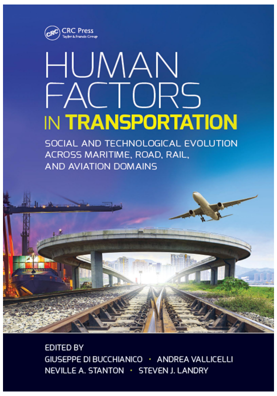

Human Factors in Transportation
Social and Technological Evolution Across Maritime, Road, Rail, and Aviation Domains

Giuseppe Di Bucchianico, Andrea Vallicelli, Neville A. Stanton, Steven J. Landry

Experimental Study for the Empirical Risk Analysis of Sociotechnical Systems in ATM

Publication details
https://test.routledgehandbooks.com/doi/10.1201/9781315370460-28
Giuseppe Di Bucchianico, Andrea Vallicelli, Neville A. Stanton, Steven J. Landry

How to cite :- Giuseppe Di Bucchianico, Andrea Vallicelli, Neville A. Stanton, Steven J. Landry. 25 Aug 2016, *Experimental Study for the Empirical Risk Analysis of Sociotechnical Systems in ATM from:* Human Factors in Transportation, Social and Technological Evolution Across Maritime, Road, Rail, and Aviation Domains CRC Press
Accessed on: 31 May 2023
https://test.routledgehandbooks.com/doi/10.1201/9781315370460-28

23

Experimental Study for the Empirical Risk Analysis of Sociotechnical Systems in ATM

Lothar Meyer, Katja Gaunitz, and Hartmut Fricke

CONTENTS

23.1 Introduction

The current methods for estimating the risk of sociotechnical systems in air traffic management (ATM) mostly rely on accident and incident reports, expert judgment, or model-based approaches. The predictive risk estimation of novel systems, in particular, is traditionally performed by the subjective adaptation of the expert's operational experiences to the expected operation after the hypothetical start-up of the target system. In this regard, the term risk complies with the definition: "*Risk is defined as the probability that an accident occurs during a stated period of time*" (Blom et al. 2003).

The most promising model-based approaches offer the advantage of coping with enormous sample spaces, by providing objective data and the statistical power to prove even very little probabilities of the accident event, for example, the target level of safety in ATM with a maximum of 1.55E-8 accidents per operating hour (Blom et al. 2001). An exhausting validation of all modeled *a priori* assumptions regarding the safety effects on a new design in realistic operating conditions is extremely challenging as there are usually no means of obtaining and transferring a direct evidence from the current systems and operations: *"errors are likely to be made when designers apply error modeling techniques"* (Johnson 1999). This might impair the external validity of the model for unknown or unexpected cases.

For the problem described above, human-in-the-loop simulations (HITLSs) offer an empirical approach that is often used for estimating the performance of sociotechnical systems in a predictive way, for example, by means of workload measures. HITLS has also been successfully used to accompany failure modes and effects analysis (FMEA) studies, namely to quantify isolated probabilities in the interaction between the operator and the working environment as well as human error probabilities (HEPs) that can be used for the quantification of model parameters (Stroeve et al. 2013). In contrast, a pure HITLS approach is rarely used only for risk analysis. This is due to the enormous efforts needed to obtain valid data as well as to the limited sample spaces that can be achieved in real-time simulation (Shorrock et al. 2001). Studies that involve operational experts perform a few hundred hours of simulation time at best (Stroeve et al. 2013), providing insufficient statistical power for a reliable elimination of rare and risk-inducing events. This is expressed by the ATM-safety iceberg in Blom et al. (2001). When applying statistical testing, the type-I error rate would be unacceptably large, in the case that an unsafe system is assumed as a null hypothesis. This error can be explained by the *Weak Law of Large Numbers*, also known as *convergence in probability* or, more specifically, Bernoulli's theorem. It describes a decreasing difference between an observable frequency and the true probability with increasing sample spaces. The difference is a quantifiable metric of the type-I error rate and can be estimated by *Chebyshev's inequality*. When assuming one operating hour as a basic unit of the population that has the end-state *accident* or *no accident*, with an underlying *binomial distribution*, the error could be estimated as

$$P(|X-\mu|\geq k)\leq \frac{n\cdot p(1-p)}{k^2}, \tag{23.1}$$

with the random variable X, the mean μ, the variance of the distribution $\sigma^2 = n \cdot p \cdot (1-p)$, and the confidence tolerance level k. Even with a sample space of 1.0E9 hours and a target safety of one accident per 1.0E9 operational hours, there is still a 13.6% probability to declare an unsafe system as safe when no accident has been detected in the experimental time. For instance, 3.0E9 operational hours are needed for gaining 95% confidence. Thus, the empirical approaches to cope with such rare events suffer from practicability to prove the novel system by means of HITLS.

Our proof-of-concept study is based on an approach by which insufficient sample spaces are compensated by intensifying the probability to detect safety indicators and by which, therefore, the power is increased with samples that are held constant. Hence, it addresses a problem definition of Swain: *"the problem remains of how raw data from training simulators can be modified to reflect real-world performance"* (Swain 1990).

As a possible solution to this problem, we developed a concept called *accelerated risk analysis* (*AccSis*), which describes a methodology to gain the desired acceleration effect needed for intensifying the probability of safety-relevant occurrences and the related safety metrics. This acceleration effect shall practically be reached by the induction of a calibrated time pressure that stimulates the occurrence of human error. Concerning the time-pressure induction, we developed a procedure following the *time-budget (TB)* principles (Bubb and Jastrzebska-Fraczek 1999, Bubb 2005) named *competitive performance* (*ComPerf*). It puts the test person under the impression of not having sufficient time to solve the problem (Chang and Mosleh 2007). This approach is motivated by the *accelerated-life-testing* (ALT) methodology, which forward the *mean time to failure* into the experimental period by means of an accelerated and calibrated stress induction during the experiment (Nelson 2009). Hence, AccSis explicitly addresses the occurrence of *right censoring* (Cox 1972).

