

Summary

Modelling Risk Control Measures in Railways

Analysing how designers and operators organise safe rail traffic

Jaap van den Top

Research question and relevance

The Dutch government and the railway industry wish to use the existing railway network more intensively. This will likely involve more interactions between trains, which could result in delays or in trains incidentally passing red signals more frequently than is currently the case. It is the main focus of this thesis to develop knowledge that helps to prevent these phenomena from happening.

Better understanding of how to cope with these interactions is required to see how the governmental wish can be realised, whilst the quality and safety levels remain acceptable. If the current performance level is not to be lowered and if disturbances cannot be taken away, the only way to achieve a higher network utilisation is to improve the control system's ability to cope with disturbances. Since these interactions are inherent to operational processes and how these are controlled, the research question is:

In what way could real time process control be improved to positively influence safety and quality of service in rail traffic operation?

Method used

To answer the research question, the following main steps were taken:

1. basic concepts from railway operation and safety science were analysed and gaps identified;
2. a framework, the so-called 'safe envelope of operations', was developed which can incorporate the main issues relevant to railway operation and safety science;
3. interviews and literature surveys were taken to identify to which extents current practices match the criteria from the framework;
4. options for improvement were explored.

Analysis of basic concepts from railway operation indicated that many functions have to co-operate in order to supply train paths. These are mainly infrastructure design, timetable design, traffic management, route allocation and train control. The cascade model can show these as a hierarchical set of functions. It indicates that in the current situation, degrees of freedom for operators become less as the time remaining to the moment of operation gets shorter. Also, it indicated that there is hardly any feedback from the lower-level functions towards the design stages.

Literature from safety science indicated that risks and benefits are two sides of the same activity: they cannot be considered in isolation, but must be seen in the context of normal operations. Also, it was shown that most existing accident models do not suffice to describe certain types of accidents in relation to normal system operation; they focus on safety only. Also, many accident models are based on probabilities of component failure or operator failure. Although useful, such models do not explain which mechanisms underlie systemic risks and what must be done to decrease the given probabilities. This requires understanding of the underlying processes.

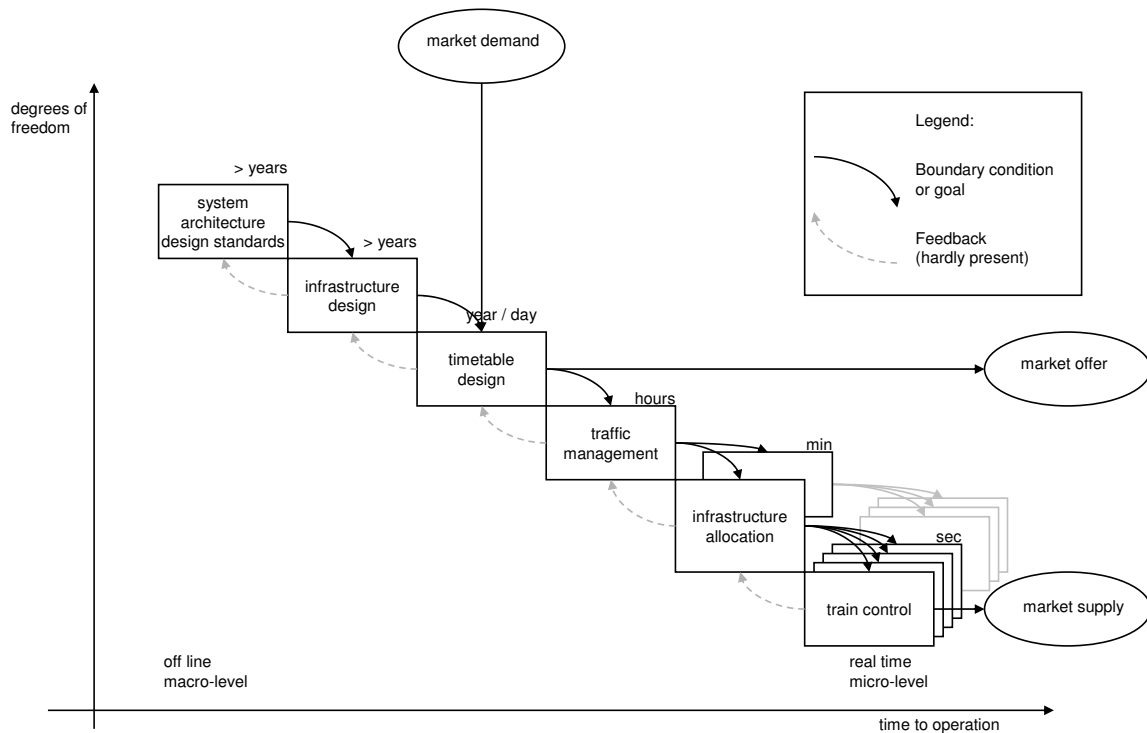


Figure S1: The cascade model.

