A Benefit/Cost/Deficit (BCD) model for learning from human errors

Frédéric Vanderhaegen, Stéphane Zieba, Simon Enjalbert, Philippe Polet

A Benefit/Cost/Deficit (BCD) model for learning from human errors

Frédéric Vanderhaegen[a,b,c], Stéphane Zieba[d], Simon Enjalbert[*,a,b,c], Philippe Polet[a,b,c]

[a] *Univ Lille Nord de France, F-59000 Lille, France*
[b] *UVHC, LAMIH, F-59313 Valenciennes, France*
[c] *CNRS, FRE 3304, F-59313 Valenciennes, France*
[d] *Laboratory for Cognitive Systems Science, Department of Risk Engineering, University of Tsukuba, Japan*

Abstract

This paper proposes an original model for interpreting human errors, mainly violations, in terms of Benefits, Costs and potential Deficits. This BCD model is then used as an input framework to learn from human errors, and two systems based on this model are developed: a case-based reasoning system and an artificial neural network system. These systems are used to predict a specific human car driving violation: not respecting the priority-to-the-right rule, which is a decision to remove a barrier. Both prediction systems learn from previous violation occurrences, using the BCD model and four criteria: safety, for identifying the Deficit or the danger; and opportunity for action, driver comfort, and time spent, for identifying the Benefits or the Costs. The application of learning systems to predict car driving violations gives a rate over 80% of correct prediction after 10 iterations. These results are validated for the non-respect of priority-to-the-right rule.

Key words: BCD model, violation, human error, neural network, case-based reasoning, learning process, human error prediction, car driving.

[*] Corresponding author. Tel.: +33 32751 1356; fax: +33 32751 1316.
Email address: `simon.enjalbert@univ-valenciennes.fr` (Philippe Polet)

1. Introduction

Adapted from Swain and Guttmann [1], the definition of human error must be linked to the capacity of the human operators to:

- not correctly perform their assigned tasks under given conditions during a time-window or at a given instant, or
- perform additional tasks that may affect the human-machine system's operations in terms of safety, quality, production and/or workload, for example.

This definition implies that the concept of human errors is related to unintentional errors and intentional errors (called violations), as well as additional tasks.
Methods for assessing unintentional or intentional human errors do exist, but the results they provide 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 [2]. The feedback needed to assess or analyze human errors is sometimes insufficient. Even if quantitative or qualitative assessments of human errors are available, they cannot be compared because they usually do not have homogeneous assessment unit [3].
In addition, risk analysis based on human error, done by designers and/or users of a given human-machine system, can provide diverse results because the people analyzing the risks have different objectives or different organisational and/or individual interests [4]. Furthermore, the task analysis may not include all the dependencies between tasks (*e.g.*, temporal dependencies, causal dependencies, functional dependencies) [5] [6]. Moreover, human error analysis methods focus mainly on the negative impact of human errors, and they usually do not take additional tasks into account and do not integrate the learning effect of human errors.
In this paper, we present a model that remedies the above problems. This model:

- is capable of analyzing unintentional human errors, violations and additional tasks,
- integrates both the negative and positive impacts of human errors, and
- functions as a framework that takes learning from human errors into account.

Our goal is to use this model to develop systems that are able to copy human behaviour and thus learn from human errors. Human errors can be explained as the possible consequences of erroneous behaviours with respect to the Benefits, Costs and potential Deficits (*i.e.*, dangers). Such a decision support system design has to use human models in order to make human error prediction feasible.
This paper first introduces our BCD model and then proposes several systems that are able to predict human violations, called barrier removals because the human operators voluntarily decide to remove (*i.e.*, not to respect) the barriers. These systems are applied to predict barrier removal in the context of car driving. This paper does not focus on human error probability but rather on the consequences of human errors.

2. The BCD model principles

2.1. The BCD model

Each person that uses risk assessment methods (*i.e.*, the designers and/or users of a given human-machine system) can define particular systems called barriers, which protect the human-machine system from undesirable events, or their consequences. These undesirable events may affect system criteria, such as safety, workload, production and/or quality. Barriers can also protect the system from undesirable events, such as human errors, or prevent them entirely. Transport signalling systems are prevention-based barriers, whereas protection grids, airbags and safety belts are protection-based barriers. Intentional deviations from the prescribed behaviour required by the system specifications are called violations [7] [8].
Barrier removal was initially defined as the voluntary inhibition of a barrier with the intention of optimizing the possible compromises between such criteria as safety, workload, production or quality, for example [9]. Thus, barrier removal, or the intentional misuse or non-respect of a barrier under appropriate conditions, is an optimizing and/or exceptional violation made without any intention to damage the human-machine system. The possible motivation for a human operator to deviate intentionally from a given prescription can be related to improving the behaviour of the human-machine system or lessening the harmful consequences of an error. This improvement can be assessed quantitatively or qualitatively with respect to several criteria.