Referring to this reasoning, this chapter presents the *time-pressure–risk model* and the related conceptual framework, named *AccSis* in Section 23.2. The primary subject of investigation is the problem of how to adapt the stochastic methods from ALT to the risk analysis of sociotechnical systems in ATM, considering the stochastic human behavior instead of stochastic processes of product aging. A HITLS experimental design is presented for the evaluation of AccSis and *ComPerf* following an innovative advanced surface movement guidance and control system (A-SMGCS) for air traffic control in the scope of a proof-of-concept study in Chapter 3. Three test persons were trained and qualified for conducting the experiments, whose individual results served for evaluating the hypothesis of the *time-pressure–risk model.* This chapter discusses the findings identified in the results of the HITLS and delivers insights in the effects of the stress-induction procedure indicated by means of the detected runway incursion (RI) frequencies and reaction times, all of which are outlined in Section 23.4.

23.2 Methodology

23.2.1 The Concept of AccSis

This conceptual framework has the objective of estimating the compliance of sociotechnical systems with a given target probability of an accepted safety metric (e.g., the accident), expressed as the alternative hypothesis $p < p_{target}$, by means of HITLS-based empirical data. Facing the problem of mitigating the statistical type-I error starts with analyzing *Chebyshev's inequality*. The mitigation can proceed as follows:

1. By increasing the number of generated samples n.
2. By modifying the simulated working conditions in the experimental design that rescales the probability by an acceleration factor a. A symmetric and linear rescaling of the target safety p_{target} and the true probability of the system p by the acceleration factor leads to $\hat{p}_{target} = a \cdot p_{target}$ and $\hat{p} = a \cdot p$ in which the alternative hypothesis is maintained. Applying the rescaling to *Chebyshev's inequality*, an effective mitigation of the type-I error can be determined as follows:

$$\frac{n \cdot \hat{p}(1-\hat{p})}{\hat{k}^2} = \frac{\hat{p}(1-\hat{p})}{n \cdot \hat{p}_{target}^2} = \frac{1}{a} \cdot \frac{p(1-p \cdot a)}{n \cdot p_{target}^2} \approx \frac{p}{a \cdot n \cdot p_{target}^2},$$

   with $\hat{k} = n \cdot \hat{p}_{target}$. When defining $p \ll 1$, one can approximately assume $p(1 - p \cdot a) \approx p$. The mitigation effect of the error can be quantified to a^{-1} and affects a virtual accumulation of the samples generated, described as $a \cdot n$.

The second approach thus constitutes an approach to face the safety-iceberg problem by describing a procedure that accelerates the convergence of the type-I error by modifying the boundary conditions of the HITLS. This affects a calibrated rescaling of the target and the system probability for safety-relevant events.

In reliability testing, the acceleration effect is practically achieved by a stress induction of, for example, thermic or mechanic stress that forwards the targeted failure event into the experimental time. In this way, the problem of *right censoring* is addressed, which describes the problem of measuring the time of an event that lies beyond the experimental time (Nelson 2009). The approach of ALT can be split into two tasks:

1. Failure stimulation—The experiment is to be conducted under varying gradations of stress, which deflect the load from design stress to accelerated stress. Three gradations of load are practically recommended for capturing sufficient samples of failure events of the product in ALT.
2. Regression analysis—The failure distributions of each load level are fitted to analytic or nonparametric distribution models. A regression model (life-stress model) is to be applied that extrapolates the trend of the distribution shape to design stress (see Figure 23.1).

The idea to adapt this concept to accelerate the occurrence of safety-relevant events in HITLS is severely impaired by the fact that human performance is a complex field that suffers from nonlinearity and nonreplicability compared to the functionality of technical products. For this reason, we identified the systematic differences between the analysis of product failure events and the commitment of errors by operators, when acting in a sociotechnical system.

- The most-significant difference is the stochastic that contrasts accident events of sociotechnical systems and technical failure events. The product lifetime is temporally limited as a result of the progress of aging that is attributed by a dependent stochastic distribution (lifetime distribution). In contrast, we assume the accident event in aviation to be the result of a failed operator's decision, which, hence, is regarded as an independent event with a limited temporal relation to the preceding operational actions and in which a distribution cannot be modeled over time when assuming a Bernoulli distribution for accidents.
- The second difference, which is the fact that the stress-inducing procedures of ALT are completely incompatible with sociotechnical systems, is related to the first one.
- The third difference is the missing accident-stress model for human behavior for the regression analysis, since state-of-the-art models, although describing the relationship between human error and stress, fail to deliver a domain-specific model curve.

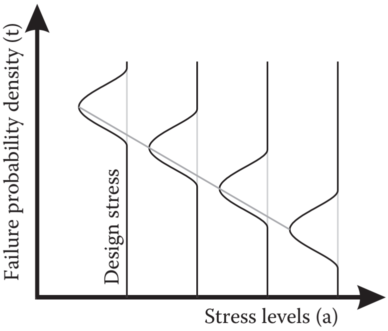

FIGURE 23.1
Stress-life relation according to the ALT concept.