Therefore, in this thesis, several existing and new concepts were combined and developed in order to understand these processes. This so-called 'safe envelope of operations' is a risk model that can combine safety risks with the purpose of an activity. It can represent operational processes and its critical boundaries and as such can help designers, managers and analysts to understand how risks and quality problems arise. The model indicates that actors controlling a system should be able to know the following items in order to fulfil their tasks safely and effectively:

- What is the current system state?
- What is the goal and which tolerances are allowed (operational envelope)?
- Can the goal be reached in time (attainability envelope)?
- Where is the boundary of safe operation (viable envelope)?
- What is the ultimate moment to exercise control (controllability envelope)?

These five elements can be visualised relative to each other, and each of these items can change over time: the envelopes are dynamic. Thus, the safe envelope model provides additional ways to think of possible accident mechanisms in dynamic systems. To keep the system within the safe envelope's boundaries, control must be exercised by the operators. Modelling this requires understanding of the technology, the humans operating the system, but also of the designers and managers who are responsible for the system as a whole. Although it appears that quite different fields of knowledge are involved, it was shown that concepts from the different fields can all be related to control loops:

- Control theory from the field of engineering;
- Detect-diagnose-act control sequence from safety science;
- Situation awareness from the field of human factors;
- Plan-do-check-adjust feedback loop from management science;
- Organisational learning from the field of management science.

It is convenient that the same principles can be applied to these different fields. This allows us to describe both safety and 'production' by taking an integrated functional approach to the system's variables that need to be controlled. This requires a clear visualisation of the boundaries within which the variables should be (this is the function of the safe envelope model) and it requires a model to describe the control processes that must keep the variables within these boundaries (for which the detect-diagnose-act sequences were used). All functions in the control model (as for example described in the cascade model) need to detect where the envelope's boundaries are, to diagnose the situation and to perform an action to influence the system state. Such interference in the system can be done either directly, or indirectly when the action provides a changed goal, a changed boundary, a new control measure to a lower level control functions, or if these are given a buffer or barrier to absorb disturbances.

In order to assess the controllability of the current railway system, several sources were used, comprising interviews with operational staff and designers, and literature from scientific and business sources. Since problems will show themselves in operations, this was the starting point for the analysis. Wherever necessary, the study looked back to earlier choices, for example in the timetable design or in systems design. To find underlying causes in the detailed data, a grounded theory approach was used.

Results

Twelve related underlying causes were found for performance problems (e.g. delays or passing red signals), which are shown in Figure S2. These underlying causes can be grouped as three vicious circles of causes and effects:

A. Individual loop

Since no clear goal is set (1), it is not really clear for many operators what exactly they should achieve. This either results in self-formulated goals (3) and self-developed practices (10), or to operators limiting themselves to their formal responsibilities (2). For several reasons, many operators are discouraged (7) and thus provide no feedback (8). Without feedback, there is insufficient organisational learning (9). Problems remain which further discourages operators.

B. Operations loop

Also, this causes operators to develop individual and therefore possibly incompatible methods (10), which are partly a cause of inadequate control information and the loss of motivation. This starts a self-reinforcing process of operators retreating to their own 'islands' with limited communication between them (12).

C. Design and management loop

This also makes it difficult for designers to understand how they can support operators with better support systems, regulations et cetera (11). This again leads to a shift from goal to means (3) and support systems that do not always adequately provide the operators with the information they need (4,5,6)

There is very little feedback on the basis of which actors in the cascade model can learn whether their decisions work in practice. This may be one of the reasons why there are different representations of the reality for different functions. For instance, in timetable design, traffic control and train driving, five different and in some situations incompatible ways of solving train delays were found. These five ways are all based on certain assumptions to make the task of the actor in question easier (at least for himself), but they do not always fit