The Benefit/Cost/Deficit (BCD) model uses indicators that assess the positive and negative consequences of deviate human behaviours in terms of criteria measuring the machine's or operator's performance or state. Positive consequences are termed Benefits, whereas negative ones are either acceptable Costs, when the undesirable events are under control, or unacceptable Deficits, when these events are out of control. In other words, a Cost is an acceptable negative consequence when human behaviour is successful and a Deficit is an unacceptable consequence when this behaviour fails and damages the human-machine system in terms of safety or some other criteria.
Whatever the deviate human behaviour state (*e.g.*, normal or degraded behaviour, intentional or unintentional deviations), the corresponding human action is assumed to be evaluated through three distinct consequences [4]:

- the expected Benefits (*i.e.*, the B values of the BCD model) due to the success of a performed action;
- the acceptable Costs (*i.e.*, the C values of the BCD model) due to the success of a performed action; these costs can be cognitive (*e.g.*, to control the potential Deficit or danger) or physical (*e.g.*, to modify the operational constraints of using a given barrier);
- the unacceptable Deficits (*i.e.*, the D values of the BCD model) related to a potential occurrence of a hazardous situation, due to the failure of a performed action.

Human behaviour can thus be explained in terms of Benefits and Costs when the behaviour is successful or in terms of Deficits when it fails. For an on-line intentional human behaviour that is under control, the Benefits and Costs are considered to be quasi-immediate, while the Deficits exist in the possible future.
The BCD model uses several functions to transform qualitative (*i.e.*, subjective) data into quantitative (*i.e.*, objective) data. This model has been validated to analyze barrier removals in different domains:

- Barrier removals in production systems, such as industrial rotary presses [9].
- Barrier removals in transport system control, such as car driving [5] or railway system control [10] [11] [12].

- Barrier removals in biomechanical systems, such as human crash behaviour [13].

The BCD model has also been used to assess human stability and the associated risks in terms of human state indicators [14]. The analysis of the consequences of a nonstandard behaviour requires a frame of reference in order to determine whether or not this deviation leads to an improvement. The usual frame of reference is the prescribed task, which is more or less detailed.

Two cases should be distinguished in estimating Benefits, Costs and Deficits: using the BCD model to compare two distinct dependent or independent situations (i.e., system state at a given time), or using the BCD model to compare different action plans, which are successive situations. The next two sub-sections give more details about comparing situations or action plans.

2.2. The BCD model for comparing pairs of situations

Indicators are required to compare dependent or independent situations [15]. Two situations are *dependent* when the state of a situation occurring at a given time is modified and leads to another short-term or long-term situation. This modification may 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 the same goals. Independent situations can thus be related to the possible action plans of the different decisional levels of a given organization in order to resolve the current situation.

The lower a given criterion's severity, the more acceptable the situation. The acceptability function, denoted $TOL_{X,i}(a)$ in Equation (1) is related to the state of the severity, denoted $s_i(a(t_a))$ for the evaluation criterion i of the situation a occurring at the time t_a according to the acceptability threshold $TH_{X,i}$ for a decision maker X:

$$TOL_{X,i}(a) \leftrightarrow (s_i(a(t_a)) < TH_{X,i}) \tag{1}$$

This acceptability threshold is used to compare different users' assessments of a situation or the assessments of several decisional levels of an organization [15]. For instance, a given situation can be acceptable for a user (or decisional level) but unacceptable for another user (or decisional level). The comparison of two dependent situation concerns two situations happening one after the other (*e.g.*, an action has been performed between the two situations)

while the comparison of two independent situations concerns two situations happening at the same time (*e.g.*, two situations from two different action plans).
A decrease in the severity of situation b with respect to the severity of situation a in terms of criterion i is denoted $G_i(a,b)$ in Equation (2):

$$G_i(a,b) \leftrightarrow (s_i(a(t_a)) > s_i(b(t_b))) \tag{2}$$

The situations a and b can be acceptable or unacceptable. When both are acceptable, the corresponding decrease is called a Benefit, denoted $B_i(a,b)$ in Equation (3):

$$B_i(a,b) \leftrightarrow G_i(a,b) \wedge (TOL_{X,i}(a) \wedge TOL_{X,i}(b)) \leftrightarrow G_i(a,b) \wedge TOL_{X,i}(a) \tag{3}$$