This chapter considers AccSis to be the subject of a long-term validation strategy due to the reasons given above. Therefore, our current research follows a stepwise validation strategy to overcome the mentioned differences, in which finally a full compliance of AccSis with the requirements of the risk analysis of sociotechnical systems shall be achieved. On the basis of this consideration, we chose the first step to be a proof-of-concept study: the controlled acceleration of safety-relevant events by intensifying human error. The first objective is, hence, to gain an understanding of the principles of applied-stress induction while being constrained by existing procedures of the working environment in ATM.

To explain our choice of human errors as the key factor, we refer to the integrated risk picture (IRP), which describes the contribution of human errors to accidents in the combination of causal factors by means of a fault-tree model (Spouge and Perrin 2006). For a sociotechnical system, the IRP can be regarded as a significant fingermark of risk, in which branches of failure catenation form the resulting accident probability. One has to bear in mind that only branches affected by the acceleration effect are taken into account for this study.

When considering causal factors in the context of AccSis, organizational, technical, and human errors can be distinguished as the principal accident causes. This complies with Reason's "a trajectory of accident opportunity" that models the human error propagation in the presence of corresponding hazards as unsafe acts (Reason 1990). Human error has been identified as the most-frequent contribution to accidents and incidents in aviation with a share of 60%–80% (Shapell and Wiegmann 1996) or 75%, respectively (Müller 2004). The focus on human error thus addresses a causal key factor of sociotechnical systems: the major contribution of human error to risk. A vast amount of causal branches must hence be covered by acceleration. Following the ceteris-paribus principles, procedures, tasks, and other boundary conditions are to be held constant during HITLS that implies a major requirement on seemingly unimportant contextual conditions of the simulation.

23.2.2 The Role of Time Pressure for the Stimulation of Human Error

Besides uncertainty, time pressure seems to be of particular relevance when considering human decision-making processes (Rastegary and Landy 1993). Rastegary defines time pressure "*… as the difference between the amount of available time and the amount of time required to resolve a decision task.*"

According empirical findings, time pressure is known to significantly affect human performance (Freedman et al. 1988). This relation points to the vital impact of time pressure on human performance, that is, on acting correctly according to the procedures. This influence can be explained by the fact that the performance of cognitive information processing is a function of time pressure that affects a minimization of cognitive effort in a cost/benefit frame of reference. It is reported that an increased selectivity of information is observable. Under time pressure, more pieces of information are used but in a shallower way (Edland and Svenson 1993).

Time pressure contributes more to HEPs than additional tasks when performing time-critical tasks (Bubb and Jastrzebska 1999). Therefore, a *TB* was defined, which puts the time-available t_a into relation to a time needed for decision t_n, as follows:

$$TB = \frac{t_n}{t_a}. \tag{23.2}$$

An increased error probability was measured by a factor of 14 under the condition of time pressure. This observation corresponds to the assumptions of the Human Reliability Assessment THERP, which considers a factor of 10 under stress conditions (Swain and Guttmann 1983).

Time pressure and human error are causally linked and can be transferred to a continuous quality metric for human actions that is ultimately classifiable as acceptable or not acceptable. Specifically, the deflection of actions below a minimum quality can be regarded as not acceptable or, in line with conventional theories, human error. Continuing, quality is linked to performance as follows:

$$P = \frac{Q}{t}, \tag{23.3}$$

with the human performance P, the quality of human action Q, and the time given t (Bubb and Jastrzebska-Fraczek 1999). Thus, time pressure affects Q, divided by time. We identified the definition of *TB* as an inherent advantage for the stimulation of human error for two reasons:

1. It induces a calibrated time pressure by setting t_a
2. Human performance is sensitive to time pressure

To summarize, the concept of accelerating the occurrence of accidents unifies many theories about accident causation and human error to a comprising causal catenation, as shown in Figure 23.2, with each of the links being already empirically validated by the elementary findings (Freedman et al. 1988, Reason 1990, Bubb and Jastrzebska-Fraczek 1999).

The introduced concept for utilizing the *TB principle* to stimulate time pressure and hence human errors to thus intensify the probability of accidents is a summative generic

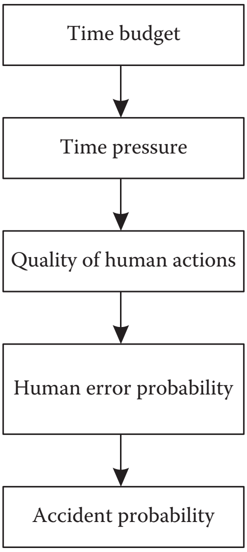

FIGURE 23.2
Causal relationship between TB and the accident probability.

description of the effect mechanism. It is necessarily a domain-specific challenge to develop a procedure that produces a *calibrated* time pressure by means of this principle.

23.2.3 Competitive Performance

Most ideas for the implementation of an induced time pressure aim at setting boundary conditions to effectively shorten the available time. Secondary tasks might, for example, shorten t_a by forcing the operator to organize task sharing and prioritization according to the time constraint. This sharing will as well change the pattern of activities and impact the IRP picture without any control of the deflection from the design stress. The same holds true for the conventional means of HITLS calibration, namely the intensification of the task load, for example, the traffic volume.

As time pressure is transformed from an objective condition to a subjective feeling, we decided to choose the approach of a "competitive arousal." Following this approach, time pressure is generated by providing a competitive environment that triggers the desire of the operator to win (Kerstholt 1994, Malhotra 2010). Our concept establishes a "competitive arousal" by forcing the operator to compete with a "calibrated reference operator" that operates under the same contextual conditions (cloned worlds) and is capable of acting according to a calibrated performance (see Figure 23.3), named ComPerf. When the human operator acts, his or her performance metric, for example, the throughput of the system, is measured and fed back for instant comparison. The headstart is the quantified indicator for the performance of the human operator compared to the reference operator. The reference operator in this instance is a model-based software agent that supports the gradations of performance.

