Toward a barrier design method based on human error success

Frédéric Vanderhaegen

Keywords: human error analysis; ACIH method; design of barrier systems; BCD model for system design; error-tolerant barrier; transportation and industrial systems

Summary

This paper discusses on human error concepts and presents barriers as supports to control human error but also as causes of human error occurrence. Because of the limits of present human error analysis methods, the ACIH approach is proposed in order to focus on the assessment of the consequences instead of the probability of human errors. Consequences are assessed in terms of benefits, costs and potential deficits or dangers with the so-called BCD model. Examples of application of the ACIH method are presented for particular intentional human errors called barrier removals, i.e. the human operators do not respect barriers integrated on field. Future application of ACIH will focus on the design and the analysis of human-machine systems integrating error-tolerant barriers, i.e. integrating barriers that can tolerate their removals by human operators.

Concepts on human errors and barrier system

A definition of human reliability may be associated to the technical reliability, i.e. it is the capacity of the human operators to achieve correctly their allocated functions, in given conditions and on a given interval of time. Nevertheless, different adaptations of such a definition occur on literature. Some authors prefer defining human reliability as the capacity of human operators to achieve correctly their allocated tasks instead of speaking on functions. The function concept relates then to the system mission whereas the task one relates to the human factor contribution to achieve the mission. This task may be more or less detailed. It can be only an objective, i.e. to supervise a process, or a very detailed procedure for which the task leads to the definition of a prescribed behaviour. For these reasons, the human error can be identified by the comparison between:

- The prescribed behaviour with the observed one, or
- The expected performance or result with the obtained one.

The human reliability concept can then be confused. It sometimes excludes the recovery of human erroneous behaviours whatever the result of this recovery process. It can also be assimilated to the technical availability, i.e. it can be the capacity of human operators to be ready to achieve their allocated tasks, in given conditions and at a given time. Moreover, related to the technical maintainability, the human reliability can be associated to the capacity of the human operator to recover their own erroneous tasks or to maintain their own knowledge. Those human characteristics cannot be applied to technical components for which the definitions of reliability, availability, maintainability or safety (i.e. RAMS concept) do not consider the possible evolution of the knowledge of these components and the possibility for these components not to respect voluntarily any prescriptions! Human operators, on the other hand, are able to decide to modify a given prescribed tasks, to create new tasks or not to achieve tasks. Therefore, a new definition of the human reliability is required. The proposed definition is adapted from Swain and Guttmann (1983). The human reliability is the capacity of human operators:

- To achieve correctly their prescribed tasks, in given conditions, during an interval of time or at a given time,

- Not to achieve any additional tasks that may damage the human-machine system, this damage may be associated to many criteria such as safety, quality, production, workload, etc.

The human error concept is the complementary of the reliability one. Therefore, it relates to the capacity of human operators:

- Not to realize correctly their allocated tasks in given conditions during a period of time or at a given time, or
- To realize additional tasks that may affect the human-machine system functioning in terms of safety, quality, production, workload, etc.

Particular supports can be designed in order to control undesirable events such as human errors that may affect criteria such safety, workload, production or quality. They are called barriers that are systems that protect the human-machine system from the occurrence or the consequences of undesirable events. A barrier is characterized by its source, i.e. who is the designer of the barrier and its target, i.e. who will use this barrier. Three levels of barrier design process can be defined. The designers of a given machine equip it with barriers with respect to the norms or risk analysis results. The employer who installs and operates this machine on an industrial site defines other barriers on field with respect to the implantation environment. Finally, the human operators who use this machine may modify some existing barriers or create new ones. The choice of the barriers of the first level (i.e. the designers of a machine) is the result of a risk analysis process. At the second level (i.e. the society that installs and operates the machine), additional barriers are linked to the operational conformity to be followed. However, the possible behaviours of human operators on field who face technical barriers or who are considered as barriers are not formally assessed (Hammerl and Vanderhaegen, 2009).