For instance, this decrease can be related to a decrease in the severity of situation b, in terms of criterion i, corresponding to a barrier removal, with respect to the severity of situation a.
An increase in the severity of situation b with respect to the severity of situation a is denoted $L_i(a,b)$ in Equation (4):

$$L_i(a,b) \leftrightarrow (s_i(a(t_a)) < s_i(b(t_b))) \tag{4}$$

Situation b can thus be acceptable or unacceptable. An acceptable increase is called a Cost, denoted $C_i(a,b)$ in Equation (5), and requires an additional acceptability constraint:

$$C_i(a,b) \leftrightarrow L_i(a,b) \wedge (TOL_{X,i}(a) \wedge TOL_{X,i}(b)) \leftrightarrow L_i(a,b) \wedge TOL_{X,i}(b) \tag{5}$$

The definition of an acceptable Cost implies that both situations are acceptable.
An unacceptable level of severity for both situations is called a Deficit, denoted $D_i(a,b)$ in Equation (6). This Deficit $D_i(a,b)$, is related to the occurrence of an unacceptable situation b with respect to an acceptable situation a:

$$D_i(a,b) \leftrightarrow L_i(a,b) \wedge (TOL_{X,i}(a) \wedge \neg TOL_{X,i}(b)) \leftrightarrow TOL_{X,i}(a) \wedge \neg TOL_{X,i}(b) \tag{6}$$

A Deficit is called *potential* when it anticipates the possible unacceptable evolution of a situation. Unacceptable situations can also be related to other specific evolutions of severity (*i.e.*, a degradation or an improvement of a

Figure 1: Evaluation of the BCD model parameters for two situations.

Figure 2: Illustration of the BCD model parameters for two situations.

Deficit of the situations or the recovery process shown in Figure 1. Figure 2 provides an illustration of the BCD model parameters for two situations.
This article is limited in scope to the application of the related BCD parameters (*i.e.*, Benefits, Cost, Deficits) and the improvement or the degradation of the Deficits. The logical value of the B, C and D functions for a given evaluation criterion i is qualitative (*i.e.*, subjective). It can be transformed into a numerical (*i.e.*, objective) value by using the function $K_{J,i}(a,b)$ given in Equation (7) and (8):

$$K_{J,i}(a,b) = s_i(b(t_b)) - s_i(a(t_a)) \tag{7}$$

$$K_{J,i}(a,b) = \begin{cases} K_{B,i}(a,b) & \text{if } B_i(a,b) \\ K_{C,i}(a,b) & \text{if } C_i(a,b) \\ K_{D,i}(a,b) & \text{if } D_i(a,b) \vee ID_i(a,b) \vee DD_i(a,b) \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

2.3. The BCD model for comparing action plans or procedures

The Benefits, Costs and Deficits can also be evaluated by comparing the prescribed behaviour (*i.e.*, the prescribed action plan or the prescribed procedure) and the actual deviate behaviour (*i.e.*, the deviate action plan or the deviate procedure). A procedure is a combination of actions, usually in the form of a sequence. Accomplishing an action generates the occurrence of a new situation. The existing barriers are usually related to an explicit or an implicit decision by the human operators to perform an action plan or a procedure that respects these barriers. Decisions to deviate from the expected behaviour may result in barrier removals. In other words, the human operators decide to perform an action plan or a procedure that removes the barriers that they are supposed to respect.
This paper applies the BCD model to compare two specific action plans or procedures: the one that respects the barrier and the one that removes the barrier. All the possible procedures can thus be compared for the same time interval. The evaluation of the BCD parameters for the two procedures $P1$

Figure 3: Evaluation of the BCD model parameters for two procedures.

Figure 4: Illustration of the BCD model parameters for two procedures.

and $P2$ is shown in Figure 3. Figure 4 illustrates the BCD model parameters for two procedures. The severity of criterion i in each action plan or procedure is denoted $s_i(P1)$ and $s_i(P2)$, respectively. The BCD evaluation requires applying the success function instead of the tolerability function. The success function, denoted $Success_{X,i}(P)$, is related to the results of P for a given criterion i for a decision-maker X with respect to the expected results, denoted $Expectedresults_{X,i}$, with an acceptable error ϵ. The quantitative function $K_{J,i}$ becomes in Equation (9) and (10):

$$K_{J,i}(P1, P2) = s_i(P2) - s_i(P1) \tag{9}$$

$$K_{J,i}(P1, P2) = \begin{cases} K_{B,i}(P1, P2) & \text{if } B_i(P1, P2) \\ K_{C,i}(P1, P2) & \text{if } C_i(P1, P2) \\ K_{D,i}(P1, P2) & \text{if } D_i(P1, P2) \vee ID_i(P1, P2) \vee DD_i(P1, P2) \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

This paper presents a new application of the BCD model. This new application uses the BCD model to learn from violations and to predict them by exploiting the BCD model parameters for several performance criteria.