If the lead of the human operator shrinks below a given threshold, a hard penalty applies to challenge the test person to compete as hard as possible. In this instance, the effort needed to finish the scenario successfully was increased by generating additional tasks or enlarging time constraints, such as the scenario's finish time. As expressed before, the implementation must carefully compensate for the changed boundary conditions stemming from applied penalties, to achieve constant and comparable contextual conditions. The advantage of controlling the available time t_a by varying the performance of the

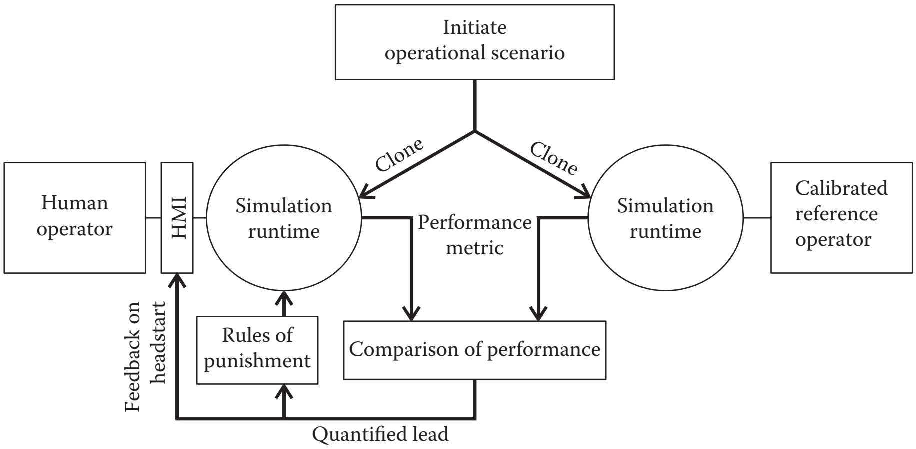

FIGURE 23.3
The concept of ComPerf.

reference operator and, therefore, by establishing the TB principles in relation to the decision times of the human operator t_n, is the inherent feedback loop that is highly suited for automatic tuning of the perceived time pressure.

23.3 Empirical Study

The introduction to the conceptual methodology of risk analysis by means of *AccSis* and the approach regarding time-pressure induction with the help of *ComPerf* were both deduced to an experimental design, in which the plausibility of the risk model, as shown in Figure 23.2, should be the subject of investigation. The controller working place (CWP) of the air traffic controller (ATCo) has been chosen as an exemplary safety-critical working environment within air traffic control. The related task is to control traffic at the airport in the function of a tower controller according to the procedures defined by ICAO PANS-ATM Doc. 4444. The principal tasks of the ATCo are defined in section 7.1.1.1 as follows: *"Aerodrome control towers shall issue information and clearances to aircraft under their control to achieve a safe, orderly and expeditious flow of air traffic on and in the vicinity of an aerodrome with the object of preventing collision(s). …"*

The hypotheses were formulated as follows:

- The time needed t_n is sensitive to the target load set by *ComPerf*
- The relative frequency of safety-relevant events is sensitive to the target load set by *ComPerf*

These hypotheses set the focus on two major causal relationships of the risk model (see Figure 23.2).

We decided to choose the RI as the target safety-relevant event instead of an accident event. In the present context of aerodrome traffic control, the RI is a precursor of an accident event and is as such selected as a risk-indicating event, defined by ICAO Doc. 4444 as the following: *"Any occurrence at an aerodrome involving the incorrect presence of an aircraft vehicle or person on the protected area of a surface designated for the landing and take-off of aircraft."*

The notion of RIs as precursors of accident events is backed by safety management principles and the statistical understanding that the occurrence of collision accidents relates to RIs in a ratio of 1:100, which would, by the way, imply a runway collision accident rate of one every 3.7 years (Birenheide 2010b).

23.3.1 Experimental Tasks and Simulation Scenarios

The chosen HITLS consists of test persons that operate a Surface Manager HMI as the primary working device (Figure 23.4). The device complies with the Eurocontrol A-SMGCS Implementation level 3 (Birenheide 2010a), with the functional exception of a missing device that prevents RI (runway incursion prevention and alerting systems, RIPASs) automatically. The tasks to be performed by the test persons are defined by ICAO Annex 11 and ICAO PANS-ATM Doc. 4444 for tower and ground-control services. The Surface Manager HMI allows for the selection of a target aircraft by pen strokes, as well as granting push-back, taxi, lineup, or take-off clearances on an airport surface surveillance radar screen presenting the entire traffic situation at Frankfurt airport (ICAO code: EDDF).

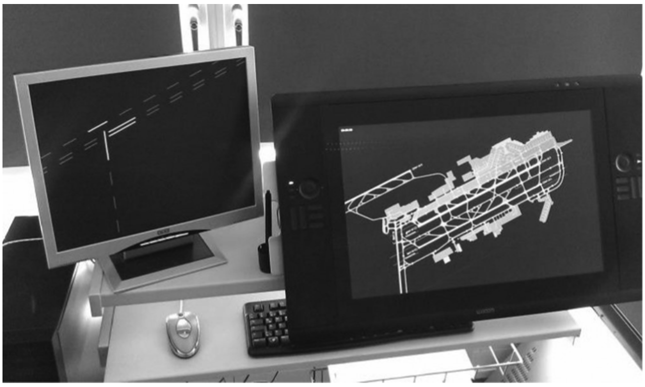

FIGURE 23.4
The Surface Movement Manager HMI consists of a ground surveillance of the airport and a secondary surveillance radar of the vicinity of the airport.