Intentional deviation from the prescribed behaviour required by the system specifications is called a violation. A so-called barrier removal was initially defined as the voluntary barrier inhibition with the intention to optimize the possible compromises between criteria such as safety, workload, production, quality, etc (Polet et al., 2003). Thus, a barrier removal, or an intentional misuse or non-respect of a barrier under appropriate conditions, is an optimizing or exceptional violation made without any intention to damage the human-machine system. Therefore, barriers may be designed for controlling human errors but barriers can also be the cause of the occurrence of particular human errors such as barrier removals. Two groups of barriers might then be considered (Polet et al., 2002):

- Barrier affecting the system integrity if they are removed. There are material barriers: if they are removed or not respected, the system is physically modified.
- Barrier unrelated to the system integrity if they are removed. There are immaterial barriers: if they are removed or not respected, the system is not modified physically.

Many methods can be used in order to analyze or assess human errors. They facilitate the definition of the barriers of the designers, the employers and/or the users. Among these methods, quantitative and cognitive based methods can be applied for off-line prospective or retrospective human error analysis and can support the design of the barriers to control on-line human errors. Nevertheless, they present operational limits:

- The results they give are not homogeneous. Studies have shown that a given method used by several groups or different methods used by a same group do not produce reliable results.
- The human behavioural model they used is sometimes difficult to apply.

- The human error based risk analysis process made by the designers, the users or the employers of a given human-machine systems can diverge because their different objectives or different organisational or individual interests they take into account.
- The analysis of the tasks does not integrate all the dependencies between tasks (e.g., temporal dependencies, causal dependencies, functional dependencies).
- They focus mainly on the analysis of non-intentional errors without taking into account intentional errors such as violations, barrier removals or additional tasks.
- They are off-line processes without taking into account the on-line risk control process requirements made by the human operators on field.
- The feedback to assess the human error probability is often insufficient.
- Even if probabilities on human errors are available, they usually cannot be compared because they do not have homogeneous assessment units.
- The process of human error analysis is much more a retrospective one than a prospective one. Instead of focusing on the incident or accident prevention, the investigation effort related to human errors is done after the occurrence of a danger due to human factors.
- The databases on accidents or incidents focus mainly on the negative contributions of the human operators reporting their errors without taking into account the possible positive ones, i.e. the contribution of human operators to avoid or recover an incident or an accident.
- Human error assessment is done without taking into account the dynamic evolution of the human-machine system, e.g. the learning effect. These methods do not consider that the system users who have to analyze the risks they are facing to and who have to control them on-line can learn from their own errors and behaviours.

For these reasons, Vanderhaegen (1999) recommends to relate human reliability analysis to the consequences of human behaviour rather to the probability occurrence in order to analyse both intentional and unintentional human errors. The explanation of the human error occurrence relates to the possible consequences of erroneous behaviours in terms of success and failure, i.e. in terms of benefits, costs and potential deficits or dangers. Human errors such as barrier removals can then be assessed and this analysis will aim at designing error-tolerant system of barriers.

The ACIH approach to design error-tolerant system of barriers

The ACIH (French acronym for Analysis of Consequences of Human Unreliability) method has several steps, Figure 1 (Vanderhaegen, 2001). Regarding the system objective and constraints, the analysis related to the system structure and functions is done and then the analysis of the human tasks involved in the achievement of the system functions aims at identifying the operational contexts and the required procedures. With these human tasks, possible human errors are identified, and their consequences are assessed with the so-called BCD model. Prescribed behaviours are compared with the anticipated possible human errors in a prospective analysis process or with the observed human errors in a retrospective analysis process. The ACIH method can be used to analyze the efficiency of the barrier system.

The BCD model is based on indicators that assess the positive and the negative consequences of intentional or unintentional deviated human behaviours on several criteria related to technical or human performance or state. Positive ones are benefits whereas negative ones are acceptable costs when the undesirable events are under control or unacceptable deficits when they are over control. A cost is then an acceptable negative consequence when the human behaviour is successful and a deficit is an unacceptable consequence when this behaviour fails and damages the human-machine system regarding the safety or other criteria.