3. The learning systems based on the BCD model

3.1. The problem statement

The learning-based tools proposed in literature are mainly iterative. They try to copy the cognitive learning process that humans use when they learn from their own errors or behaviour. Such automated tools that learn from errors are usually related to iterative learning control (ILC) and its different mechanisms. Iterative learning control detects and minimizes tracking errors, starting with a low initial knowledge level for the possible consequences of the action [16] [17] [18]. After a limited number of iterations, the ILC algorithm adjusts the learning parameters in order to reduce errors between an input signal and an output signal. The tasks governed by ILC are usually repetitive

activities, such as robot movements.
Two main classes of iterative learning control schemes can be identified:

- The feedforward-feedback scheme [19], or the previous cycle learning scheme [20], is an extension of the first class of iterative learning control; it uses the previous iterations to calculate the current one.

- The current cycle learning scheme [20], or the direct adaptive scheme [21], integrates the previous iterations and the error between the desired output signal and the current output signal in order to refine the input signal evaluation.

This paper deals with the current cycle learning scheme. We use the BCD model to take the positive and negative consequences of human behaviour into account in order to propose predictive systems that can learn from these consequences. Two iterative learning control implementations are developed below: a case-based reasoning system and a neural network based system. Both systems were tested to study the feasibility of predicting barrier removals.
For a given iteration, the correct prediction evaluation compares the real observed decision, denoted u_i, with the decision predicted by a reinforced iterative learning tool, denoted u_i^* in Figure 5. An iteration is based on input data vectors, denoted e_i, and the previous iterations that are modelled by the previous input vectors and their associated decisions (e_{i-1}, u_{i-1}). The vector e_i contains a series of triplets (b_k, c_k, d_k) for a given criterion k. FFor each iteration, the vector contains the same number of data related to m criteria: $(b_1, c_1, d_1, b_2, c_2, d_2, ..., b_m, c_m, d_m)$. Each parameter is defined as an interval of values, denoted $\Omega = [X_{min}, X_{max}]$, and the output signal u is defined as an interval of values, denoted $\Psi = [0, 1]$.

Figure 5: The prediction process statement.

3.2. ILC implementation on a case-based reasoning system

The case-based reasoning is very appropriate when no formalized knowledge is available. The principle is reasoning by analogy: when problems are similar, their solutions (*e.g.*, the actions to be performed) are also similar. There are two kinds of case-based reasoning [22]): case-based reasoning for

interpretation and case-based reasoning for problem solving. This paper focuses on a case-based reasoning system for interpretation.
As shown in Figure 6, case-based reasoning can be broken down into four steps. The first step involves developing a case database. A case is represented by a vector pair $< problem, solution >$, and this vector pair is related to the pair $< e, u >$. The second step is to find the problem in the base that is the most similar to a new problem to be solved. This implies interpreting the current problem with respect to current knowledge stored in a database. In interpretative case-based reasoning, similarity functions have to be defined. The third and fourth steps involve adapting the solution and if necessary, improving it. As mentioned in the previous paragraph, the case-

Figure 6: The steps for case-based reasoning [23] [24].

based reasoning system requires defining a similarity function, denoted S, that identifies the pair (e, u) for which e is similar to the input vector e_i in the knowledge stored in the database, denoted E_i in Figure 7. Knowing all the n^{th} previous pairs (e_{i-n}, u_{i-n}), this function S finds a possible input signal u with respect to the input vector e_i. The Euclidean distance value is used to find the vector e, as shown in Equation (11):

Figure 7: Case-based reasoning system for interpretation.

$$\|(e_i)^T - (e)^T\| = \min_{0 \leq k < n} \|(e_i)^T - (e_k)^T\| \tag{11}$$

When the vector e is found, the corresponding decision u is considered as the prediction of u_i, denoted u_i^*. The learning process involves managing the knowledge base E_i by integrating the previous pairs (e_{i-1}, u_{i-1}).

