, 4.6.1 [3] . For any hazard H_i, $risk(H_i)$ is defined as the product of the hazard probability and the hazard consequence: $risk(H_i) = p(H_i) * c(H_i)$. It should be clear that this is a simplification, perceiving a large risk as equivalent to the sum of many small risks, or a large damage as equivalent to the sum of many small damages.

Formal domains and operators. A hazard consequence $c(H_i)$ usually is defined to be a scalar like n_C = number of dead by train collision (see EN 50126, D.2.2). However, a vector of consequences or a set of qualitative severity levels might be more appropriate.

Quantitative or qualitative approaches. Some risk acceptance principles (like MEM) will lead to a *quantitative* risk analysis based on numerical targets like TIRF (tolerable individual risk of fatality). On the other hand, *qualitative* principles and methods refer to *Boolean values* (like fail-safe / not fail-safe) or small sets of *discrete values* (like the safety integrity levels SIL 0, SIL 1 - 4 [3-5]). Note that quantitative approaches are not necessarily more transparent, objective, reproducible, or stable than qualitative approches.

Granularity. Are real numbers necessary, or are a few distinct levels sufficient for dealing with the risk parameters? Will the result be more precise or more stable, if real numbers are used?

Explicit or implicit risk acceptance. The acceptable risk can either be stated explicitly as a quantity, or it can be implied in the amount of required risk reduction (say, in the safety integrity level of required safety functions).

Allocation and traceability. A quantitative „tolerable hazard rate“ or „target failure measure“ can be allocated and traced down to the stochastic properties of hardware components, but not to the „systematic“ correctness of system requirement specification, system architecture, module design, software coding, etc. According to prEN 50129 A.3 [5], systematic faults cannot be quantified. For most quantitative approaches, this poses a severe validity problem.

Modularity. Can we directly apply the RAP to subsystems, components or isolated safety functions? Or do we have to perform a complete risk analysis of the whole system in any case? Can we decide on a newly found risk (say, a remote risk caused by a software defect detected during commissioning) incrementally - or do we have to reiterate the whole risk analysis?

Scalability. A RAP should be applicable to systems of different scale, from a whole turn-key transport system like Bangkok Sky Train or Shanghai Transrapid, down to a small isolated safety function (like determining the set of acceptable signal aspects in case of a single lamp failure of the distant signal 3Vb in Drei Annen Hohne).

Interoperability. Cars are not designed just for particular towns or countries, depending on the local traffic density there. Likewise, we prefer RAPs which do not depend heavily on local risk parameters like numbers of interlocking elements or vehicles, traffic density, headways, etc.

4 Dynamic Aspects

The choice of RAPs may have dynamic effects on the „state of the art“ in signalling. For instance, the application of GAMAB or MGS may have a *retarding effect*, since there is no incentive for improving safety. Rather, the designer is motivated to achieve the same old safety level more efficiently. Some safety authorities are not willing to accept this approach. They compare it with the development of standards in other fields (e.g. civil engineering, fire protection). Nobody would justify modern solid-state interlockings without any track circuits or axle counters, on the grounds that there are so many old interlockings without automatic train detection.

Another criterion is *stability*. Is the chosen RAP able to cope with changes of the signalling plant (e.g. added lines, field elements, or vehicles) and with functional changes? Can it cope with gradual increases in train frequencies or passenger numbers, or reduction of headways? A related property is interoperability (see Section 3).

An often neglected dynamic effect is *risk compensation*, or *risk homeostasis* [8]. One aspect of risk homeostasis is that people will get used to technical safety measures and will compensate the actual risk reduction by lowering their vigilance and their adherence to rules („compliance“). For instance, the compulsory use of safety belts and safety helmets did *not* accelerate the reduction of road traffic fatalities as you might expect, but retarded it in comparison with countries *without* compulsory use [8]. Some people just drove faster, thereby compensating the risk reduction and, moreover, increasing the hazard to other traffic participants, especially to pedestrians (who, by the way, use neither safety belts nor helmets).

5 Sample Evaluation of the MGS Principle

The MGS principle „Deviations from the generally accepted rules of technology are allowed if *at least the same level of safety* can be demonstrated“ has the following properties:

1. It can be formalized.
2. It allows and requires to state the risk explicitly.
3. It admits both qualitative and quantitative approaches.
4. The quantitative approach first calculates the individual risk of a system complying with the generally accepted rules of technology and then takes this value as a risk limit for the new, deviating system.
5. MGS may retard progress by perpetuating just the old safety level.
6. MGS does not guarantee legal reasonableness, since simple and cost-effective changes to the new system might increase its safety, but might be ignored.
7. MGS does not guarantee cost-efficient new solutions, since the generally accepted rules of technology might be statistically safer than necessary.

MGS needs careful definition of the *items* (systems, operation modes, functions etc.) to be compared between the old, standard system and the new, deviating system. If the items are too small, it is easy to find a hazard scenario that is safely controlled by the old system, but not by the new, thus rejecting the new system. If the items are too large, MGS may finally turn into the GAMAB principle, provoking an elaborate and costly risk analysis of the whole signalling system.

So, MGS seems to be well suited for the comparison of base components (e.g. a programmable logic controller - PLC - as a substitute for relay circuits) and for the comparison of a single functional complex (e.g. the intrusion detection device for station tracks under driverless operation as a substitute for the driver’s vigilance).

Note, however, that MGS literally applies only to „deviations from the generally accepted rules of technology“. This kind of application is described in detail by Wittkowski [16]. By the way, Wittkowski does not consider MGS to be sufficient for all relevant cases; in fig. 2, he calls for additional measures against high risks and adds the principle of proportionality[10] as the final risk acceptance criterion.