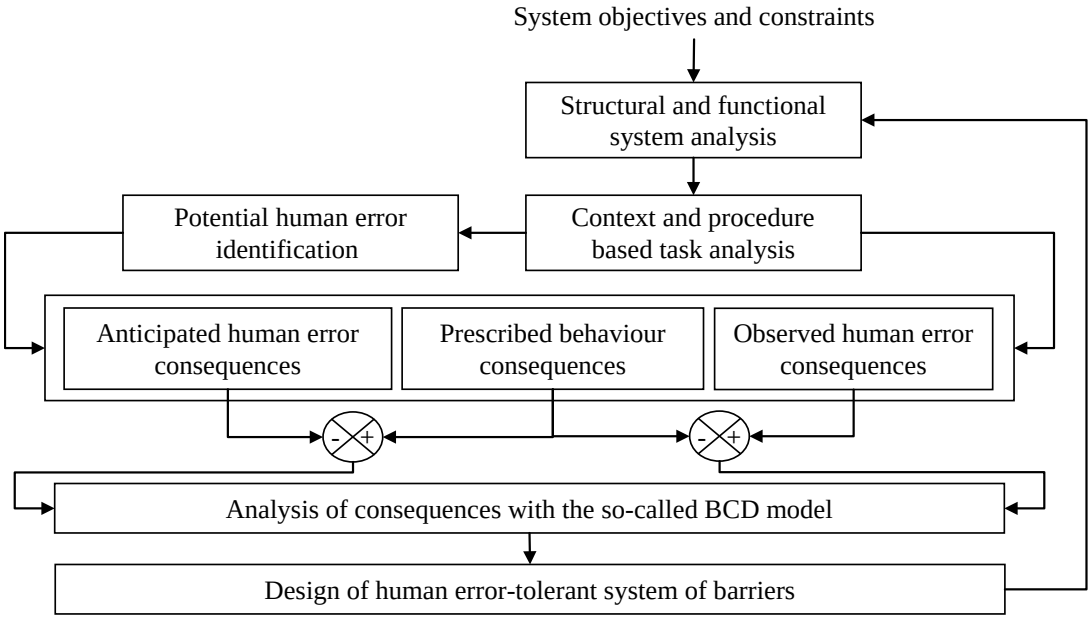

Fig. 1: The ACIH method steps

Therefore, whatever the deviated human behavior states (e.g., normal or degraded behaviours, intentional or unintentional deviations), the corresponding human action occurrence is supposed to be valuable by three distinct consequences on several evaluation criteria (Polet et al, 2002):

- The expected benefits (i.e., the B values of the BCD model) due to the success of the performed action.
- The acceptable costs (i.e., the C values of the BCD model) due to the success of the performed action. It can relate to a cognitive cost to control the potential deficit or danger or to a physical one to modify the operational constraints of the use of a given system such a barrier.
- The unacceptable possible deficit (i.e., the D values of the BCD model) related to the potential occurrence of a hazardous situation, in case of an unsuccessful action.

A human behaviour can be explained in terms of benefits and costs when it is a success or in terms of deficits when it fails (Vanderhaegen, 2004). For an on-line intentional human behaviour under control, the benefits and costs are considered as quasi-immediate whereas the deficits are potential. Indicators are then required to compare dependent or independent situations. Two situations are dependent when the state of a situation occurring at a given time is modified and leads to another short or long term situation. This modification can be due to the dynamic evolution of the process or to a strategic or tactical action. Two situations are independent when they can occur at the same time but concern two different paths to achieve same goals. Independent situations can then relate to the possible action plans of different decisional levels of a given organization to solve a current situation. Whatever the hierarchical level, the BCD model is able to transform qualitative or subjective data into quantitative or objective ones using several functions. The analysis of the consequences of a deviated behaviour requires a reference in order to determine if this deviation lead to an improvement or not. The usual reference is the prescribed behaviour. The use of the BCD model aims then to compare globally different action plans or procedures, i.e., to compare different series of successive situations, Figure 2. A procedure or an action plan is a combination, usually a sequence, of actions and the achievement of an action generates the occurrence of a new situation.