3.3. ILC implementation on a neural network system

Like case-based reasoning system, the neural network system requires a similarity function, denoted S, which identifies the vector $(e \cup u)$, denoted (e, u), for which e is similar to the input vector e_i in the knowledge base, denoted E_i. This database E_i is modified with respect to a reinforcement function, denoted R, to handle the database content shown in Figure 8. To simplify the explanation of the process, the functions that have the same

Figure 8: The iterative learning formalism based on the BCD parameters.

purpose but a different number or type of parameter have the same name (*i.e.* the S and R functions). Knowing all the previous vectors (e, u) of E_i obtained from the previous iterations, the function S finds a possible input signal u with respect to the input vector e_i. The Euclidean distance value is used to find the vector e, as written in Equation (12):

$$\begin{array}{ll} S: & \Omega^{3m} \to \Psi \\ & e_i \to u_i^* = S(e_i), u_i^* = u'/(e \cup u) \cap u' = u, \\ & \forall e_k \in E_i, \|(e_i)^T - (e)^T\| = \min \|(e_i)^T - (e_k)^T\| \end{array} \tag{12}$$

When a vector e is found, the corresponding decision u is considered as the prediction of u_i, denoted u_i^*. Then, the function S is adapted and integrates other parameters. This function tries to find the vector $(e_{i-1}^+ \cup u_{i-1}^+)$, denoted (e_{i-1}^+, u_{i-1}^+), which corresponds to the vector (e_k, u_k) in E_i with minimum differences between the parameters of the vector (e_{i-1}, u_{i-1}), as written in Equation (13):

$$\begin{array}{ll} S: & \Omega^{3m} * \Psi \to \Omega^{3m} * \Psi \\ & (e_{i-1}, u_{i-1}) \to (e_{i-1}^+, u_{i-1}^+) = S(e_{i-1}, u_{i-1}), \\ & \forall (e_k, u_k) \in E_i, \|(e_{i-1}, u_{i-1})^T - (e_{i-1}^+, u_{i-1}^+)^T\| = \min \|(e_{i-1}, u_{i-1})^T - (e_k, u_k)^T\| \end{array} \tag{13}$$

A written in Equation 14, the obtained error ϵ_1 is then processed with the function R in order to reinforce the impact of the vector (e_{i-1}, u_{i-1}). As illustrated in Figure 9, this function R handles the weight parameters related to the vector (e_{i-1}^+, u_{i-1}^+) with a predefined function Δ that allocates a weight with respect to the value ϵ_1, such as the function developed by Kohonen [25]:

Figure 9: Allocation of a weight $\Delta(\epsilon)$ regarding an error value ϵ.

$$\begin{array}{ll} R: & \Omega^{3m} * \Psi \to \Omega^{3m} * \Psi \\ & (e_{i-1}^+, u_{i-1}^+) \to (e_{i-1}^x, u_{i-1}^x) = R(e_{i-1}^+, u_{i-1}^+), \\ & (e_{i-1}^x, u_{i-1}^x)^T = (e_{i-1}^+, u_{i-1}^+)^T + \Delta[(e_{i-1}, u_{i-1})^T - (e_{i-1}^+, u_{i-1}^+)^T] \end{array} \tag{14}$$

The error ϵ_1 is written as shown in Equation (15):

$$\epsilon_1^T = (e_{i-1}, u_{i-1})^T - (e_{i-1}^+, u_{i-1}^+)^T \tag{15}$$

In a second step, the errors ϵ_2 between the vectors $(e,u) \neq (e^x_{i-1}, u^x_{i-1})$ in E_i and the reinforced vector (e^x_{i-1}, u^x_{i-1}) are processed with the function R in order to obtain a new database based on the new vectors, $(e^x, u^x) = R(e,u)$ and (e^x_{i-1}, u^x_{i-1}), as written in Equation (16).

$$
\begin{aligned}
R: \quad & \Omega^{3m} * \Psi \to \Omega^{3m} * \Psi \\
& (e,u) \to (e^x, u^x) = R(e,u), \\
& \forall (e,u) \neq (e^+_{i-1}, u^+_{i-1}), (e^x, u^x) = (e,u) + \Delta[(e^x_{i-1}, u^x_{i-1})^T - (e,u)^T]
\end{aligned} \tag{16}
$$

The error ϵ_2 is written as shown in Equation (17):

$$\epsilon_2^T = (e^x_{i-1}, u^x_{i-1})^T - (e,u)^T \tag{17}$$

As written in Equation (18), the obtained vectors (e^x, u^x) and the vector (e^x_{i-1}, u^x_{i-1}) are gathered into a new database, denoted E^x_i, that replaces E_i:

$$E^x_i = \{(e^x, u^x) \forall (e,u) \neq (e^+_{i-1}, u^+_{i-1}), (e^x, u^x) = R(e,u)\} \cup (e^x_{i-1}, u^x_{i-1}) \tag{18}$$

4. Application to the car driving domain

4.1. The experimental protocol

Figure 10: The car driving simulator.