The generated traffic consists of inbound and outbound a/c traffic movements at Frankfurt airport on the three active runways (RWYs) in direction 25, operating 25L as a landing-only RWY, 18 as take-off-only RWY, and 25R in a mixed mode. This complies with the former operational concept before RWY north started operating. RWY dependencies are given for RWY 18 and RWY 25R, as well as for RWY 18 and RWY 25L. The dependency between 25R and 25L was considered according to the reduced RWY separation and semi-mixed parallel RWY operations. The random traffic generator initially distributes 160 movements over 240 simulated minutes per execution run according to a given set of stochastic parameters with uniformly distributed destination routes or departure gates (including north and south area stands) and RWYs. We accelerated the simulation speed by a factor of 2. The routes of the ground movements are initialized by the *Floyd and Warshall algorithm*, which optimizes routes according to a given operational concept and ensures a similar task load for all experimental executions. The software aircraft/pilot agents are capable of self-separating on taxiways and to solve taxi obstruction and crossing conflicts autonomously according to the rules laid out in ICAO Annex 2—Rules of the Air. The execution scenario demands that the test persons work on both ground and tower positions parallelly, that they control the whole airport at a severely increased task load (160 movements at doubled real time).

The concept of *ComPerf* was adapted to the experiment by the application of a simple controller agent, who is capable of acting as an ATCo. The evaluation of the agent's decisions by a traffic-movements predictor affects the resulting operation to be conflict-free to a verifiable degree. No prioritization is implemented, since the agent handles all the movements simultaneously and independently. The agent is configurable by a reaction time t_r per clearance, which calibrates the performance concerning the number of aircraft handled per time. By setting t_n, the decision making of the controller agent allows the human operator a controllable advantage in the context of the performance comparison of *ComPerf*. The human operators' time necessary for decision-making t_n is hence set into competition with t_r, by which the TB principles are established when defining $t_r = t_a$. Setting a desired rapidness, t_r of decision making can consequently be assumed as a target load for the human operator.

The absolute number of traffic movements, which depart from the simulated airport or reach their designated stand, has been chosen to be the key performance metric for

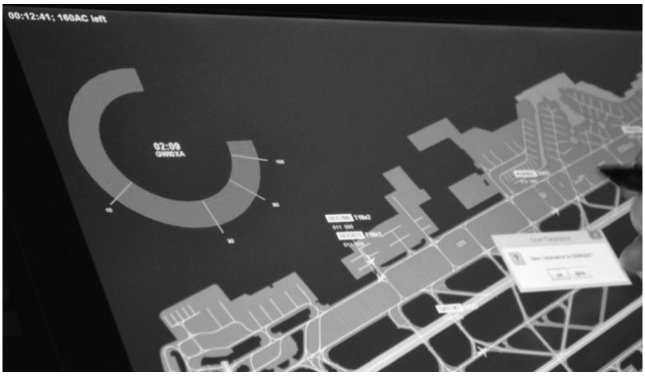

FIGURE 23.5
The clock on the ground surveillance display feedbacks the lead to the human operator.

ComPerf. Leaving the system is defined by the moment of (1) granting the clearance for takeoff for outbound movements or (2) granting the last taxi clearance before entering the aircraft stand. The comparison calculates the performance lead of the human operator by comparing these metrics to the autonomous software competitor. Presuming that the test person would not take any action, a time can be calculated for which the lead becomes zero. This can be regarded as a quantified headstart, calculated on the basis of a fast-time simulation of the controller-agent's world that establishes the complete timeline of the agent, including timestamps of all the operational events, in very little time.

The countdown was visually and acoustically fed back to the human operator by the visualization of a clock on the ground surveillance display (Figure 23.5) and by an alarm noise. The noise indicated the lead time, graded from 300 to 180, 30, and 10 simulated seconds, accompanied by an increasing playback volume. A lead of zero was accompanied by an unpleasant alarm noise, indicating the time-error (TE) condition that results in penalty. The visualization of the headstart consisted of a circle-like clock that covered 6 min as a full circle with a logarithmic time axis.

The penalty was implemented as an increase of the aircraft queue by two additional movements. This consequently increased the duration of the experiment indirectly by the time necessary for handling and finalizing the movements. As the simulated world of the agent is synchronized with the test person's world, the duration of the experiment effectively lies in the test person's hand. This mechanism is regarded as a sufficient measure of motivation for winning the competition, since we presume that all test persons are not only motivated to successfully compete with the controller agent but, moreover, to finish the simulation in time (and be done).

23.3.2 Test Persons and Training

For the empirical study, we acquired three students of the study program "Transport Engineering" in the 4th year of their diploma to act as novice test persons. We educated them according to the tasks described above and trained them by means of the test setup. Every test person successfully completed a training consisting of 10 h and final tests that indicated whether the rules of RWY separation could be mastered according to the trained procedures.

TABLE 23.1

Target Load Parameters

ComPerf A/B	t_r **(s)**
A	30
B	20

23.3.3 Measurements

The measurements consisted of three metrics, namely, the necessary time, the frequency of RI, and the frequency of TEs, which fulfilled our requirements to capture reactions to the gradations of load according to our hypotheses.