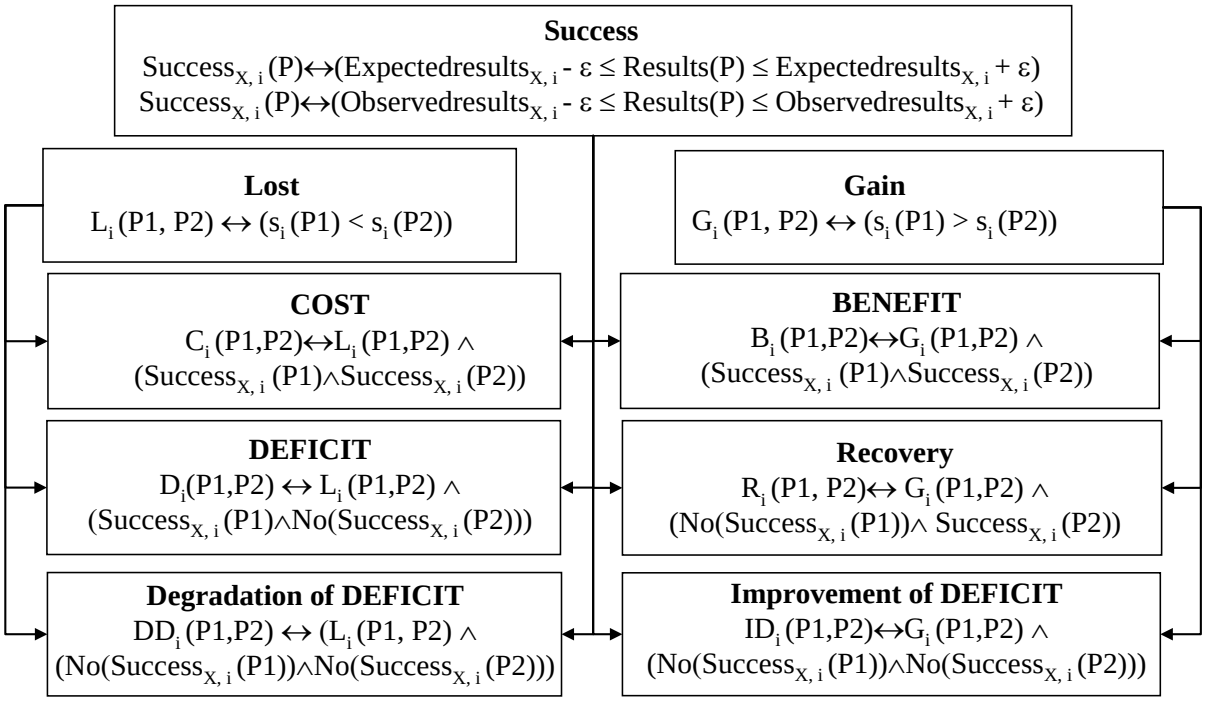

Fig. 2: Assessment of the BCD model parameters between plans or procedures.

The estimation of the benefits, the costs and the deficits can be assessed by comparison between the prescribed procedure or action plan with the anticipated or observed deviated procedure or action plan. The existing barriers on field relate usually to an explicit or an implicit prescribed action plan or procedure: the human operators achieve an action plan or a procedure that respects these barriers. Deviated procedures or action plans may relate to barrier removals, i.e., the human operators achieve an action plan or a procedure that removes the barriers they are supposed to respect. All the possible action plans or procedures can then be compared on the same interval of time and the assessment of the BCD parameters of two alternatives *P1* and *P2* is defined on Figure 2. The severity related to the criterion *i* of both plans or procedures are noted $s_i(P1)$ and noted $s_i(P2)$ respectively. The BCD assessment requires the success function application. The success function noted $Success_{X,i}(P)$ relates to the results of *P* on a given criterion *i* for a decision-maker *X* regarding the expected or observed ones noted $Expectedresults_{X,i}$ or $Observedresults_{X,i}$ respectively, with an acceptable error ε. Several parameters are assessed: the cost, the benefit, the deficit, the recovery, the improvement or the degradation of the deficit.