The experimental protocol involves using the driving simulator shown in Figure 10. During the experiments, drivers faced different situations and barriers (*i.e.*, the driving rules) [26]. In this paper, we examined the priority-to-the-right barrier presented in Figure 11. In fact, there were two priority-to-the-right barriers examined in this study:

- A first priority-to-the-right barrier in a rural context, in which the speed limit is 90 km/h.
- A second priority-to-the-right barrier in an urban context, in which the speed limit is 50 km/h.

Figure 11: The priority-to-the-right barrier.

In both situations, the driver S is approaching an intersection where another car is arriving on the right side. Normally, he/she has to give way to this vehicle coming from the right. Forty-four subjects participated in this experiment. Out of 131 situations - 66 for first priority-to-the-right barrier situation (4 for familiarization purposes) and 65 for the second priority-to-the-right barrier situation (4 for familiarization purposes) - 33 barrier removals were observed. Barrier removal means that the drivers did not give way to the vehicle coming from their right.
In the experiment, the severity was expressed in terms of four performance criteria: safety, opportunity for action, driver comfort and time spent [27]. It was possible to associate the situation evaluation with the BCD model parameters and the observed behaviour (priority respected or not). In order to compare the correct prediction rate using the systems based on the BCD model (*i.e.*, the case-based reasoning system and the neural network system), the input vectors e contained the Benefits, Costs and Deficits for each barrier removed and/or each barrier respected. The BCD model parameters were evaluated for the four criteria stated above.

4.2. The performance criteria

Four performance criteria were chosen to analyze the barrier removal in terms of Benefits, Costs and potential Deficits: safety, opportunity for action, driver comfort and time spent. With respect to the actual position of the car on the road, several points can be identified: admissible points on which the driver is authorized to drive and potential collision points on which the driver is not authorized to drive (Figure 12). At a given instant t_i, the position and the environment (e.g., road geometry, road limits) of the car are known.

Figure 12: Examples of admissible points and potential collision points allocated to a given car driver.

4.2.1. Safety

The safety indicator is used to calculate the severity of the incident based on a predictive time-window. New positions are predicted using a sampling rate for time and steering angle (*e.g.*, 2 seconds and 22,5°). These predicted positions produce a tree-like structure, in which the root is the current position of the vehicle at instant t_i, and the leaves are the predicted positions at instant $t_i + 2$ seconds. In this tree-like structure, each node n_j is characterized by a possible position on the road and an instant t_j. If the position is not a collision point, this position is judged *safe* (*i.e.*, an admissible point); otherwise, the position is judged *dangerous* (*i.e.*, a potential collision point). In the latter case, the dangerousness of a point is inversely proportional to the time (*i.e.*, time of collision). As written in Equation (19), the higher the value of s_{safety}, the more severe (*i.e.*, less safe) the incident.

$$s_{safety} = Card(n_j | position(n_j) = "dangerous") \times \sum(\frac{1}{t_j}) \tag{19}$$

Figure 13 illustrates two situations: in the situation A, s_{safety} (*i.e.*, the

Figure 13: Examples of admissible points and potential collision points for a given driver, with a time-window of 0.5s

severity of the incident in terms of safety) is 0; in the situation B, s_{safety} is 2 (1/0,5). It is possible to compare situations A and B in terms of Benefit and Deficit:

- for the Benefit, B_{safety}(A,B) = false;
- for the Deficit, D_{safety}(A,B) = true, in that a potential collision before 1 second has elapsed is unacceptable.

4.2.2. Opportunity for action

Inside the predictive time-window, the more dangerous points that can be reached, the less chance the driver has to choose a safe action. The incident severity in terms of opportunity for action is given in Equation (20):

$$s_{OpportunitiesForActions} = 1 - \frac{Card(n_j | position(n_j) = "dangerous")}{Card(n_j)} \tag{20}$$

4.2.3. Driver comfort

The comfort criterion is divided in several sub-criteria: the number of braking actions, the number of accelerations, the number of steering actions and the available space around the car. For the first three sub-criteria, the lower the number of manoeuvres, the more comfortable the driving situation. For the last sub-criteria, the more the available the space around the car, the more comfortable the driving situation. An example of the evaluation of the available space around the car is illustrated in Figure 14.

Figure 14: Assessment of the available space to move freely.