Firstly, we recorded RI events as the principal safety metric during the experiment. RIs were automatically detected as soon as rules of the reduced RWY separation minima and parallel RWY operations described in ICAO PANS-ATM Doc. 4444 were violated. Secondly, the necessary time t_n is regarded as an indicator of the cognitive decision time and is the measured time period from the request of clearance by the aircraft until the clearance is granted by the human operator. Third, the frequency of TE was recorded, quantifying the number of penalties applied when the lead was zero and the TB, therefore, equaled >1.

23.3.4 Calibration of Target Load

The calibration procedures were performed prior to the experiment and consisted of a trajectory that varies t_r over time through a predefined bandwidth between 0 and 150 s. The calibration procedure is explained in more detail in Meyer et al. (2014). Two target load levels were quantified as parameters for the controller agent, which defined two experimental configurations (Table 23.1).

23.3.5 Executive Planning

The experiments were conducted according to a sequence plan that varied the configurations and its target load in a systematic order. It is assumed that the quality of the novice person's decisions is continuously increasing due to the experience gained during the series of experiments. To prevent the measurements from being affected by the training effect, the sequence plan follows an alternating order of the configurations.

23.4 Results

23.4.1 The Necessary Time

According to the stated hypothesis for correlation, it was expected that the decreasing reaction times of the controller agent (increasing the target load) would effect an accelerated working speed of the human operator (hence, a decreased time needed t_n). With respect to this expectation, the three test persons showed unclear reactions in the time needed to grant clearances. This is indicated by the measurements ($n > 1000$), illustrated

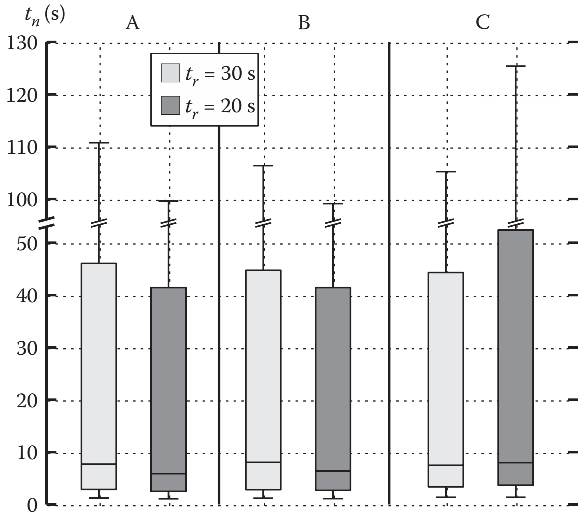

FIGURE 23.6
Reaction times of t_n over the varying target loads of t_r.

TABLE 23.2
U Test of Reaction Times t_n

Test Person	p-Value (%)
A	1.11
B	9.89
C	5.09

in Figure 23.6, which contrasts t_n as box plots according to the selected target loads for each test person. The measurements of test persons A and B indicate the tendencies of an accelerated working speed. In contrast, test person C shows a tendency to maintain his or her working speed.

For testing these observations objectively, the Mann–Whitney–U test provides a probability (p-value) for two independent nonparametric samples on its central tendency to belong to the same population.

The test results (Table 23.2) show no clear rejection of the null hypothesis for all test persons. Only the distribution of test person A exhibits a significant increase in reaction time, indicated by a value of $p < 5\%$. Test person B shows the same tendency. The reaction of test person C is contrary to A and B.

23.4.2 Runway Incursion

RIs were measured as an absolute frequency per target load and test person. The frequency was divided by the number of take-off clearances granted by the human operator. This should compensate for those varying periods of the execution scenarios due to the extensions by the applied penalties. A common tendency can be found through all the measurements (Figure 23.7). This confirms a sensitivity of the target load to the resulting

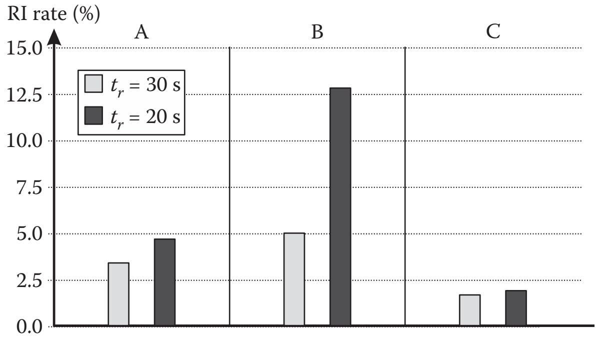

FIGURE 23.7
RI rate.

frequency of safety-relevant events. Therein, test person B shows the largest increase and C the lowest, indicating a decrease in the quality of decision making.

The measured frequency of RI was subject to a learning curve, indicating an increase in competence and therefore also in quality over the course of the experiment.

23.4.3 TE and TB

The absolute frequency of TE indicates the compliance of the working speed with the given target load. For this reason, TE is a metric that shows the ability of the human operator to respond to the induced time pressure. The test persons show a two-track reaction on the increasing load (Figure 23.8). Test persons A and C showed less reactions to the increased target load than test person B, while test person C shows a smaller overall susceptibility to the induction procedure. From the view point of the human operator, permitting a higher frequency of TEs might be an attractive means to effectively extend the available time t_a while accepting that the experimental period is extended by the penalty. Thus, a correlation between the frequency of TE and the mean *TB* (Table 23.3) can be expected.

Dividing the samples of t_n by the reaction time t_r delivers the TB samples whose mean values are summarized in Table 23.3.

The Spearman-correlation rank coefficient was 60% (p-value: 20.8%). Even when no significance can be proven, it provides an indication of a strong relation between *TE* and *TB*.