The ACIH approach was applied to analyse barrier removals in different domains by using the BCD model parameters, i.e. the benefits, the costs, the potential deficits or dangers integrating the improvement and the degradation of the deficits or dangers:

- Retrospective analysis of barrier removals during the use of a production system such as an industrial rotary press (Polet et al., 2003).
- Retrospective analysis of barrier removals during the control of transport system such as railway system (Vanderhaegen, 2009; Zhang et al., 2004).
- Prospective analysis to predict barrier removals in road domain (Chaali-Djelassi and Vanderhaegen, 2006) or in railway domain (Vanderhaegen et al., 2009)
- Prospective analysis with the assessment of the probability of success of railway procedure removals in order to attenuate the BCD model parameter assessment (Chaali-Djelassi et al., 2007).

Conclusion

Both the ACIH method and the BCD model were presented. The ACIH method aims at assessing the consequence of human erroneous behaviours such as intentional barrier removals and these behaviours are analysed in terms of success and failure, i.e. of benefits, costs and potential deficits or dangers with the BCD model. Several studies have shown that the ACIH method and the BCD model are applicable. Future studies will use the ACIH method to design error-tolerant barrier system.

Acknowledgements

The present research work has been supported by the International Campus on Safety and Intermodality in Transportation (CISIT), the European Community, the Délégation Régionale à la Recherche et à la Technologie, the Ministère de l'Enseignement Supérieur et de la Recherche, the Région Nord Pas de Calais and the Centre National de la Recherche Scientifique. The authors gratefully acknowledge the support of these institutions.

References

Chaali-Djelassi, A. and Vanderhaegen, F. (2006). *Predication of violations in road transportation system*. Proceedings of the 4th International Conference on Safety and Reliability, Krakow, May 30- June 02 2006, pp. 57-68. ISSN 1895-8281.

Chaali-Djelassi, A., Vanderhaegen, F., Cacciabue, P.-C., Cassani, M. (2007). *Barrier removal prediction based on a new approach. Application to a degraded train speed procedure*. Paper presented at the 26th European Annual Conference on Human Decision Making and Manual Control, 21-22 June 2007, Copenhagen, Denmark.

Swain, A. D., and H. E. Guttmann (1983). *Handbook of Reliability Analysis with emphasis on Nuclear Plant Applications*. NUclear REGulatory Commission, NUREG/CR-1278, Washington D.C.

Hammerl, M., Vanderhaegen, F. (2009). *Human factors in the railway system safety analysis*. Rail Human Factor Conference, Lille, France, 3-5 March 2009.

Polet, P., Vanderhaegen, F., Amalberti, R. (2003). Modelling Border-line tolerated conditions of use (BTCUs) and associated risks. *Safety Science*, 41, 111-136.

Polet, P., Vanderhaegen, F., Wieringa, P. A. (2002). Theory of safety-related violations of system barriers. *Cognition, Technology & Work*, 4, 171-179.

Vanderhaegen, F. (1999). APRECIH : a human unreliability analysis method – application to railway system. *Control Engineering Practice*, 7, 1395-1403.

Vanderhaegen, F. (2001). A non-probabilistic prospective and retrospective human reliability analysis method – application to railway system. Reliability Engineering and System Safety, 71, 1-13.

Vanderhaegen, F. (2004). *The Benefit-Cost-Deficit (BCD) model for human error analysis and control*. Proceedings of the 9th IFAC/IFORS/IEA symposium on Analysis, Design, and Evaluation of Human-Machine Systems, Atlanta, GA, USA, 7-9 September 2004.

Vanderhaegen, F. (2009). *Rail simulation to study human reliability*. Rail Human Factor Conference, Lille, France, 3-5 March 2009.

Vanderhaegen, F., Polet, P., Ziéba, S. (2009). A reinforced iterative formalism to learn from human errors and uncertainty. *Engineering Applications of Artificial Intelligence*, 22, 4-5, 654-659.

Zhang, Z., Polet, P., Vanderhaegen, F. and Millot, P. (2004). Artificial neural network for violation analysis. *Reliability Engineering and System Safety*, 84, 1, 3-18.