4.2.4. Time spent

This indicator, t_s, corresponds to the difference between the acceptable time limit t_a and the time really spent and measured t_m in Equation (21). The acceptable time limit t_a is obtained by dividing the distance d by the speed limitation of the road S.

$$t_s = t_m - t_a \tag{21}$$

After the experiments, the values of the above criteria were evaluated for the priority-to-the-right barriers. The average values were then used as a threshold to determine the Benefits or the Costs: a value under the average threshold is associated to a Benefit, and a value over the average threshold is associated to a Cost. The safety criterion is related to the occurrence of a Deficit, an improvement of a Deficit or a decrease in a Deficit. The other criteria are related to Benefits or Costs. These criteria were evaluated for specific geographical zones, such as the zone where the car is located with respect to the priority-to-the-right barrier and the zone where the car is situated on the crossroads.

4.3. The results

The two systems based on the BCD model (case-based reasoning and neural network) were then used to study the rate of correct predictions, by comparing the predicted u_i and the real observed u_i^* (*i.e.*, the prediction values they produced and the real observed values). The predictions involve determining whether or not a driver will remove a given barrier. Results for the two systems for the first priority-to-the-right barrier situation are reported in Figure 15. The predictions were for 66 iterations, with 4 iterations used for familiarization purposes. Results for the second priority-to-the-right

Figure 15: Prediction results for the first priority-to-the-right barrier removal.

barrier situation are reported in Figure 16. The predictions were for 65 iterations, with 4 iterations used for familiarization purposes. The results shown in Figure 15 and 16 prove that our systems are able to learn from barrier removals interpreted in term of Benefits, Costs and Deficits. The

Figure 16: Prediction results of the second priority-to-the-right barrier removal.

rate of correct predictions increases progressively for both priority-to-the-right barrier situations and converges between 70% and 100%. Both these systems give acceptable correct prediction rates, and they are complementary because their prediction rates sometimes concerns different cases.

5. Conclusion

This paper described the BCD model framework in order to analyze human behaviour in terms of Benefits, Costs and Deficits, comparing the consequences of situations and action plans or procedures. Using this BCD model framework, we defined and validated both a case-based reasoning system and a neural network system for predicting barrier removals. These systems were applied to car driving, identifying Benefits, Costs and Deficits with respect to four performance criteria: safety, opportunity for action, driver comfort, and time spent. The rate of correct predictions obtained by these systems increased with the evolving learning process based on the BCD model parameters. Therefore, this study has shown that it is possible to design systems capable of learning from human errors interpreted in terms of Benefits, Costs and Deficits in off-line experimental conditions.
The maximum obtained rate is over 80% after only 10 iterations. Further studies aim at determining specific situations that are correctly predicted by each learning system. The rate of correct prediction could then be increased by merging the knowledge of each system.
Two main perspectives are possible:

- The improvement of the BCD model. Several evolution of the model may be studied: the integration of probability of success or failure of barrier removal, the weighting of the BCD parameters, the selection of optimal action plan, and the integration of organizational factor such as cooperative or competitive factors [28].

- The improvement of prediction systems. Some researches have to be done at different levels in order to study the feasibility of several alternatives: the identification of minimum relevant data in the input framework, the extension of the application domain, the merging of knowledge of different systems, the learning from organizational factor to design for instance human error tolerant barrier [29], the application of allocation criteria such as those defined on [30] related to the prediction systems, and the on-line prediction by implementing onboard learning systems.

6. Acknowledgments

The present research work has been supported by
The International Campus on Safety and Intermodality in Transportation,
The Nord-Pas-de-Calais Region,
The European Community,
The Regional Delegation for Research and Technology,
The Ministry of Higher Education and Research,
The Scientific Research Group on Supervisory, Safety and Security of Complex Systems,
The European Research Group on Human-Machine Systems in Transportation,
And the National Center for Scientific Research.
The authors gratefully acknowledge the support of these institutions.

References

[1] A. Swain, H. Guttmann, Handbook of human reliability analysis with emphasis on nuclear power plant applications-final report, NUREG/CR-1278, Sandia National Labs., Albuquerque, NM (USA), 1983.

[2] B. Kirwan, The validation of three human reliability quantification techniquesTHERP, HEART and JHEDI: part iiiPractical aspects of the usage of the techniques, Applied Ergonomics 28 (1) (1997) 27–39.

[3] B. Reer, Review of advances in human reliability analysis of errors of commissionPart 2: EOC quantification, Reliability Engineering and System Safety 93 (8) (2008) 1105–1122.

[4] P. Polet, F. Vanderhaegen, P. Wieringa, Theory of safety-related violations of system barriers, Cognition, Technology & Work 4 (3) (2002) 171–179.