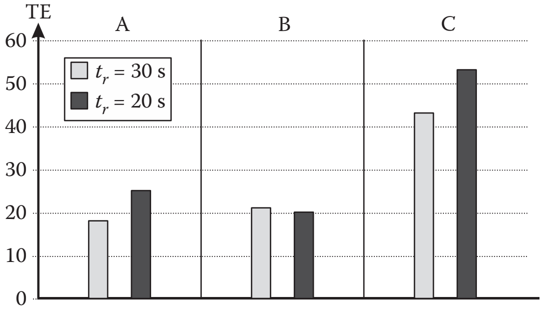

FIGURE 23.8
Frequency of TEs.

TABLE 23.3

Mean TBs and TEs

	A		B		C	
$t_r(s)$	30	20	30	20	30	20
Mean TB	1.30	1.82	1.29	1.76	1.37	2.20
Sum TE	18	25	21	20	43	53

23.5 Conclusion

In summary, the results clearly show the reactions of the test persons to increased stress, as well as a lowered quality of work for all test persons, at the same time. However, the data gathered from this small test group are not sufficient to validate the risk model (Figure 23.2), since the number of samples does not provide the required power for statistical testing. Nevertheless, the findings serve satisfactorily for giving clues on the success of the induction procedure and the plausibility of the hypothesis in the scope of the proof-of-concept study.

The tendency observed indicates an increase of the probability of safety-relevant events and human error when increasing the target load (Figure 23.7). Thus, an increased uncertainty during decision making can be concluded from the data gathered from the three test persons. The high variance of the amplitude of the *RI rate* may be explained by considering the frequency of *TE* as a crucial influencing factor of uncertainty. This consideration manifests in all measurable actions of the human operator, forcing him or her to balance between the quality and working speed in a subjective speed–accuracy trade-off (SAT) that is in line with the principles described by *Fitts' law* (Fitts 1954). This trade-off is illustrated as an example in Figure 23.8, which shows the individual operating points on the RI-rate–TE chart and the balance of the available performance between the two claims of the task definition defined in section 7.1.1.1, ICAO PANS-ATM Doc. 4444.

A plausible analytical description of the trade-off can be derived from the relation between quality, performance, and time needed, introduced in Section 23.2.2. Assuming the *RI rate* as a reciprocal metric of quality and the frequency of *TE* as a valid measure of the TB (cf. correlation test), the relationship can be expressed as

$$\frac{t_n}{Q} = \frac{1}{P} \Leftrightarrow RI_{rate} \cdot TB \cdot t_r = \frac{1}{P} \Leftrightarrow \boxed{RI_{rate} \cdot TE = \frac{1}{P \cdot t_r}}. \tag{23.4}$$

This term describes the product of t_r and P as the reciprocal function of a surface, with the measured factors forming the dimensions of the related rectangle. With a view on Figure 23.8, the shift of the operating points of the test persons A and C approximately follows this relationship, leading to a measure of performance that takes the target load t_r into account, while maintaining constant proportions of the trade-off. Figure 23.9 illustrates this relation.

Test person B obviously switched the strategic and tactical priorities, which led to a shift in the trade-off's proportions in favor of the *RI rate*. The cause of this rebalancing might be explained by using Hollnagel's *Contextual Control Model* (Hollnagel 1993). In this model, test person B's behavior could refer to a *scrambled control* mode, indicating an overload situation or insufficient training of the actions that do not follow a stable action pattern.

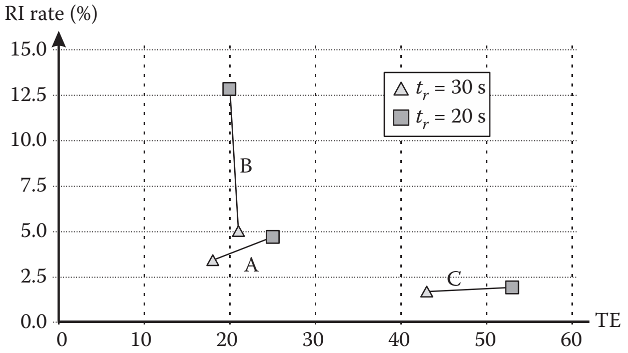

FIGURE 23.9
RI-rate–TE chart of stress reactions and the trade-off.

In contrast, the test persons A and C show stress reactions that indicate the systematic relations between the *RI rate* and *TE*. This relation might explain the effective acceleration of safety-relevant events that occur under the present time constraints.

The lowest degree of sensitivity most clearly show the reactions of test person C. Being insensitive to time-pressure arousal is an essential mark of quality air traffic control. Permitting *TE* fully complies with the training, by which the test person is qualified to act correspondingly to their role model. An alternative conclusion can be deduced from the reactions of test persons A and C, whose proportions of the trade-off remained stable, while time pressure increased. This can be best observed in the time needed by test person C (Figure 23.7), who showed the robustness against stress induction by maintaining the working speed in the best way. This observation complies with the findings of Rastegary (Rastegary and Landy 1993), by which test person C can be classified as a time-urgent individual, whose sensitivity to time pressure provides for a constant performance over a larger interval.

As a conclusion, it can be said that the risk of a sociotechnical system under the conditions of design stress might best be represented by an individual operating point on the RI-rate–TE plane. The estimation of this point can then be achieved by using stress reactions of varying accelerated-stress conditions for the regression analysis. Since, currently, there is no valid regression model known that can analytically relate to time pressure, *TE*, and the safety metrics, the results of the stress reaction, illustrated in Figure 23.8, can give clear indications of the qualities needed for the model, which are

- Basis performance: any individual provides a specific basis performance for design-stress conditions, consisting of specific proportions of the *RI rate* and *TE*.
- Variance of the proportions: as the reactions of test person B show, the proportions of the trade-off depend on the present time pressure. The Contextual Control Model might provide qualitative indications of the curve progression of the proportions of the trade-off over an increased time pressure.