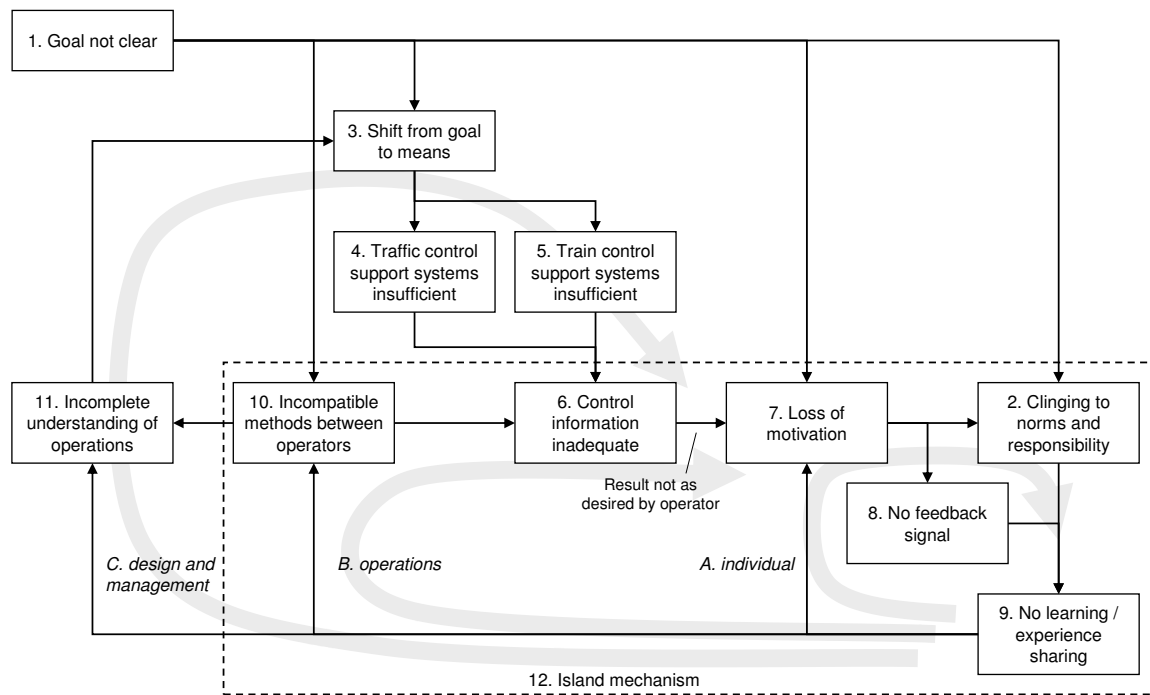


Figure S2: Cause-effect diagram for underlying causes of performance problems.

together. Many operators retreat to their own island in order to make their task manageable, but by retreating, they may cut off others from relevant information.

With regard to the traffic control process it is remarkable that this process is difficult to predict, even though the traffic controller's main function is to solve conflicts in advance. It would be helpful if a traffic controller had tools that provide him with an oversight of the situation as it is expected to be in 15 – 60 minutes and if conflicts in that picture can easily be found. The effort of a traffic controller would then not be spent on finding out what the situation actually is and where problems are about to occur, but on selecting solutions to solve these problems in advance. Although operator experience partially fills the gap of incomplete information, the result is that normally no updated, conflict-free, real time schedule is made in response to smaller deviations from the timetable. Thus, it is not guaranteed that there is a feasible production schedule, a fact that itself makes the process unpredictable.

Also the drivers have difficulties to perform their main task, which is to operate their train closely to their scheduled path. Apart from the fact that new conflict-free schedules are normally not made in response to smaller disruptions, the driver is also not informed about possibly changed schedules. Also, he cannot see in sufficient detail whether he does adhere and how he can adhere better to the (re)scheduled path, which requires a precision in fractions of minutes. These facts result in an unnecessarily high number of restrictive (yellow or red) signal approaches. Thus, preventing that signals are passed at danger is not just a human factor or technical (signalling) issue, but also an information exchange issue between traffic controllers reworking the plan and drivers executing it so that trains can be controlled such that they meet signals preferably only when they have cleared the route behind them.

Although timetable designers and traffic controllers usually rely on the driver to obey signals, imperfections in the route setting system, the signalling system and the train protection system, and the impossibility of faultless operator behaviour make this a dangerous assumption. Driver expectancy cannot be banned, since this is essential human behaviour. The focus should not be on prevention of expectancies, but effort should instead be spent to

preventing that these expectancies can become wrong, which is also an information exchange issue. One of the main problems is that the signalling system in several cases provides information that does not inform the driver about the unique scenario that he is going to enter, which leaves room for (wrong) interpretation. This puts stronger emphasis on the driver's ability to detect a red signal from a sufficient distance and relate it to his track, but it was shown that this is not sufficiently well possible in some cases.