[5] M. Čepin, DEPEND-HRA - A method for consideration of dependency in human reliability analysis, Reliability Engineering and System Safety 93 (10) (2008) 1452–1460.

[6] F. Vanderhaegen, D. Jouglet, S. Piechowiak, Human-reliability analysis of cooperative redundancy to support diagnosis, IEEE Transactions on Reliability 53 (4) (2004) 458–464.

[7] J. Reason, Human error. 1990.

[8] A. Carpignano, M. Piccini, Cognitive theories and engineering approaches for safety assessment and design of automated systems: a case study of a power plant, Cognition, Technology & Work 1 (1) (1999) 47–61.

[9] P. Polet, F. Vanderhaegen, R. Amalberti, Modelling border-line tolerated conditions of use (BTCU) and associated risks, Safety Science 41 (2-3) (2003) 111–136.

[10] P. Polet, F. Vanderhaegen, P. Millot, Human behaviour analysis of barrier deviations using a benefit-cost-deficit model, Advances in Human-Computer Interaction 2009.

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

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

[13] F. Robache, H. Morvan, P. Polet, M.-P. Pacaux-Lemoine, F. Vanderhaegen, The BCD model for biomechanical application, in: Proceedings of the 25th European Annual Conference on Human Decision Making and Manual Control, EAM'06. Valenciennes, France, 2006.

[14] M.-P. Pacaux-Lemoine, F. Vanderhaegen, BCD model for human state identification, in: Proceedings of the 10th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design, and Evaluation of Human-Machine Systems, Seoul, Korea, 2007.

[15] F. Vanderhaegen, The benefit-cost-deficit (BCD) model for human analysis and control, in: Proceedings of the Ninth IFAC/IFORS/IEA Symposium on Analysis, Design, and Evaluation of Human-Machine Systems, Atlanta, GA, USA, 2004, pp. 7–9.

[16] J. Xu, R. Yan, Iterative learning control design without a priori knowledge of the control direction, Automatica 40 (10) (2004) 1803–1809.

[17] C. Chien, C. Yao, An output-based adaptive iterative learning controller for high relative degree uncertain linear systems, Automatica 40 (1) (2004) 145–153.

[18] M. Norrlöf, S. Gunnarsson, A note on causal and CITE iterative learning control algorithms, Automatica 41 (2) (2005) 345–350.

[19] T. Lee, K. Tan, S. Lim, H. Dou, Iterative learning control of permanent magnet linear motor with relay automatic tuning, Mechatronics 10 (1-2) (2000) 169–190.

[20] J. Xu, T. Heng Lee, H. Zhang, Analysis and comparison of iterative learning control schemes, Engineering Applications of Artificial Intelligence 17 (6) (2004) 675–686.

[21] Y. Wang, C. Chien, C. Teng, Direct adaptive iterative learning control of nonlinear systems using an output-recurrent fuzzy neural network, IEEE Transactions on Systems, Man, and Cybernetics, Part B 34 (3) (2004) 1348–1359.

[22] I. Bichindaritz, Apprentissage de concepts dans une mémoire dynamique : Raisonnement a partir de cas adaptable la tache cognitive, Ph.D. thesis, Paris V, France (1994).

[23] J.-P. Haton, Le Raisonnement Humain en Intelligence Artificielle, Modèles, Techniques, et Architecture pour les Systèmes à Base de Connaissances.

[24] S. Slade, Case-based reasoning: a research paradigm, AI magazine 12 (1) (1991) 42.

[25] T. Kohonen, Data Management by Self-Organizing Maps, Lecture Notes in Computer Science 5050 (2008) 309.

[26] A. Chaali-Djelassi, F. Vanderhaegen, Predication of violations in road transportation system, in: Proceedings of the 4th International Conference on Safety and Reliability, Krakow, 2006, pp. 57–68.

[27] A. Chaali-Djelassi, P. Polet, F. Vanderhaegen, Ponderation functions of the car driver decisions, in: Proceedings of the 6th Multidisciplinary International Conference on Quality and Dependability, QUALITA'05, Bordeaux, France, 2005.

[28] F. Vanderhaegen, S. Chalmé, F. Anceaux, P. Millot, Principles of cooperation and competition: application to car driver behavior analysis, Cognition, Technology & Work 8 (3) (2006) 183–192.

[29] F. Vanderhaegen, Human-error-based design of barriers and analysis of their uses, Cognition, Technology & Work 12 (2) (2010) 133–142.

[30] F. Vanderhaegen, Multilevel organization design: The case of the air traffic control, Control Engineering Practice 5 (3) (1997) 391–399.