By way of modeling the relations analytically for use in a regression analysis, the findings on the field of *speed-versus-accuracy* studies might be helpful. Furthermore, enhancing the related SAT curves by using ATM-related safety metrics instead of human error events might lead to a successful adaptation of the models known to use in the current context.

References

Birenheide, M. 2010a. Definition of A-SMGCS Implementation Levels. Article. EUROCONTROL Headquarters. https://www.eurocontrol.int/articles/advanced-surface-movement-guidance-and-control-systems-smgcs (accessed August 7, 2015).

Birenheide, M. 2010b. Generic cost–benefit analysis of A-SMGCS levels 1 and 2. Article. EUROCONTROL Headquarters. https://www.eurocontrol.int/articles/advanced-surface-movement-guidance-and-control-systems-smgcs (accessed August 7, 2015).

Blom, H. A. P., G. J. Bakker, P. J. G. Blanker et al. 2001. Accident risk assessment for advanced air traffic management. *Progress in Astronautics and Aeronautics* 193: 463–80.

Blom, H. A. P., G. J. Bakker, M. H. C Everdij et al. 2003. Stochastic analysis background of accident risk assessment for air traffic management. Hybridge Report D2.2. National Aerospace Laboratory NLR. http://hybridge.nlr.nl (accessed August 7, 2015).

Bubb, H. 2005. Human reliability: A key to improved quality in manufacturing. *Human Factors and Ergonomics in Manufacturing and Service Industries* 15.4: 353–68.

Bubb, H., and I. Jastrzebska-Fraczek. 1999. Human error probability depending on time pressure and difficulty of sequential tasks. *Safety and Reliability—Proceedings of ESREL 99,* 1: 681–86.

Chang, Y. H. J., and A. Mosleh. 2007. Cognitive modeling and dynamic probabilistic simulation of operating crew response to complex system accidents. Part 2: IDAC performance influencing factors model. *Reliability Engineering and System Safety* 92.8: 1014–40.

Cox, D. R. 1972. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)* 34.2: 187–220.

Edland, A., and O. Svenson. 1993. Judgment and decision making under time pressure—Studies and findings. In *Time Pressure and Stress in Human Judgment and Decision Making*, eds. O. Svenson, and A. J. Maule, 27–40, New York and London: Plenum Press.

Fitts, P. M. 1954. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology* 47.6: 262–69.

Freedman, J. L., D. R. Edwards, and J. E. McGrath. 1988. Time pressure, task performance, and enjoyment. *The Social Psychology of Time: New Perspectives* 91: 113–33.

Hollnagel, E. 1993. *Human Reliability Analysis: Context and Control*. London: Academic Press.

Johnson, C. 1999. Why human error modeling has failed to help systems development. *Interacting with Computers* 11.5: 517–524.

Kerstholt, J. H. 1994. The effect of time pressure on decision-making behaviour in a dynamic task environment. *Acta Psychologica* 86.1: 89–104.

Malhotra, D. 2010. The desire to win: The effects of competitive arousal on motivation and behavior. *Organizational Behavior and Human Decision Processes* 111.2: 139–46.

Meyer, L., K. Gaunitz, and H. Fricke. 2014. Investigating time pressure for the empirical risk analysis of socio-technical systems in ATM. *Advances in Human Aspects of Transportation: Part I* 1: 33–45.

Müller, M. 2004. Risk and risk management in aviation. *Zeitschrift für ärztliche Fortbildung und Qualitätssicherung* 98.7: 559–65.

Nelson, B. N. 2009. *Accelerated Testing: Statistical Method, Test Plans, and Data Analysis*. New Jersey: John Wiley & Sons.

Rastegary, H., and F. J. Landy. 1993. The interactions among time urgency, uncertainty, and time pressure. In *Time Pressure and Stress in Human Judgment and Decision Making*, eds. O. Svenson, and A. J. Maule, 217–39, New York and London: Plenum Press.

Reason, J. 1990. *Human Error*. New York, USA: Cambridge University Press.

Shapell, S., and D. Wiegmann. 1996. U.S. naval aviation mishaps 1977–92: Differences between single- and dual-piloted aircraft. *Aviation, Space, and Environmental Medicine* 67(1): 65–9.

Shorrock, S. T., B. Kirwan, H. MacKendrick, and R. Kennedy. 2001. Assessing human error in air traffic management systems design: Methodological issues. *Le travail humain* 64.3: 269–89.

Spouge, J., and E. Perrin. 2006. 2005/2012 integrated risk picture for air traffic management in Europe. Eurocontrol EEC Note 2006/05, https://www.eurocontrol.int/eec/public/standard_page/DOC_Report_2006_009.html (accessed August 7, 2015).

Stroeve, S. H., H. A. P. Blom, and G. J. Bakker. 2013. Contrasting safety assessments of a runway incursion scenario: Event sequence analysis versus multi-agent dynamic risk modelling. *Reliability Engineering and System Safety* 109: 133–49.

Swain, A. D. 1990. Human reliability analysis: Need, status, trends and limitations. *Reliability Engineering and System Safety* 29.3: 301–13.

Swain, A. D., and H. E. Guttman. 1983. *Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications*. NUREG/CR-1278. Albuquerque: Sandia National Laboratories.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com