Route knowledge, which is in fact also an expectancy, is in these cases the last line of defence in order to assure safety, but is only has a weak formal meaning. Moreover, the human memory works more reliably in response to external cues and should thus not be used to replace these. Therefore, although route knowledge is useful as general experience, it should not be relied on to ensure safe operations.

The belief that driver expectancy is the cause of many problems gives the impression that the system design is failure-free, and does not need to be changed, but this is not the case. Several examples were shown in which changes to the traffic control system, the signalling system, the train protection system or the radio system made the driver's task more difficult in an unexpected way. Since this was not the designers' intention, it must be concluded that their influence on the driver's task is not fully understood. Often, new technology was introduced without a clear view on how this would precisely support the operator's processes, or new technology replaced older technology and thereby, informal working routines were rendered infeasible.

Possible improvements

In the first place, the wish to 'better' use the system requires a clear cut answer to the question when the use is 'good', or in other words, what the system should achieve. The second question that should be asked is how the operational processes must be designed in order to achieve this result. This asks for redesigned processes in the first place, such as improved communication between traffic controllers and train drivers and improved situation awareness. This again will often require new technology in order to support the processes. The intended operating process should be designed first and the required technology should be derived from that, not vice versa.

The production processes, their planning and the means of production, such as the timetable, the rolling stock capabilities, the infrastructure layout and the safety system limitations must be geared to each other in order to fully exploit their possibilities. This asks for an integrated approach: if one of these elements is considered as a given and immutable fact, improvements in the other elements cannot be fully exploited. Since process control defines when which resources will be used for which process, it is inherently linked to the achievement of the goals and thus to the system's capacity. The current paradigm in timetable design, in which it is assumed that trains do not interact and a unique solution to operate them is nailed down a year in advance, does not take the effect of process control into account and therefore needs to be replaced. What is required instead, is a new timetabling paradigm in which the traffic pattern is structured in such a way that the real time process control has more and explicit freedom to reschedule the traffic pattern on the short term. The customer service intention, which should remain explicit, can then serve as a clear goal for the traffic control process, so that the quality of service is likely to be improved. Better operator support would steer the system more efficiently. This would then allow for a more efficient use of slack time and the ability to cut on slack times, thus leading to a higher utilisation. Since the system would then be operated closer to the boundary of what it can handle, constant monitoring of whether operational processes are running as desired and without side-effects would be necessary.

This means that not only real time traffic control is necessary, but also the traffic control and train control processes themselves must be monitored and analysed.

On an abstract level, solving quality and safety problems is all about avoiding surprises (undesired, unintended and unexpected results) during operations. To improve (safety) performance, the correct functioning of the processes performed needs to be made sure. Such insights can be gained by closing the feedback loop from operational level to design and management. Problems during operations should be freely reported by operators and actively be looked for by the organisation. For example, the organisation must measure how often safety-critical processes are executed, how often they failed and how often failures were stopped by the safety systems. Examples are the number of route setting commands refused by the interlocking, or the number of interventions by the train protection system. Such indicators make clear how well the operational process is actually kept away from its critical boundaries. Such a process-related analysis allows the finding of more possible causes for trains running through red signals than an approach in which causes are considered to be certain properties of signals or of drivers, since it also takes their interaction into account. As long as no proper analysis of current operations takes place, knowledge about the actual processes will not improve and neither will the processes themselves. Feedback of actual performance data to timetable design has recently been successfully started with; it is recommended that such feedback loops are also closed on other disciplines.

In spite of the errors addressed earlier, the current signalling system, in combination with a high level of professionalism of train drivers and traffic controllers, has resulted in railways being one of the safest transportation systems. However, if the public demands on the system increase and trains have to be planned even more closely together, the system will be taxed beyond its limits. Right now there are insistent signs, in the form of operators losing oversight over the situation and increasing numbers of signals passed at danger, that indicate that the limits of the current control philosophy are approaching. Therefore, a much more sophisticated real-time control philosophy is required, which must enhance operators' situations awareness so that all actors share the same expectation. The signalling system can then, again, be used as a backup system – the function it was designed for. Failures and unexpected surprises should be looked for and freely reported, so that operational processes can be geared to each other and so that designers can better understand how actual operations are being performed. This will not only improve safety performance, but it will also make it more likely that the promises to the customer can be fulfilled.