The German ordinances BOStrab [1], EBO [2], and MbBO [12] *do not justify* the use of MGS as a general risk acceptance principle with the interpretation „a new system is considered safe if it has *at least the same level of safety* as an existing, approved system.“ This interpretation of MGS might perpetuate safety levels which are out-dated, arbitrary, or coincidental. Another possible pitfall of MGS is that essential differences between the new and the existing system are disregarded (including operational and environmental conditions) and that both systems are wrongly treated as equivalent.

[10] dt. Verhältnismäßigkeit

When applying MGS to an automatic system replacing a manual system, *risk compensation (risk homeostasis)* must be considered. For example, assume that under a manual block system, the vigilance[11] and compliance[12] of the personnel is sufficiently high. This results in the known hazard rate of the old system. A new, automatic block system is designed by applying the MGS principle to produce - theoretically - the same hazard rate. The vigilance and compliance of the personnel decreases during successful operation of the new system, however. Failures of the new system (e.g. a signal stuck at green) might then go unnoticed, resulting in a higher actual hazard rate. Therefore, the new automatic system must actually have a lower hazard rate than the old manual system.

6 Summary and Outlook

As a conclusion, it is obvious that some common RAPs, even those proposed in CENELEC standards, have severe deficiencies and limitations.

Also, several evaluation criteria - in fact the more important legal and social aspects - suggest that RAPs should consider not only the hazards, their probabilities and consequences, but also the properties of the corrective safety measures (e.g. costs, feasibility, reasonableness, transparency) and their dynamic aspects (e.g. risk compensation).

Simple safety targets like TIRF (tolerable individual risk of fatality) and THR (tolerable hazard rate) are not sufficient for risk acceptability. Risk acceptance must refer to solution space properties, too. Admittedly, this does not make risk analysis easier. Future work could be structured as follows:

1. Single out the most important evaluation criteria. “Most important” are criteria imposed by law, and criteria with direct relation to safety.

2. Design combinations of risk acceptance principles that meet those “most important” evaluation criteria.

3. Design risk acceptance graphs or flowcharts (as in [16]) which implement those combinations of risk acceptance principles.

[11] dt. Wachsamkeit
[12] dt. Befolgung von Vorschriften

Acknowledgements

I am grateful to Walter Eberhardt (Regierung von Mittelfranken) and Dr. Jürgen Kappus (Bezirksregierung Düsseldorf) for helpful comments on RAPs.

References

[1] BOStrab, 1987, Verordnung über den Bau und Betrieb der Straßenbahnen (BOStrab).

[2] EBO, 1967, Eisenbahn- Bau- und Betriebsordnung (EBO), 8. Mai 1967, last change 21.06.2002, BGBl. I p. 2191, BGBl. III 933-10.

[3] CENELEC, 1999, EN 50126, Railway applications - The specification and demonstration of reliability, availability, maintainability and safety (RAMS).

[4] CENELEC, 2001, EN 50128, Railway applications - Communications, signalling and processing systems - Software for railway control and protection systems.

[5] CENELEC, 2002, prEN 50129, Railway applications - Communication, signalling and processing systems - Safety related electronic systems for signalling, final draft, May 2002.

[6] Grottker, U., 1986, Die Gefährdungswahrscheinlichkeit als Sicherheitskennwert technischer Systeme am Beispiel des Eisenbahnbetriebs in Abhängigkeit der Verspätungsverteilungen von Zugfahrten (Schriftenreihe IVEV, Nr. 35, Technische Universität Braunschweig).

[7] Krebs, H., B. Le Trung, E. M. El Koursi & P. Firpo, 2000, Minimale Endogene Mortalität – ein universelles Sicherheitskriterium, *Eisenbahntechnische Rundschau ETR* 49 (2000) Heft 12, 816 - 821.

[8] Krieg, R., 2000, Risiken der Technik: ein kritischer Dialog über neue, am Menschen orientierte Denk- und Lösungsansätze (Raabe Fachverlag für Wissenschaftsinformation, Stuttgart, ISBN 3-88649-364-4).

[9] Kuhlmann, A., 1981, Einführung in die Sicherheitswissenschaft (Friedr. Vieweg & Sohn, Wiesbaden / Verlag TÜV Rheinland, Köln).

[10] Leißner, F., L. H. Hansen, R. Beck & K. Kammel, 2003, Erkenntnisse aus der Risikoanalyse für die ETCS-Pilotanwendung, *Signal und Draht* (95) 6/2003, 6 - 10.

[11] Marburger, P., 1979, Die Regeln der Technik im Recht (Carl Heymann Verlag, Köln).

[12] MbBO, 1997, Verordnung über den Bau und Betrieb der Magnetschwebebahnen (Magnetschwebebahn-Bau-und Betriebsordnung - MbBO).

[13] Pierick, K., 1980, Die „Allgemein anerkannten Regeln der Technik“ beim Einsatz von

Datenverarbeitungsanlagen mit Sicherheitsverantwortung im spurgeführten Verkehr, *Die Bundesbahn* 11/1980, 765 - 768.

[14] Schnieder, E., 1999, Methoden der Automatisierung (Vieweg Studium Technik, Braunschweig, Wiesbaden).

[15] Wittenberg, K.-D., 2002, Sicherheits- und Betreiberverantwortung im Eisenbahnbetrieb - Teil 1, *Signal und Draht* (94) 12/2002, 37 - 42; Teil 2, *Signal und Draht* (95) 5/2003, 32 - 41.

[16] Wittkowski, A., 1999, Abweichen von anerkannten Regeln der Technik, *Signal und Draht* (91) 3/1999, 10 - 14.